Cross-Sectional Methods for Empirical Asset Pricing

Valentina Raponi

June 14, 2019
Abstract

This thesis develops new methods in empirical asset pricing which are valid when a large number of assets is available for the analysis. The work is divided in three main chapters, each of them focusing on different aspects and issues typically related to asset-pricing models.

The first chapter introduces a methodology for estimating and testing beta-pricing models when a large number of assets is available for investment but the number of time-series observations is fixed. We first consider the case of correctly specified models with constant risk premia, and then extend our framework to deal with time-varying risk premia, potentially misspecified models, firm characteristics, and unbalanced panels. We show that our large cross-sectional framework poses a serious challenge to common empirical findings regarding the validity of beta-pricing models. Firm characteristics are found to explain a much larger proportion of variation in estimated expected returns than betas.

The second chapter investigates the effect of model misspecification on mean-variance portfolios and show how asset-pricing theory and asymptotic analysis can mitigate misspecification. The analysis is founded on the Arbitrage Pricing Theory (APT), because it allows for pricing errors. The APT is extended to show it can capture not just small pricing errors unrelated to factors but also large pricing errors from mismeasured and missing factors. The key insight is that, instead of treating misspecification directly in the mean-variance portfolio, it is better to first decompose the portfolio into a "beta" portfolio that depends only on factor risk premia and an "alpha" portfolio that depends only on pricing errors. Then, as the number of assets increases, we show that the weights of the alpha portfolio dominate those of the beta portfolio, leading to mean-variance portfolio weights that are immune to beta misspecification. For the alpha portfolio, misspecification is treated by imposing the APT restriction, which serves as an identification condition and a shrinkage constraint. Using simulations, we illustrate how our theoretical insights lead to a significant improvement in the out-of-sample performance of mean-variance portfolios.

The third chapter analyzes the large cross-sectional properties of the standard two-pass methodology, when useless factors are included in the beta-pricing specification. When the number of time-series observations, $T$, is assumed to be fixed, and contrary to the conventional large-$T$ framework, we find that the simple two-pass OLS estimator of risk premia exhibits desirable asymptotic properties that can be used to detect useless factors. In particular, we derived correctly-sized $t$-ratios, $F$-tests and goodness-of-fit measures that allow us to implement a powerful statistical strategy to test for factors that can be potentially irrelevant for the analysis. The results hold also under the assumption of potential model misspecification. The validity of our results is assessed by means of simulation exercises.
# Contents

1 Testing Beta-Pricing Models Using Large Cross-Sections  
   Abstract .......................................................... 4  
   1.1. Introduction .................................................. 5  
   1.2. The Two-Pass Methodology .................................. 6  
   1.3. Asymptotic Analysis under Correctly Specified Models  
      1.3.1 Baseline case ........................................... 10  
      1.3.2 Time-varying case ..................................... 16  
   1.4. Asymptotic Analysis under Potentially Misspecified Models  
      1.4.1 Testing for model misspecification ................. 16  
      1.4.2 Estimation under potential model misspecification  
           1.4.3 Misspecification due to priced characteristics  
      .. 22  
   1.5. Unbalanced Panels .......................................... 24  
   1.6. Monte Carlo simulations .................................. 26  
      1.6.1 Percentage errors and root mean squared errors of  
           the estimates ........................................... 30  
      1.6.2 Rejection rates of the t-tests ................. 32  
      1.6.3 Rejection rates of the specification test ....... 33  
   1.7. Empirical Analysis ......................................... 30  
      1.7.1 Data ................................................... 48  
      1.7.2 Specification testing ................................ 48  
      1.7.3 Risk premia estimates ................................ 49  
      1.7.4 Characteristics ...................................... 51  
      1.7.5 CAPM, Fama and French (1993) Three-Factor Model,  
           and Fama and French (2015) Five-Factor Model ...... 53  
   1.8. Conclusion .................................................. 56  
   Appendix I ....................................................... 60  
   1.A. Lemmas ...................................................... 68  
   1.B. Proofs of Propositions and Theorems .................. 69  
   1.C. Explicit Form of $U_c$ .................................. 70  
   1.D. Random Betas ............................................... 71  
   1.E. Nonparametric Estimation of Risk Premia on Traded Factors .. 71  

2 Portfolio Choice with Model Misspecification  
   Abstract .......................................................... 72  
   2.1. Introduction ................................................ 73  
   2.2. Related Literature ......................................... 74  
   2.3. Generalizing the APT ...................................... 75  
      2.3.1 Notation ............................................... 75  
      2.3.2 Linear factor model for asset returns with  
           misspecification ........................................ 75  
      2.3.3 Arbitrage Pricing Theory (APT) ..................... 76  
      2.3.4 Extending the APT .................................... 76
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.5 Different forms of model misspecification</td>
<td>128</td>
</tr>
<tr>
<td>2.4. Mitigating Model Misspecification</td>
<td></td>
</tr>
<tr>
<td>2.4.1 Decomposing the mean-variance portfolio</td>
<td>129</td>
</tr>
<tr>
<td>2.4.2 Mitigating misspecification in the beta component of returns</td>
<td>130</td>
</tr>
<tr>
<td>2.4.3 Mitigating misspecification in the alpha component of returns</td>
<td>134</td>
</tr>
<tr>
<td>2.5. Numerical Illustration</td>
<td></td>
</tr>
<tr>
<td>2.5.1 Simulation design and performance evaluation</td>
<td>138</td>
</tr>
<tr>
<td>2.5.2 Model misspecification in the beta component of returns</td>
<td>142</td>
</tr>
<tr>
<td>2.5.3 Model misspecification in the alpha component of returns</td>
<td>143</td>
</tr>
<tr>
<td>2.6. Conclusion</td>
<td>145</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>146</td>
</tr>
<tr>
<td>2.A. Proofs for Theorems</td>
<td></td>
</tr>
<tr>
<td>2.A.1 Lemma on decomposition of the Sharpe ratio</td>
<td>147</td>
</tr>
<tr>
<td>2.A.2 Extension of Roll (1980)</td>
<td>148</td>
</tr>
<tr>
<td>2.A.3 Equivalent representations for the portfolio $w_N^t$</td>
<td>149</td>
</tr>
<tr>
<td>2.A.4 APT restriction in terms of projection errors</td>
<td>150</td>
</tr>
<tr>
<td>2.A.5 Proof of Theorem 2.4</td>
<td>151</td>
</tr>
<tr>
<td>2.A.6 Proof of Theorem 2.5</td>
<td>152</td>
</tr>
<tr>
<td>2.A.7 Proof of Theorem 2.6</td>
<td>153</td>
</tr>
<tr>
<td>2.A.8 Proof of Theorem 2.8</td>
<td>154</td>
</tr>
<tr>
<td>2.A.9 Proof of Theorem 2.9</td>
<td>155</td>
</tr>
<tr>
<td>2.A.10 Proof of Theorem 2.10</td>
<td>156</td>
</tr>
<tr>
<td>2.A.11 Proof of Theorem 2.11</td>
<td>157</td>
</tr>
<tr>
<td>2.A.12 Estimation for unbounded residual variation case with mismeasured factors</td>
<td>158</td>
</tr>
<tr>
<td>3 Detecting Spurious Factors using Cross-Sectional Regressions</td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>161</td>
</tr>
<tr>
<td>3.1. Introduction</td>
<td>162</td>
</tr>
<tr>
<td>3.2. Two pass-regression with useless factors when $N$ is large and $T$ is fixed</td>
<td>163</td>
</tr>
<tr>
<td>3.2.1 Two-pass regression with useless factors under correct model specification</td>
<td>164</td>
</tr>
<tr>
<td>3.2.2 The base case under potential model misspecification</td>
<td>165</td>
</tr>
<tr>
<td>3.3. Including useless factors in a two-pass regression with known useful factors</td>
<td>166</td>
</tr>
<tr>
<td>3.4. Two-pass regression with useless factors and omitted useful factors</td>
<td>167</td>
</tr>
<tr>
<td>3.5. Simulation results</td>
<td></td>
</tr>
<tr>
<td>3.5.1 Base case</td>
<td>168</td>
</tr>
<tr>
<td>3.5.2 Correctly specified models with useful and useless factors</td>
<td>169</td>
</tr>
<tr>
<td>3.6. Conclusion</td>
<td>170</td>
</tr>
<tr>
<td>Appendix 3</td>
<td>171</td>
</tr>
<tr>
<td>3.A. Assumptions</td>
<td></td>
</tr>
<tr>
<td>3.B. Lemmas</td>
<td>172</td>
</tr>
<tr>
<td>3.C. Proofs of Theorems</td>
<td>173</td>
</tr>
<tr>
<td>References</td>
<td>262</td>
</tr>
</tbody>
</table>
List of Tables

1.1 Bias and root mean squared error of the OLS and bias-adjusted Shanken (1992) estimators in a one-factor model ($\Sigma$ spherical) .................................................. 34
1.2 Bias and root mean squared error of the OLS and bias-adjusted Shanken (1992) estimators in a one-factor model ($\Sigma$ diagonal) ............................. 35
1.3 Bias and root mean squared error of the OLS and bias-adjusted Shanken (1992) estimators in a one-factor model ($\Sigma$ full, $\delta = 0.5$) ............. 36
1.4 Bias and root mean squared error of the OLS and bias-adjusted Shanken (1992) estimators in a one-factor model ($\Sigma$ full, $\delta = 0.25$) ............. 37
1.5 Size of $t$-tests in a one-factor model ($\Sigma$ spherical) ........................................ 38
1.5 (Continued) Size of $t$-tests in a one-factor model ($\Sigma$ spherical) ...... 39
1.6 Size of $t$-tests in a one-factor model ($\Sigma$ diagonal) ........................................ 40
1.6 (Continued) Size of $t$-tests in a one-Factor model ($\Sigma$ diagonal) .......... 41
1.7 Size of $t$-tests in a one-factor model ($\Sigma$ full, $\delta = 0.5$) ............ 42
1.7 (Continued) [0pt] Size of $t$-tests in a one-factor model ($\Sigma$ full, $\delta = 0.5$) .... 43
1.8 [0pt] Size of $t$-tests in a one-factor model ($\Sigma$ full, $\delta = 0.25$) ............... 44
1.8 (Continued) Size of $t$-tests in a one-factor model ($\Sigma$ full, $\delta = 0.25$) ... 45
1.9 [0pt] Rejection rates of the specification test in a one-factor model ($T = 36$) 46
1.10 [0pt] Rejection rates of the specification test in a one-factor model ($T = 72$) 47
1.11 Percentage difference between estimated risk premia .................. 115
1.12 Betas versus Characteristics ................................................................. 116

2.1 Out-of-Sample Sharpe Ratios with Beta Misspecification ................. 144
2.2 Out-of-Sample Sharpe Ratios with Alpha Misspecification ................. 145

3.1 Bias and RMSE of the OLS Estimator in a one-factor model with a useless factor ($\Sigma$ scalar) ................................................................. 181
3.2 Bias and RMSE of the OLS Estimator in a one-factor model with a useless factor ($\Sigma$ Diagonal) ............................................................. 182
3.3 Bias and RMSE of the OLS Estimator in a one-factor model with a useless factor ($\Sigma$ Full - $\delta = 0.5$). .................................................. 183
3.4 Bias and RMSE of the OLS Estimator in a one-factor model with a useless factor ($\Sigma$ Full - $\delta = 0.25$) .................................................. 184
3.5 Empirical size of $t$-tests in a one-factor model with a useless factor ($\Sigma$ Scalar) 185
<table>
<thead>
<tr>
<th>Table Number</th>
<th>Table Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6</td>
<td>Empirical size of t-tests in a one-factor model with a useless factor (Σ Diagonal)</td>
<td>186</td>
</tr>
<tr>
<td>3.7</td>
<td>Empirical size of t-tests in a one-factor model with a useless factor (Σ Full - δ = 0.5)</td>
<td>187</td>
</tr>
<tr>
<td>3.8</td>
<td>Empirical size of t-tests in a one-factor model with a useless factor (Σ Full - δ = 0.25)</td>
<td>188</td>
</tr>
<tr>
<td>3.9</td>
<td>Empirical size of F-tests in a one-factor model with a useless factor (Σ scalar)</td>
<td>189</td>
</tr>
<tr>
<td>3.10</td>
<td>Empirical size of F-tests in a one-factor model with a useless factor (Σ Diagonal)</td>
<td>189</td>
</tr>
<tr>
<td>3.11</td>
<td>Empirical size of F-tests in a one-factor model with a useless factor (Σ Full - δ = 0.5)</td>
<td>190</td>
</tr>
<tr>
<td>3.12</td>
<td>Empirical size of F-tests in a one-factor model with a useless factor (Σ Full - δ = 0.25)</td>
<td>190</td>
</tr>
<tr>
<td>3.13</td>
<td>Bias and RMSE of the OLS Estimator in a correctly specified model with useful and useless factors (Σ scalar)</td>
<td>192</td>
</tr>
<tr>
<td>3.14</td>
<td>Bias and RMSE of the OLS Estimator in a correctly specified model with useful and useless factors (Σ Diagonal)</td>
<td>193</td>
</tr>
<tr>
<td>3.15</td>
<td>Bias and RMSE of the OLS Estimator in a correctly specified model with useful and useless factors (Σ Full, δ = 0.5)</td>
<td>194</td>
</tr>
<tr>
<td>3.16</td>
<td>Empirical Size of t-tests in a correctly specified model with useful and useless factors (Σ Scalar)</td>
<td>195</td>
</tr>
<tr>
<td>3.17</td>
<td>Empirical Size of t-tests in a correctly specified model with useful and useless factors (Σ Diagonal)</td>
<td>196</td>
</tr>
<tr>
<td>3.18</td>
<td>Empirical Size of t-tests in a correctly specified model with useful and useless factors (Σ Full - δ = 0.5)</td>
<td>197</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Specification testing for the Fama and French (2015) five-factor model ............ 54
1.2 Specification testing for the liquidity-augmented Fama and French (2015) five-factor model ............................................................. 55
1.3 Specification testing for the Fama and French (2015) five-factor model using the Gibbons et al. (1989a) and Gungor and Luger (2016) tests ............................. 56
1.4 Estimates and confidence intervals for the market risk premium .................. 57
1.5 Estimates and confidence intervals for the time-varying market risk premium . 58
1.6 Estimates and confidence intervals for the liquidity risk premium ............... 59
1.7 Estimates and confidence intervals for the time-varying liquidity risk premium 60
1.8 Estimates and confidence intervals for the characteristic premia ............... 61
1.9 Specification testing for CAPM ............................................................... 62
1.10 Specification testing for the liquidity-augmented CAPM .......................... 63
1.11 Specification testing for CAPM using the Gibbons et al. (1989a) and Gungor and Luger (2016) tests ...................................................... 64
1.12 Estimates and confidence intervals for the market risk premium (CAPM) ...... 65
1.13 Estimates and confidence intervals for the time-varying market risk premium (CAPM) ........................................................................ 66
1.14 Estimates and confidence intervals for the liquidity risk premium (liquidity-augmented CAPM) ............................................................... 67
1.15 Estimates and confidence intervals for the time-varying liquidity risk premium (liquidity-augmented CAPM) .................................................. 68
1.16 Specification testing for the Fama and French (1993) three-factor model .... 69
1.17 Specification testing for the liquidity-augmented Fama and French (1993) three-factor model .............................................................. 70
1.18 Specification testing for the Fama and French (1993) three-factor model using the Gibbons et al. (1989a) and Gungor and Luger (2016) tests .................. 71
1.19 Estimates and confidence intervals for the market risk premium (FF3) ........ 72
1.20 Estimates and confidence intervals for the size premium (FF3) ................... 73
1.21 Estimates and confidence intervals for the value premium (FF3) ............... 74
1.22 Estimates and confidence intervals for the liquidity risk premium (liquidity-augmented FF3) ............................................................... 75
1.23 Estimates and confidence intervals for the time-varying market risk premium (FF3) .............................................................. 76
1.24 Estimates and confidence intervals for the time-varying size premium (FF3) . 77
1.25 Estimates and confidence intervals for the time-varying value premium (FF3) 78
1.26 Estimates and confidence intervals for the time-varying liquidity risk premium (liquidity-augmented FF3) .................................................. 79
1.27 Estimates and confidence intervals for the size premium (FF5) ................... 80
1.28 Estimates and confidence intervals for the value premium (FF5) ............... 81
1.29 Estimates and confidence intervals for the profitability premium (FF5) ........ 82
1.30 Estimates and confidence intervals for the investment premium (FF5) ........ 83
1.31 Estimates and confidence intervals for the time-varying size premium (FF5) . 84
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.32</td>
<td>Estimates and confidence intervals for the time-varying value premium (FF5)</td>
<td>85</td>
</tr>
<tr>
<td>1.33</td>
<td>Estimates and confidence intervals for the time-varying profitability premium (FF5)</td>
<td>86</td>
</tr>
<tr>
<td>1.34</td>
<td>Estimates and confidence intervals for the time-varying investment premium (FF5)</td>
<td>87</td>
</tr>
<tr>
<td>2.1</td>
<td>Weighted sum of the squared pricing errors</td>
<td>127</td>
</tr>
<tr>
<td>2.2</td>
<td>Mean-variance and minimum-variance frontier portfolios</td>
<td>130</td>
</tr>
<tr>
<td>2.3</td>
<td>Decomposition of the mean-variance portfolio</td>
<td>132</td>
</tr>
<tr>
<td>2.4</td>
<td>Typical weights of the $w^*_N$ and $w^d_N$ portfolios</td>
<td>133</td>
</tr>
<tr>
<td>2.5</td>
<td>Relative average magnitude of weights of $w^*_N$ and $w^d_N$ portfolios</td>
<td>135</td>
</tr>
<tr>
<td>2.6</td>
<td>Sharpe ratio of $w^*_N$ and $w^{bunch}_N$ relative to that of $w^v_N$ portfolio</td>
<td>138</td>
</tr>
</tbody>
</table>
Declaration of originality

I hereby certify that this thesis constitutes my own work and that all material, which is not my own work, has been properly acknowledged.

Valentina Raponi
Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution - Non Commercial - No Derivatives 4.0 International Licence (CC BY-NC-ND). Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.
Acknowledgments

This thesis is the summary of tough yet, a wonderful and an unforgettable journey, and I feel that I still have miles to go. Throughout this long journey, I have gained a lot by learning to persevere despite hardship. Although only my name is printed on the cover of this PhD thesis, many people around me deserve credits for its completion. Therefore, at this point acknowledgments becomes not duties but rather a personal pleasure.

I shall eternally be grateful to my supervisor, Prof. Paolo Zaffaroni. I am indebted to him for he has not only guided me throughout the course of my thesis work but he has been a constant source of encouragement, support and motivation for me. He has been always there providing his heartfelt guidance at all times and has given me invaluable assistance, inspiration and suggestions. Without him, this thesis would not have been possible and I could not have wished for a better advisor. It has been a privilege learning from him.

In my journey, I have found a teacher, a friend, an inspiration, a role model and a pillar of support in Prof. Raman Uppal. Thank you, Raman, for being with me throughout and helping me come out of the difficult times with a smile on my face. I am extending my heartfelt thanks to his wife and daughter, for their unconditional acceptance, patience and hospitality over these past years.

My profound gratitude goes to Prof. Cesare Robotti, who has been a second advisor and an extremely helpful co-author at Imperial College. I am sincerely grateful to him for sharing his undoubted experience with me, encouraging my research and allowing me to grow as a research scientist. His advice on both research as well as on my career have been invaluable.

There is a long list of other faculty members who have contributed to this thesis. I would particularly like to thank Pasquale Della Corte for all the advice and guidance he has provided in the past five years. He always has had good suggestions to improve the presentation of my papers, and I am sure this thesis would have looked quite different had it not been for his input.

A warm thank you goes to my friends at Imperial College, Adelina, Engin, Mobeen and Robert, for making this journey an amazing experience. They have been my second family during these past five years and have all made an essential contribution in helping me reach this stage in my life. I thank them for putting up with me in difficult moments where I felt lost and surrendered and for goading me on to follow my dream of getting this degree. Thank you guys for sharing your life with me and for your friendship.

I feel short of words to describe the unconditional help, support, love and encouragement I have always received from my parents and sister. This thesis is only a small repayment of all the efforts they took for me. Although I never managed to explain what my research was actually about, they have been the biggest source of my strength. Without them I would not have been where I am today and what I am today. Thank you for being my first supporters who always believed in me. Thank you for teaching me that my job in life was to learn, to be happy, and to know and understand myself; only then could I know and understand others. This one is for you!

Thank you all, also to people and friends I have surely forgotten. I wish they knew their names are missing only in this list, but not in my heart. Thank you to those who will continue to be part of my life and to those who are not by my side anymore. I hope one day they will understand what this thesis actually meant to me.

Valentina
Chapter 1

Testing Beta-Pricing Models Using Large Cross-Sections

Valentina Raponi, Imperial College London
Cesare Robotti, University of Warwick
Paolo Zaffaroni, Imperial College London
Abstract

We propose a methodology for estimating and testing beta-pricing models when a large number of assets is available for investment but the number of time-series observations is fixed. We first consider the case of correctly specified models with constant risk premia, and then extend our framework to deal with time-varying risk premia, potentially misspecified models, firm characteristics, and unbalanced panels. We show that our large cross-sectional framework poses a serious challenge to common empirical findings regarding the validity of beta-pricing models. Firm characteristics are found to explain a much larger proportion of variation in estimated expected returns than betas.

Keywords: beta-pricing models; ex post risk premia; two-pass cross-sectional regressions; time-varying risk premia; model misspecification; firm characteristics; specification test; unbalanced panel; large-N asymptotics.

JEL classification: C12, C13, G12.
1.1. Introduction

Traditional econometric methodologies for estimating risk premia and testing beta-pricing models hinge on a large time-series sample size, \( T \), and a small number of securities, \( N \). At the same time, the thousands of stocks that are traded on a daily basis in financial markets provide a rich investment universe and an interesting laboratory for risk premia and cost of capital determination.\(^1\) Moreover, although we have approximately a hundred years of US equity data, much shorter time series are typically used in empirical work to mitigate concerns of structural breaks and to bypass the difficult issue of modelling explicitly the time variation in risk premia. Finally, when considering non-US financial markets, only short time series are typically available.\(^2\) Importantly, when \( N \) is large and \( T \) is small, the asymptotic distribution of any traditional risk premium estimator provides a poor approximation to its finite-sample distribution, thus rendering the statistical inference problematic.\(^3\)

The main contribution of this paper is that it provides a methodology built on the large-\( N \) estimator of Shanken (1992), which allows us to perform valid inference on risk premia and assess the validity of the beta-pricing relation when \( N \) is large and \( T \) is fixed, possibly very small.\(^4\) Our novel methods are first illustrated for correctly specified models with constant risk premia and then extended to deal with time variation in risk premia, potential model misspecification, firm characteristics in the risk-return relation, and unbalanced panels. We also demonstrate that methodologies specifically designed for a large \( T \) and fixed \( N \) environment are no longer applicable when a large number of assets is used. Proposition 1.3 below demonstrates the perils of inadvertently using the Fama and MacBeth (1973) \( t \)-ratios with the Shanken (1992) correction in our large \( N \) setting.

As emphasized by Shanken (1992), when \( T \) is fixed, one cannot reasonably hope for a consistent estimate of the traditional \( \text{ex ante} \) risk premium. For this reason, we focus on the \( \text{ex post} \) risk premia, which equal the \( \text{ex ante} \) risk premia plus the unexpected factor outcomes.\(^5\)

We start by considering the baseline case of a correctly specified beta-pricing model with constant risk premia when a balanced panel of test asset returns is available. We show that the estimator of Shanken (1992) is free of any pre-testing biases and that no data has to be sacrificed for the preliminary estimation of the bias. (See Proposition 1 below.) Next, we establish the asymptotic properties of the estimator, namely its \( \sqrt{N} \)-consistency and asymptotic normality. We derive an explicit expression for the estimator's asymptotic covariance matrix and show how this expression can be used to construct correctly sized confidence intervals for the risk premia. Our technical assumptions are relatively mild and easily verifiable. In particular, we allow for a substantial degree of cross-correlation among returns (conditional on the factors' realizations), and our assumptions are even weaker than the ones behind the Arbitrage Pricing Theory (APT) of Ross (1976).

In the first extension of the baseline methodology, we demonstrate that the estimator continues to exhibit attractive properties even when risk premia vary over time. In particular, it accurately describes the time-averages of the (time-varying) risk premia over a \( \text{fixed} \) time interval. We also derive a suitably modified version of the estimator that permits valid inference on risk premia at any given point in time. Noticeably,\(^1\)

\(^1\)For example, one can download the returns on 18,474 US stocks for December 2013 from the Center for Research in Security Prices (CRSP), half of which are actively traded.

\(^2\)For example, Table 1 in Hou et al. (2011) shows that, at most, only about thirty years of equity return data is available for emerging economies in Latin America, Europe-Middle East-Africa, and Asia-Pacific regions.

\(^3\) The alternative approach of increasing the time-series frequency, although appealing, can lead to complications and is not always implementable. Potential problems with this approach include non-synchronous trading and market microstructure noise. Furthermore, for models that include non-traded (macroeconomic) risk factors, high-frequency data is not available.

\(^4\)Our methodology offers an alternative to the common practice of employing a relatively small number of portfolios for the purpose of estimating and testing beta-pricing models. Although the use of portfolios is typically motivated by the attempt of reducing data noisiness, it can also cause loss of information and lead to misleading inference due to data aggregation. (See, for example, Brennan et al. (1998), Berk (2002), and Ang et al. (2018), among others.)

\(^5\)The \( \text{ex post} \) risk premium is a parameter with several attractive properties. It is unbiased for the \( \text{ex ante} \) risk premium, and the beta-pricing model is still linear in the \( \text{ex post} \) risk premia under the assumptions of either correctly specified or misspecified models. Finally, the corresponding \( \text{ex post} \) pricing errors can be used to assess the validity of a given beta-pricing model when \( T \) is fixed. Naturally, when \( T \) becomes large, any discrepancy between the \( \text{ex ante} \) and \( \text{ex post} \) risk premia vanishes because the sample mean of the factors converges to its population mean.
in our analysis we do not need to take a stand on the form of time variation in risk premia. Our time-varying risk premium estimator can accommodate non-traded as well as traded factors. For the latter, the traditional estimator based on the factors' rolling sample mean is asymptotically valid for the true risk premium at a given point in time only for specific sampling schemes, and it requires a very large $T$ to work when time variation is allowed for. (See Appendix 1E for details.)

Next, we allow for the possibility that the beta-pricing model is misspecified. We provide a new test of the validity of the beta-pricing relation and derive its large-$N$ distribution under the null hypothesis that the model is correctly specified.

Moreover, we show that our test enjoys nice size and power properties. We then establish the statistical properties of the estimator when the beta-pricing model is misspecified. This extension is particularly relevant when we reject the model's validity based on the outcome of the specification test, but we are still interested in estimating the risk premia of a model with a possibly incomplete set of factors. Finally, we study an important case of deviations from exact pricing, that is, the cross-sectional dependence of expected returns on firm characteristics. The asymptotic covariance matrix of the normally distributed characteristic premia estimator is derived in closed form, unlike most approaches in this literature that typically rely on simulation-based arguments for inference purposes. Our method can be used to determine whether the beta-pricing model is invalid and to quantify the economic importance of the characteristics when there are deviations from exact pricing. By employing a new measure, which is immune to the often-mentioned cross-correlations between estimated betas and characteristics, we are able to determine the relative contribution of betas and characteristics to the overall cross-sectional variation in expected returns.

In the last methodological extension of our baseline analysis, we consider the case of unbalanced panels. This is a useful extension because eliminating observations for the sole purpose of obtaining a balanced panel could result in unnecessarily large confidence intervals for the risk premia and loss of power of the specification test.

We demonstrate the usefulness of our methodology by means of several empirical analyses. The three prominent beta-pricing specifications that we consider are the Capital Asset Pricing Model (CAPM), the three-factor Fama and French (1993) model (FF3), and the recently proposed five-factor Fama and French (2015) model (FF5). We also consider variants of these models augmented with the non-traded liquidity factor of Pástor and Stambaugh (2003). Our proposed methods under potential model misspecification uncover a significant pricing ability for all the traded factors in each of the three models, even when using a relatively short time window of three years. In contrast, the risk premia estimates often appear to be statistically insignificant when using the traditional large-$T$ approaches. Based on our methodology, the liquidity factor appears to be priced in only about one-fifth of the three-year rolling samples examined. We also document strong patterns of time variation in risk premia, for both traded and non-traded factors. In addition, our specification test rejects all beta-pricing models (with and without the liquidity factor), even when a short time window is used. Alternative methodologies, such as the finite-$N$ approach of Gibbons et al. (1989a) and the more recent test of Gungor and Luger (2016), seem to have substantially lower power in detecting model misspecification. Finally, our results indicate that five prominent firm characteristics (book-to-market ratio, asset growth, operating profitability, market capitalization, and six-month momentum) are important determinants of the cross-section of expected returns of individual assets. Although the characteristic premia estimates are not always found to be statistically significant, it seems that these characteristics jointly explain a fraction of the overall cross-sectional dispersion in expected returns that is about 30 times larger than the fraction explained by the estimated factors' betas, regardless of the beta-pricing model under consideration.

Our paper is related to a large number of studies in empirical asset pricing and financial econometrics. The traditional two-pass cross-sectional regression (CSR) methodology for estimating beta-pricing models, developed by Black (1972) and Fama and Macbeth (1973), is valid when $T$ is large and $N$ is fixed. Shanken (1992) shows how the asymptotic standard errors of the second-pass CSR risk premia estimators are affected by the estimation error in the first-pass betas and provides standard errors that are robust to the errors-in-

---

6Since our test is specifically designed for scenarios in which $N$ is large, it alleviates the concerns of Lewellen et al. (2010), Harvey et al. (2016), and Barillas and Shanken (2017) about a particular choice of test assets in the econometric analysis.
variables (EIV) problem. Shanken and Zhou (2007) derive the large-$T$ properties of the two-pass estimator in the presence of global model misspecification. A different form of misspecification, not explored in this paper, can also occur when some of the factors have zero, or almost zero, betas, a situation that is referred to as the spurious or “useless” factors problem. Lack of identification of the risk premia also arises when at least one of the betas is cross-sectionally quasi-constant, as documented by Ahn et al. (2013) with respect to the market factor empirical betas, a case also ruled out here.

Building on Litzenberger and Ramaswamy (1979), Shanken (1992) (Section 6) proposes a large-$N$ estimator of the ex post risk premium and shows that it is asymptotically unbiased when $N$ diverges and $T$ is fixed. However, Shanken (1992) does not prove the consistency and asymptotic normality of this risk premium estimator. Differently from Litzenberger and Ramaswamy (1979), Shanken (1992) demonstrates unbiasedness without imposing a rigid structure on the covariance matrix of the first-pass residuals.

Following these seminal contributions, other methods have been recently proposed to take advantage of the increasing availability of large cross-sections of individual securities. Our paper is close to Gagliardini et al. (2016) in the sense that both studies provide inferential methods for estimating and testing beta-pricing models. However, their work is developed in a joint-asymptotics setting, where both $T$ and $N$ need to diverge. Moreover, they focus on a slightly different parameter of interest (obtained as the difference between the ex ante risk premia and the factors’ population mean), which can be derived from the ex post risk premium by netting out the sample mean of the factor. Like us, Gagliardini et al. (2016) need a bias adjustment because in their setting $N$ is diverging at a much faster rate than $T$. Moreover, while Gagliardini et al. (2016) assume random betas, as a consequence of their sampling framework with a continuum of assets, in our analysis we prefer to keep the betas nonrandom. This is for us mostly a convenience assumption since we show in the Appendix 1.D that allowing for randomness of the betas in a large-$N$ environment leaves our theoretical results unchanged. Gagliardini et al. (2016) characterize the time variation in risk premia by conditioning on observed state variables, whereas we leave the form of time variation unspecified. Like us, they show how to carry out inference when the beta-pricing model is globally misspecified. Finally, Gagliardini et al. (2016) allow for a substantial degree of cross-sectional dependence of the returns’ residuals. Although our setup and assumptions differ from theirs (mainly because in our framework only $N$ diverges), we also allow for a similar form of cross-sectional dependence in the residuals’ covariance matrix.

Bai and Zhou (2015) investigate the joint asymptotics of the unmodified OLS and GLS CSR estimators of the ex ante risk premia. Although the CSR estimators are asymptotically unbiased when $T$ diverges, they propose an adjustment to mitigate the finite-sample bias. Their bias adjustment differs from the one suggested by Litzenberger and Ramaswamy (1979) and Shanken (1992), and studied in this paper, because it relies on a large $T$ for its validity. However, their simulation results suggest that their bias-adjusted estimator performs well for various values of $N$ and $T$. Moreover, since $T$ must be large in their setting, Bai and Zhou (2015) bias-adjustment is asymptotically negligible, implying that the asymptotic distribution of their CSR estimators is identical to the asymptotic distribution of the traditional OLS and GLS CSR estimators. In contrast, we show that the asymptotic distribution of the risk premia estimator must necessarily change in the fixed-$T$ case, where the traditional trade-off between bias and variance emerges. Moreover, consistent estimation of the asymptotic covariance matrix of our risk premia estimator requires a different analysis.

---


8 See also Hou and Kimmel (2006) and Kan et al. (2013).

9 Several methods have been developed to deal with this particular form of model misspecification. See, for example, Jagannathan and Wang (1998), Fan and Zhang (1999a), Fan and Zhang (1999b), Kleibergen (2009), Ahn et al. (2013), Gospodinov et al. (2014), Burnside (2015), Bryzgalova (2016a), Gospodinov et al. (2017), Ahn et al. (2018), Gospodinov et al. (2018), Kleibergen and Zhan (2018a), and Kleibergen and Zhan (2018b), among others.

10 In the same paper, Shanken (1992) provides the well-known standard errors correction for ordinary least squares (OLS) and generalized least squares (GLS) estimators of the ex post risk premia, but his correction is only valid when $T$ is large and $N$ is fixed. (See his Section 3.2.)

11 In contrast, recall that in the traditional analysis of the CSR estimator (where $T$ diverges and $N$ is fixed), no bias adjustment is required.

12 Gagliardini et al. (2016) show that the bias adjustment in their framework is not asymptotically negligible when $N$ diverges at a much faster rate than $T$, a case not explicitly studied in Bai and Zhou (2015).
because only $N$ is allowed to diverge. Bai and Zhou (2015) focus exclusively on the case of a balanced panel under the assumption of correctly specified models. Unlike us, they do not account for time variation in the risk premia and do not analyze model misspecification.

Giglio and Xiu (2017) propose a modification of the two-pass methodology based on principal components that is robust to omitted priced factors and mis-measured observed factors, and establish its validity under joint asymptotics.

Kim and Skoulikis (2018) employ the so-called regression calibration approach used in EIV models to derive a $\sqrt{N}$-consistent estimator of the ex post risk premia in a two-pass CSR setting. Finally, Jegadeesh et al. (2018) propose instrumental-variable estimators of the ex post risk premia, exploiting the assumed independence over time of the return data.

As for specification testing, Pesaran and Yamagata (2012) extend the classical test of Gibbons et al. (1989a) to a large-$N$ setting. Besides accommodating only traded factors, the feasible version of their tests requires joint asymptotics and $N$ needs to diverge at a faster rate than $T$. Gungor and Luger (2016) propose a nonparametric testing procedure for mean-variance efficiency and spanning hypotheses (with tests of the beta-pricing restriction as a special case), and they derive (exact) bounds on the null distribution of the test statistics using resampling techniques. Their procedure, which is designed for traded factors only, is valid for any $N$ and $T$, even though they show that the power of their test increases when both $N$ and $T$ diverge. Gagliardini et al. (2016) derive the asymptotic distribution of their specification test under joint asymptotics and, like us, they allow for general factors. Finally, Gagliardini et al. (2018) propose a diagnostic criterion for detecting the number of omitted factors from a given beta-pricing model and establish its statistical behavior under joint asymptotics.

Having detailed our contributions and related them to the existing literature, we now discuss when our methodology should be used, from three different angles. With respect to the sampling scheme, our methodology is theoretically justified when $T$ is fixed and $N$ diverges. In contrast, the limiting results for the traditional CSR estimators cited above are valid when $T$ diverges with a fixed $N$ as well as when both $T$ and $N$ diverge. Proposition 1.3 in the paper warns us about using these traditional methods under our reference sampling scheme. Moreover, based on numerous Monte Carlo experiments, previous studies have found that the large-$T$ approximations of the CSR estimators are reliable only when five or more decades of data are used. (See Chen and Kan (2004) and Shankey and Zhou (2007), among others.) Therefore, our methodology could be useful also in scenarios where the time-series dimension is relatively large.

Starting from traded factors and assuming that the true risk premia are constant and the model is correctly specified, the sample means of the factors' excess returns or return spreads could be used as risk premia estimators of the true factors' means. However, a sufficiently large $T$ is required for the sample means to converge to their population counterparts. For non-traded factors, for example, macroeconomic variables, a panel of test asset returns is required to pin down the factors' risk premia, as the time series of the factors do not suffice. Mimicking portfolio excess returns could also be used in place of the non-traded factors, with the population means of the mimicking portfolio excess returns serving as the true risk premia. However, the mimicking portfolio projection requires $N < T$, which is violated under our reference sampling scheme.

---

13Building on Jagannathan et al. (2010), the Kim and Skoulikis (2018) estimator can be seen as an alternative to the Shranken estimator, the only difference being that in Kim and Skoulikis (2018) the first- and second-pass regressions are evaluated on non-overlapping time periods.

14Besides the classical econometric challenges associated with the choice of potentially weak instruments, these instrumental-variable approaches require a relatively larger $T$ in order to achieve the same statistical accuracy of the Shranken (1992) estimator. Moreover, the construction of the instruments in Jegadeesh et al. (2018) hinges upon the assumption of stochastic independence over time of the return data. The same assumption is also required in Kim and Skoulikis (2018). In contrast, it can be shown that the Shranken (1992) estimator retains its asymptotic properties even when the data is not independent over time. In fact, an arbitrary degree of serial dependence of the return data can be allowed for.

15See Breeden et al. (1989), Chan et al. (1998), and Lamont (2001), among others, for empirical studies based on the mimicking portfolio methodology. Baldassari and Robotti (2008) demonstrate by means of Monte Carlo simulations the greater accuracy of the mimicking portfolio risk premia estimates relative to the CSR risk premia estimates associated with the corresponding non-traded factors.

16When $N > T$, one could obtain the first $\tilde{N}$ principal components from a large panel of test assets returns, and then construct the mimicking portfolio for the non-traded factor using these $\tilde{N}$ assets (assuming that $\tilde{N} < T < N$). Although this
Finally, when the risk premia are time-varying, the argument for using our methodology appears even more compelling. Note that the considerations above regarding alternative estimation procedures for the traded factors case hold for both constant and time-varying risk premia. In particular, the (rolling) sample mean of the excess return on the traded factor (or of the return spread) will capture, in general, the average, over \( T \) observations, of the true time-varying risk premium associated with the factor. Alternatively, one can adopt the sampling scheme typical of nonparametric methods, with the implication that now the (rolling) sample mean will capture the time-varying risk premium and not just its average. However, a very large \( T \) would be necessary to obtain accurate estimates and a certain degree of smoothness, over time, of the true time-varying risk premium would be required. (See the Appendix 1.E for further details.) Our method for time-varying risk premia works for any \( T \) and makes no smoothness assumption.

To summarize, compelling reasons for using our methodology arise when \( T \) is fairly small (and, in particular, smaller than \( N \)). When considering models with non-traded factors, and when interest lies in the time variation in risk premia on traded and non-traded factors. In addition, our methodology can handle potential model misspecification (due, for example, to omitted pervasive factors) and, in particular, it provides a natural framework to determine whether the rejection of the beta-pricing relation is due to priced firm characteristics. Finally, we can easily accommodate unbalanced panels in the analysis.

The rest of the paper is organized as follows. Section 1.2 surveys the two-pass OLS CSR methodology, introduces our main assumptions, and sets the notation. Section 1.3 presents the asymptotic results for constant and time-varying risk premia estimates under correctly specified models. Section 1.4 generalizes our theory to potentially misspecified beta-pricing models with and without firm characteristics. Section 1.5 is for the unbalanced panel case. In Section 1.6, we present Monte Carlo simulation results. In Section 1.7, we investigate the empirical performance of FF5. Section 1.8 concludes. The technical proofs and additional material are in the Appendices 1.A–1.E.

1.2. The Two-Pass Methodology

This section introduces the notation and summarizes the two-pass OLS CSR methodology. We assume that the asset returns \( R_t = [R_{1t}, \ldots, R_{Nt}]' \) are governed by the following beta-pricing model:

\[
R_{it} = \alpha_i + \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \epsilon_{it} = \alpha_i + \beta_i' f_t + \epsilon_{it},
\]

where \( i \) denotes the \( i \)-th asset, with \( i = 1, \ldots, N \), \( t \) refers to time, with \( t = 1, \ldots, T \), \( \alpha_i \) is a scalar parameter representing the asset specific intercept, \( \beta_i = [\beta_{i1}, \ldots, \beta_{iK}]' \) is a vector of multiple regression betas of asset \( i \) with respect to the \( K \) factors \( f_t = [f_{1t}, \ldots, f_{Kt}]' \), and \( \epsilon_{it} \) is the \( i \)-th return's idiosyncratic component. In matrix notation, we can write the model above as

\[
R_t = \alpha + B f_t + \epsilon_t, \quad t = 1, \ldots, T,
\]

where \( \alpha = [\alpha_1, \ldots, \alpha_N]' \), \( B = [\beta_{11}, \ldots, \beta_{NN}]' \), and \( \epsilon_t = [\epsilon_{1t}, \ldots, \epsilon_{Nt}]' \). Let \( \Gamma = [\gamma_0, \gamma_1]' \), where \( \gamma_0 \) the zero-beta rate and \( \gamma_1 \) is the \( K \)-vector of ex ante factor risk premia, and denote by \( X = [1_N, B] \) the beta matrix augmented with \( 1_N \), an \( N \)-vector of ones. The following assumption of exact pricing is used at various points in the analysis below.

**Assumption 1.1**

\[
E[R_t] = X \Gamma.
\]

Eq. (1.3) follows, for example, from no-arbitrage (see Condition A in Chamberlain (1983a)) and a well-diversified mean-variance frontier (Definition 4 in Chamberlain (1983a)).\(^{17}\)

---

\(^{17}\) It should be noted that the mere absence of arbitrage is not sufficient for exact pricing, that is, nonzero pricing errors can coexist with no-arbitrage, as in the case of the APT of Ross (1976).
Averaging Eq. (1.2) over time, where we set \( \tilde{R} = \frac{1}{T} \sum_{t=1}^{T} R_t = [\tilde{R}_1, \ldots, \tilde{R}_N]' \), \( \tilde{\epsilon} = \frac{1}{T} \sum_{t=1}^{T} \epsilon_t \), and \( \tilde{f} = [\tilde{f}_1, \ldots, \tilde{f}_K]' = \frac{1}{T} \sum_{t=1}^{T} f_t \), imposing Assumption 1.1, and noting that \( E[R_t] = \alpha + BE[f_t] \) from Eq. (1.2), yields
\[
\tilde{R} = X\Gamma' + \tilde{\epsilon},
\]
where \( \Gamma' = [\gamma_0, \gamma_1']' \), and
\[
\gamma_1' = \gamma_1 + \tilde{f} - E[f_t].
\]
From Eq. (1.4), average returns are linear in the asset betas \textit{conditional} on the factor outcomes through the quantity \( \gamma_1' \), which, in turn, depends on the factors' sample mean innovations, \( \tilde{f} - E[f_t] \). The random coefficient vector \( \gamma_1' \) in Eq. (1.5) is referred to as the vector of ex post risk premia.\(^{18}\)

Eq. (1.5) shows that \( \Gamma \) and \( \Gamma' \) will coincide when \( \tilde{f} = E[f_t] \), which happens for \( T \to \infty \). When \( T \) is small, ex ante and ex post risk premia can differ substantially, as emphasized in the empirical section of the paper, although \( \gamma_1' \) remains an unbiased measure for the ex ante risk premia, \( \gamma_1 \).\(^{19}\)

Note that Eq. (1.4) cannot be used to estimate the ex post risk premia \( \Gamma' \) since \( X \) is not observed. For this reason, the popular two-pass OLS CSR method first obtains estimates of the betas by running the following multivariate regression for every \( i \):
\[
R_i = \alpha_i 1_T + F\beta_i + \epsilon_i,
\]
where \( R_i = [R_{i,1}, \ldots, R_{i,T}]' \), \( \epsilon_i = [\epsilon_{i,1}, \ldots, \epsilon_{i,T}]' \), \( F = [f_{i,1}, \ldots, f_{i,T}]' \) is the \( T \times K \) matrix of factors, and \( 1_T \) is a \( T \)-vector of ones. Then, the OLS estimates of \( B \) are given by
\[
\hat{B} = R'\hat{F}(\hat{F}'\hat{F})^{-1} = B + \epsilon'P,
\]
where \( \hat{B} = [\hat{\beta}_1, \ldots, \hat{\beta}_N]' \), \( R = [R_{1,1}, \ldots, R_{N,N}] \), \( \epsilon = [\epsilon_1, \ldots, \epsilon_N] \), and \( P = \hat{F}(\hat{F}'\hat{F})^{-1} \) with \( \hat{F} = [\hat{f}_1, \ldots, \hat{f}_T]' = (I_T - \frac{1}{T} 1_T1_T')F = F - 1_T1_T' \), where \( I_T \) is the identity matrix of order \( T \). The corresponding matrix of OLS residuals is given by \( \hat{\epsilon} = [\hat{\epsilon}_1, \ldots, \hat{\epsilon}_N] = R - 1_T\hat{R}' - \hat{F}\hat{B}' \).

We then run a single CSR of the sample mean vector \( \hat{R} \) on \( \hat{X} = [1_N, \hat{B}] \) to estimate the risk premia. Note that we have two alternative feasible representations of Eq. (1.4), that is,
\[
\tilde{R} = \hat{X}\Gamma + \eta,
\]
with residuals \( \eta = [\tilde{\epsilon} + B(\tilde{f} - E[f_t]) - (\hat{X} - X)\Gamma]' \), and
\[
\tilde{R} = \hat{X}\Gamma' + \eta',
\]
with residuals \( \eta' = [\tilde{\epsilon} - (\hat{X} - X)\Gamma']' \). The OLS CSR estimator applied to either Eq. (1.8) or Eq. (1.9) yields
\[
\hat{\Gamma} = \begin{bmatrix}
\hat{\gamma}_0 \\
\hat{\gamma}_1
\end{bmatrix} = (\hat{X}'\hat{X})^{-1}\hat{X}'\tilde{R}.
\]
However, when \( T \) is fixed, \( \hat{\Gamma} \) cannot be used as a consistent estimator of the ex ante risk premia, \( \Gamma \), in Eq. (1.8) and of the ex post risk premia, \( \Gamma' \), in Eq. (1.9). The reason is that neither \( \hat{B} \) converges to \( B \), nor \( \tilde{f} \) converges to \( E[f_t] \) unless \( T \to \infty \). Focusing on the representation in Eq. (1.9), the OLS CSR estimator can be corrected as follows. Denote by \( \text{tr}(\cdot) \) the trace operator and by \( 0_K \) a \( K \)-vector of zeros. In addition, let
\[
\hat{\sigma}^2 = \frac{1}{N(T-K-1)}\text{tr}(\hat{\epsilon}'\hat{\epsilon}).
\]
The bias-adjusted estimator of Shanken (1992) is then given by
\[
\hat{\Gamma}^* = \begin{bmatrix} \hat{\gamma}_0^* \\ \hat{\gamma}_1^* \end{bmatrix} = (\hat{\Sigma}_X - \hat{\Lambda})^{-1} \frac{\hat{X}'\hat{R}}{N}, \tag{1.12}
\]
where
\[
\hat{\Sigma}_X = \frac{\hat{X}'\hat{X}}{N} \quad \text{and} \quad \hat{\Lambda} = \begin{bmatrix} 0 \\ 0_K \end{bmatrix} \quad \text{and} \quad \hat{\alpha}^2 (\hat{F}'\hat{F})^{-1}.
\tag{1.13}
\]

The formula for the estimator \(\hat{\Gamma}^*\) exhibits a multiplicative bias adjustment through the term \((\hat{\Sigma}_X - \hat{\Lambda})^{-1}\).\(^{20}\)

This prompts us to explore the analogies of \(\hat{\Gamma}^*\) with the more conventional class of additive bias-adjusted OLS CSR estimators. To this end, it is useful to consider the following expression for the OLS CSR estimator, \(\hat{\Gamma}\), obtained from Bai and Zhou (2015) in their Theorem 1:
\[
\hat{\Gamma} = \Gamma + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \begin{bmatrix} 0 \\ 0_K \end{bmatrix} \delta^2 (\hat{F}'\hat{F})^{-1} \Gamma + O_p \left(\frac{1}{\sqrt{N}}\right)
\]
\[
= \Gamma + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \hat{\Lambda} \Gamma + O_p \left(\frac{1}{\sqrt{N}}\right) \Gamma^p. \tag{1.14}
\]

This formula suggests a simple way to construct an additive bias-adjusted estimator of \(\Gamma^p\); that is,
\[
\hat{\Gamma}^\text{bias-adj} = \hat{\Gamma} + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \hat{\Lambda} \hat{\Gamma}^\text{prelim}, \tag{1.15}
\]
where \(\hat{\Gamma}^\text{prelim}\) is an arbitrary preliminary estimator of \(\Gamma^p\).\(^{21}\) The next proposition shows that, by imposing that the preliminary estimator, \(\hat{\Gamma}^\text{prelim}\), and the bias-adjusted estimator, \(\hat{\Gamma}^\text{bias-adj}\), coincide, the unique solution to Eq. (1.15) is the Shanken (1992) estimator \(\hat{\Gamma}^*\) in Eq. (1.12).

**Proposition 1.1** Assume that \(\hat{\Sigma}_X - \hat{\Lambda}\) is nonsingular. Then, the Shanken (1992) estimator \(\hat{\Gamma}^*\) in Eq. (1.12) is the unique solution to the linear system of equations:
\[
\hat{\Gamma}^* = \hat{\Gamma} + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \hat{\Lambda} \hat{\Gamma}^*. \tag{1.16}
\]

**Proof:** See Appendix B.

Therefore, \(\hat{\Gamma}^*\) is the unique additive bias-adjusted OLS CSR estimator that does not require the preliminary estimation of the risk premia. As a computational precaution, it is possible that the EIV correction in Eq. (1.12) overestimates, making the matrix \((\hat{\Sigma}_X - \hat{\Lambda})\) almost singular for a given \(N\) and potentially leading to extreme values for the estimator. To alleviate this risk, our suggestion is to multiply the matrix \(\hat{\Lambda}\) by a scalar \(k (0 \leq k \leq 1)\) and to substitute \((\hat{\Sigma}_X - k\hat{\Lambda})^{-1}\) with \((\hat{\Sigma}_X - k\hat{\Lambda})^{-1}\) in Eq. (1.12), effectively yielding a shrinkage estimator.\(^{22}\) If \(k\) is zero, we obtain the OLS CSR estimator \(\hat{\Gamma}\), whereas if \(k\) is one, we obtain

\(^{20}\)Eq. (15) in Shanken (1992) differs slightly from our Eq. (1.12). The reason is that we do not impose the traded-factor restriction of Shanken (1992) in our setting.

\(^{21}\)For example, Bai and Zhou (2015) propose using the OLS CSR \(\hat{\Gamma}\) itself as the preliminary estimator, plugging it into the formula above in place of \(\hat{\Gamma}^\text{prelim}\). However, this adjustment is justified only when \(T \to \infty\). In general, the use of a preliminary estimator would decrease the precision of the bias-adjusted estimator and, in addition, it would make its properties harder to study.

\(^{22}\)Our asymptotic theory would require \(k = 1\) to converge to unity at a suitably slow rate as \(N\) increases. We omit the details to simplify the exposition.
the Shanken (1992) estimator $\hat{\Gamma}$. In our simulation experiments, we find that this shrinkage estimator is virtually unbiased, leading to $k = 1$. In contrast, in our empirical application in Section 1.7, shrinking is applied to roughly 75% of the cases (the average $k$ is 0.58) when $T = 36$ and to 5% of the cases (the average $k$ is 0.71) when $T = 120$. Our shrinkage adjustment can also alleviate the documented evidence of cross-sectional quasi-homogeneity for the loadings associated with certain risk factors, in particular for the market factor (see Ahn et al. (2013)).

Before turning to the challenging task of deriving the large-$N$ distribution of the Shanken (1992) estimator (and the associated standard errors), we discuss the perils of using the traditional $t$-ratio (specifically designed for a large-$T$ environment) when $N$ diverges. We first introduce the necessary assumptions and then present our results in Proposition 1.3 below.

**Assumption 1.2** As $N \to \infty$, 

$$
\frac{1}{N} \sum_{i=1}^{N} \beta_i \to \mu_\beta \quad \text{and} \quad \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' \to \Sigma_\beta, \quad (1.17)
$$

such that the matrix 

$$
\begin{bmatrix}
\frac{1}{N} \\
\mu_\beta \\
\Sigma_\beta
\end{bmatrix}
$$

is positive-definite. \hspace{1cm} (1.18)

Assumption 1.2 states that the limiting cross-sectional averages of the betas, and of the squared betas, exist. The second part of Assumption 1.2 rules out the possibility of spurious factors and situations in which at least one of the elements of $\beta_i$ is cross-sectionally constant. (See Ahn et al. (2013).) It implies that $X$ has full (column) rank for $N$ sufficiently large. To simplify the exposition, we assume that the $\beta_i$ are nonrandom.

**Assumption 1.3** The vector $\epsilon_t$ is independently and identically distributed (i.i.d.) over time with 

$$
E[\epsilon_t | F] = 0_N \quad (1.19)
$$

and a positive-definite matrix, 

$$
\text{Var}[\epsilon_t | F] = \begin{pmatrix}
\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\
\sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{N1} & \sigma_{N2} & \cdots & \sigma_N^2
\end{pmatrix} = \Sigma, \quad (1.20)
$$

where $0_N$ is a $N$-vector of zeros, and $\sigma_{ij}$ denotes the $(i,j)$-th element of $\Sigma$, for every $i, j = 1, \ldots, N$ with $\sigma_i^2 = \sigma_{ii}$.

---

23 The choice of the shrinkage parameter $k$ can be based on the eigenvalues of the matrix $\left( \hat{\Sigma}_X - k\hat{\Lambda} \right)$ as follows. Starting from $k = 1$, if the minimum eigenvalue of this matrix is negative and/or the condition number of this matrix is larger than 20 (as suggested by Greene (2003), p. 60), then we lower $k$ by an arbitrarily small amount. In our empirical application we set this amount equal to 0.05 and perform shrinkage whenever the absolute value of the relative change between the Shanken (1992) and the OLS CSR estimators is greater than 100%. We iterate this procedure until the minimum eigenvalue is positive and the condition number becomes less than 20. Gagliardini et al. (2016) rely on similar methods to implement their trimming conditions. Alternatively, one could use cross-validation to set the value of $k$.

24 Ahn et al. (2013) propose the so-called invariance bias (IB) coefficient as a measure of cross-sectional homogeneity. Applying their measure to our data on FF5, we find that the IB coefficient corresponding to the market factor equals 0.74 and 0.81 for rolling samples of size $T = 36$ and $T = 120$, respectively (averages across rolling samples). The IB coefficient is equal to 0.93 when considering the whole sample. According to Ahn et al. (2013), these values signal a very moderate risk of multicollinearity due to cross-sectional homogeneity. Similar values of the IB coefficient associated with the loadings on the market factor are obtained when estimating CAPM and FF3.

25 See Gagliardini et al. (2016) for a treatment of the beta-pricing model with random betas. In Appendix 1.D, we discuss the consequences of relaxing the nonrandomness of the $\beta_i$. 
The i.i.d. assumption over time is common to many studies, including Shanken (1992). However, our large \( N \) asymptotic theory, in principle, permits the \( \epsilon_{it} \) to be arbitrarily correlated over time, but the expressions would be more complicated. Conditions (1.19) and (1.20) are verified if the factors \( f_t \) and the innovations \( \epsilon_s \) are mutually independent for any \( s, t \). Noticeably, Condition (1.20) is not imposing any specific structure on the elements of \( \Sigma \). In particular, we are not assuming that the returns' innovations are uncorrelated across assets or exhibit the same variance. However, our large-\( N \) asymptotic theory needs to discipline the degree of cross-correlation among the residuals, although still allowing for a substantial degree of heterogeneity in the cross-section of asset returns. (See Assumption 1.5 below.)

As for the factors, we impose minimal assumptions because our asymptotic analysis holds conditional on the factors' realizations.

**Assumption 1.4** \( E[|f_t|] \) does not vary over time. Moreover, \( \tilde{F}'\tilde{F} \) is a positive-definite matrix for every \( T \geq K \).

**Assumption 1.5** As \( N \to \infty \),

(i)

\[
\frac{1}{N} \sum_{i=1}^{N} (\sigma_i^2 - \sigma^2) = o \left( \frac{1}{\sqrt{N}} \right),
\]

for some \( 0 < \sigma^2 < \infty \).

(ii)

\[
\sum_{i,j=1}^{N} |\sigma_{ij}| \mathbb{I}_{(i \neq j)} = o(N),
\]

where \( \mathbb{I}_{(\cdot)} \) denotes the indicator function.

(iii)

\[
\frac{1}{N} \sum_{i=1}^{N} \mu_{4i} \to \mu_4,
\]

for some \( 0 < \mu_4 < \infty \) where \( \mu_{4i} = E[\epsilon_{it}^4] \).

(iv)

\[
\frac{1}{N} \sum_{i=1}^{N} \sigma_{it}^4 \to \sigma_4,
\]

for some \( 0 < \sigma_4 < \infty \).

(v)

\[
\sup_{i} \mu_{4i} \leq C < \infty,
\]

for a generic constant \( C \).

(vi)

\[
E[\epsilon_{it}^4] = 0.
\]

(vii)

\[
\frac{1}{N} \sum_{i=1}^{N} \kappa_{4,iii} \to \kappa_4,
\]

for some \( 0 \leq |\kappa_4| < \infty \), where \( \kappa_{4,iii} = \kappa_4(\epsilon_{it}, \epsilon_{it}, \epsilon_{it}, \epsilon_{it}) \) denotes the fourth-order cumulant of the residuals \( \{\epsilon_{it}, \epsilon_{it}, \epsilon_{it}, \epsilon_{it}\} \).
(iii) For every $3 \leq h \leq 8$, all the mixed cumulants of order $h$ satisfy

$$
\sup_n \sum_{i_2, \ldots, i_h = 1}^N |\kappa_{i_1 i_2 \ldots i_h}| = o(N),
$$

for at least one $i_j \ (2 \leq j \leq h)$ different from $i_1$.

Assumption 1.5 essentially describes the cross-sectional behavior of the model disturbances. In particular, Assumption 1.5(i) limits the cross-sectional heterogeneity of the return conditional variance. Assumption 1.5(ii) implies that the conditional correlation among asset returns is sufficiently weak. Assumptions 1.5(i) and 1.5(ii) allow for many forms of strong cross-sectional dependence, as emphasized by the following proposition, which considers the case in which the $\epsilon_{it}$ obey a factor structure.

**Proposition 1.2** Assume that

$$
\epsilon_{it} = \lambda_t u_t + \eta_{it},
$$

where

$$
\sum_{i=1}^N |\lambda_i| = O(N^\delta), \quad 0 \leq \delta < 1/2,
$$

and (without loss of generality) for some fixed $q < N$ and some constant $C$,

$$
\lambda_1 + \cdots + \lambda_q \sim C N^{\frac{\delta}{2}},
$$

with $u_t \ \text{i.i.d.} \ (0,1)$ and $\eta_{it} \ \text{i.i.d.} \ (0, \sigma_\eta^2)$ over time and across units, where the $u_t$ and the $\eta_{it}$ are mutually independent for every $i, s, t$. Then,

(i) Assumption 1.5(i) and 1.5(ii) are satisfied with $\sigma^2 = \sigma_\eta^2$,

(ii) The maximum eigenvalue of $\Sigma$ diverges as $N \to \infty$.\textsuperscript{26}

**Proof:** See Appendix B.

Note that the boundedness of the maximum eigenvalue is the most common assumption on the covariance matrix of the disturbances in beta-pricing models. (See, e.g., the generalization of the APT by Chamberlain (1983b).) Our assumptions are weaker than the ones for the APT because the maximum eigenvalue can now diverge. This implies that the row-column norm of $\Sigma$, $\sup_{1 \leq i \leq N} \sum_{j=1}^N |\sigma_{ij}|$, diverges.\textsuperscript{27} Eq. (1.29) is adopted in our Monte Carlo experiments. Other special cases nested by Assumption 1.5 for which the cross-covariances $\sigma_{ij}$ are nonzero are network and spatial measures of cross-dependence and a suitably modified version of the block-dependence structure of Gagliardini et al. (2016).\textsuperscript{28}

In Assumption 1.5(iii), we simply assume the existence of the limit of the conditional fourth-moment, averaged across assets. In Assumption 1.5(iv), the magnitude of $\sigma_T$ reflects the degree of cross-sectional heterogeneity of the conditional variance of the asset returns. Assumption 1.5(v) is a bounded fourth-moment condition uniform across assets, which implies that $\sup_{T} \sigma_{T}^2 \leq C < \infty$. Assumption 1.5(vi) is a convenient symmetry assumption, but it is not strictly necessary for our results. Without 1.5(vi) the asymptotic distribution would be more involved, due to the presence of terms such as the third moment of the disturbance (averaged across assets). Assumption 1.5(vii) allows for non-Gaussianity of the asset returns.

\textsuperscript{26} The maximum eigenvalue of $\Sigma$ is given by $\sup_{x \neq 0} \|x\| = \sqrt{\Sigma x x'}$.

\textsuperscript{27} Assumption 1.5 allows for the maximum eigenvalue of $\Sigma$ to diverge at rate $o\left(\sqrt{N}\right)$. (See the proof of Proposition 1.2 for details.) Gagliardini et al. (2016) can allow for a faster rate, $o(N)$, of divergence of the maximum eigenvalue of $\Sigma$ because both $T$ and $N$ diverge in their double-asymptotics setting.

\textsuperscript{28} Gagliardini et al. (2016) Assumption BD.2 on block sizes and block numbers requires that the largest block size shrinks with $N$ and that there are not too many large blocks; that is, the partition in independent blocks is sufficiently fine-grained asymptotically. They show formally that such block-dependence structure is compatible with the unboundedness of the maximum eigenvalue of $\Sigma$. 
when $|\kappa_4| > 0$. For example, this assumption is satisfied when the marginal distribution of asset returns is a Student $t$ with degrees of freedom greater than four. However, when estimating the asymptotic covariance matrix of the Shanken (1992) estimator, one needs to set $\kappa_4 = 0$ merely for identification purposes, as explained in Lemma 1.6 in Appendix A. However, higher-order cumulants are not constrained to be zero, implying that $\kappa_4 = 0$ is not equivalent to Gaussianity. We are now ready to state our Proposition 1.3.

**Proposition 1.3** Under Assumptions 1.1-1.5 and as $N \to \infty$, the Fama and MacBeth (1973) $t$-ratios for $\hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_K, \ldots, \hat{\gamma}_K]$ based on the correction of Shanken (1992) satisfy the following relations.

(i) For the ex ante risk premia $\Gamma = [\gamma_0, \gamma_1, \ldots, \gamma_K]$, we have

$$|t_{FM}(\hat{\gamma}_0)| = \frac{|\hat{\gamma}_0 - \gamma_0|}{SE_{FM}^{\hat{\gamma}_0}} \to_p \infty$$

and

$$|t_{FM}(\hat{\gamma}_k)| = \frac{|\hat{\gamma}_k - \gamma_k|}{SE_{FM}^{\hat{\gamma}_k}} \to_p \left| J_k - E[J_k] \right| \sigma_k / \sqrt{T} - \frac{t_{k,K}^2 A^{-1} C \gamma_0^2}{\sigma_k / \sqrt{T}}$$

for $k \geq 1$. \hspace{1cm} (1.33)

(ii) For the ex post risk premia $\Gamma^P = [\gamma_0, \gamma_1^P, \ldots, \gamma_K^P, \ldots, \gamma_K^P]$, we have

$$|t_{FM,P}(\gamma_0)| = \frac{|\hat{\gamma}_0 - \gamma_0|}{SE_{FM,P}^{\gamma_0}} \to_p \infty$$

and

$$|t_{FM,P}(\hat{\gamma}_k)| = \frac{|\hat{\gamma}_k - \gamma_k^P|}{SE_{FM,P}^{\gamma_k}} \to_p \infty \text{ for } k \geq 1,$$ \hspace{1cm} (1.35)

where $SE_{FM}^{\hat{\gamma}_0}$ and $SE_{FM,P}^{\gamma_0}$ are the Fama and MacBeth (1973) standard errors with the Shanken (1992) correction corresponding to the ex ante and ex post risk premia, respectively (see Appendix B for details), and where $J_k$ is $k$-th column of the identity matrix $I_K$, $\sigma_k^2$ is the $(k,k)$-th element of $\hat{F}F^T$, $A = \Sigma_{\beta} - \mu_\beta \mu_\beta^T + C$, and $C = \sigma^2 (\hat{F}F)^{-1}$.

**Proof:** See Appendix B.

In summary, Proposition 1.3 shows that a methodology designed for a fixed $N$ and a large $T$, such as the one based on the Fama and MacBeth (1973) standard errors with the Shanken’s correction, is likely to lead to severe over-rejections when $N$ is large, thus rendering the inference on the beta-pricing model invalid.\(^{29}\) Our Monte Carlo simulations corroborate this finding. Moreover, Proposition 1.3 shows that when $N$ and $T$ are large, there is no need to apply the correction of Shanken (1992) to the Fama and MacBeth (1973) standard errors.

### 1.3. Asymptotic Analysis under Correctly Specified Models

In this section, we establish the limiting distribution of the Shanken (1992) bias-adjusted estimator, $\hat{\Gamma}^*$, and explain how its asymptotic covariance matrix can be consistently estimated.

#### 1.3.1 Baseline case

Our baseline case assumes that the beta-pricing model is correctly specified, that the risk premia are constant, and that the panel is balanced. This corresponds to the setup of Shanken (1992).

\(^{29}\)In particular, the $t$-ratio of the OLS CSR estimator for a particular element of the ex ante risk premium vector, $\gamma_1$, equals the standardized sample mean of the associated factor plus a bias term. When $T$ is allowed to diverge, the convergence of this $t$-ratio to a standard normal is re-obtained, but, for any given $T$, the deviations from normality can be substantial.
Let \( \Sigma_X = \begin{bmatrix} 1 & \mu'_\beta \\ \Sigma_\beta & \Sigma_\beta \end{bmatrix} \), \( \sigma^2 = \lim \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \), \( U_c = \lim \frac{1}{N} \sum_{i,j=1}^{N} E \left[ \text{vec}(\epsilon_i \epsilon'_j - \sigma_i^2 I_T) \text{vec}(\epsilon_j \epsilon'_j - \sigma_j^2 I_T) \right] \), \( M = I_T - D(D'D)^{-1}D' \), where \( \mu_\beta, \Sigma_\beta \), and \( \sigma_\beta^2 \) are defined in our assumptions above, \( U_c \) is described in Appendix C, \( D = [I_T, F] \), \( Q = \frac{1}{T} - P \gamma'_T \), \( Z = (Q \otimes P') + \frac{\text{vec}(M)}{T} \gamma'_T P' P \), and \( \otimes \) and \( \text{vec}(\cdot) \) denote the Kronecker product operator and the vec operator, respectively.

We make the following further assumption to derive the large-\( N \) distribution of the Shanken (1992) estimator.

**Assumption 1.6** As \( N \to \infty \), we have

\( i \)

\( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i \xrightarrow{d} N(0_T, \sigma^2 I_T). \) \( (1.36) \)

\( ii \)

\( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon'_i - \sigma_i^2 I_T) \xrightarrow{d} N(0_{p^2}, U_c). \) \( (1.37) \)

\( iii \) For a generic \( T \)-vector \( C_T \),

\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( C_T \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \epsilon_i \xrightarrow{d} N(0_{K+1}, V_c), \]

where \( V_c = c \sigma^2 \Sigma_X \) and \( c = C_T^T C_T \). In particular, \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( C_T \otimes \beta_i \right) \epsilon_i \xrightarrow{d} N(0_K, V_c) \), where \( V_c = c \sigma^2 \Sigma_\beta \).

Primitive conditions for Assumption 1.6 can be derived but at the cost of raising the level of complexity of our proofs. For instance, when Eqs. (1.29)-(1.30) hold, then Eq. (1.36) follows by Theorem 2 of Kuersteiner and Prucha (2013) when the \( \eta_i \) satisfy their martingale difference assumptions. (See their Assumptions 1 and 2.) This result extends easily to Eqs. (1.37)-(1.38) under suitable additional assumptions. (Details are available upon request.) We are now ready to state our first theorem.

**Theorem 1.1** As \( N \to \infty \), we have

\( i \) Under Assumptions 1.1–1.5,

\[ \hat{\Gamma}^* - \Gamma^* = O_P \left( \frac{1}{\sqrt{N}} \right). \] \( (1.39) \)

\( ii \) Under Assumptions 1.1–1.6,

\[ \sqrt{N} \left( \hat{\Gamma}^* - \Gamma^* \right) \Rightarrow_d N \left( 0_{K+1}, V + \Sigma_X^{-1} W \Sigma_X^{-1} \right), \]

where

\[ V = \sigma^2 \left[ 1 + \gamma'_T \left( \frac{T'}{T} \right)^{-1} \gamma_T \right] \Sigma_X^{-1} \] \( (1.41) \)

and

\[ W = \begin{bmatrix} 0 & 0_K \\ 0_K & Z' U Z \end{bmatrix}. \] (1.42)
**Proof:** See Appendix B.

The expression in Eq. (1.40) is remarkably simple and has a neat interpretation. The first term of this asymptotic covariance, \( V \), accounts for the estimation error in the betas, and it is essentially identical to the large-\( T \) expression of the asymptotic covariance matrix associated with the OLS CSR estimator in Shanken (1992). (See his Theorem 1(ii).) The term \( \frac{\bar{\sigma}_V^2 \Sigma_{X}^{-1}}{T} \) in Eq. (1.41) is the classical OLS CSR covariance matrix, which one would obtain if the betas were observed. The term \( c = \gamma_1^{P'} \left( \hat{F}' \hat{F} / T \right)^{-1} \gamma_1^P \) is an asymptotic EIV adjustment, with \( c \bar{\sigma}_V^2 \Sigma_{X}^{-1} \) being the corresponding overall EIV contribution to the asymptotic covariance matrix. As Shanken (1992) points out, the EIV adjustment reflects the fact that the variability of the estimated betas is directly related to the residual variance, \( \sigma^2 \), and inversely related to the factors' variability, \( \left( \hat{F}' \hat{F} / T \right)^{-1} \). The last term of the asymptotic covariance, \( \Sigma_{X}^{-1} W \Sigma_{X}^{-1} \) in Eq. (1.40), arises because of the bias adjustment that characterizes \( \hat{\Gamma}^* \). The \( W \) matrix in Eq. (1.42) accounts for the cross-sectional variation in the residual variances of the asset returns through \( U \). This term will vanish when \( T \to \infty \). In Appendix C, we provide an explicit expression for \( U \), and we show that \( U \) only depends on the fourth-moment structure of the \( \epsilon_{it} \), that is, on \( \kappa_4 \) and \( \sigma_4 \).\(^{36}\) The \( \sqrt{N} \)-rate of convergence established in Theorem 1.1-(i) coincides with the rate of convergence established by Gagliardini et al. (2016) with respect to their \( \sqrt{N}T \)-consistent estimator of \( \nu = \hat{\gamma}_1^* - \bar{\gamma} \) when \( T \) is fixed.

To conduct statistical inference, we need a consistent estimator of the asymptotic covariance matrix, which we present in the next theorem. Let \( M^{(2)} = M \odot M \), where \( \odot \) denotes the Hadamard product operator. In addition, define

\[
\tilde{Z} = (Q \otimes P) + \frac{\text{vec}(M)}{T - K - 1} \hat{\gamma}_1^* P' P \quad \text{with} \quad \hat{Q} = \frac{1}{T} \bar{P} \hat{\gamma}_1^*.
\]

\[ (1.43) \]

**Theorem 1.2** Under Assumptions 1.1-1.5 and the identification condition \( \kappa_4 = 0 \), as \( N \to \infty \), we have

\[
\hat{V} + \left( \hat{\Sigma}_{X} - \bar{\lambda} \right)^{-1} \hat{W} \left( \hat{\Sigma}_{X} - \bar{\lambda} \right)^{-1} \to_p \nu + \Sigma_{X}^{-1} W \Sigma_{X}^{-1},
\]

\[ (1.44) \]

where

\[
\hat{V} = \frac{\hat{\sigma}_V^2}{T} \left[ 1 + \hat{\gamma}_1^{P'} \left( \frac{\hat{F}' \hat{F}}{T} \right)^{-1} \hat{\gamma}_1^* \right] (\hat{\Sigma}_{X} - \bar{\lambda})^{-1},
\]

\[ (1.45) \]

\[
\hat{W} = \begin{bmatrix} 0 & \hat{0}'_{K} \\ \hat{0}_{K} & \hat{Z} \hat{U}, \hat{Z} \end{bmatrix},
\]

\[ (1.46) \]

and \( \hat{U} \) is a consistent estimator of \( U \), (see Appendix C), obtained replacing \( \sigma_4 \) with

\[
\hat{\sigma}_4 = \frac{1}{N} \sum_{i=1}^{T} \sum_{j=1}^{N} \epsilon_{it}^2.
\]

\[ (1.47) \]

**Proof:** See Appendix B.

A remarkable feature of the result above is that a consistent estimate of the asymptotic covariance matrix of \( \hat{\Gamma}^* \) can be obtained while leaving the residual covariance matrix \( \Sigma \) unspecified. In fact, with \( \Sigma \) having in general \( N(N + 1)/2 \) distinct elements and our asymptotic theory being valid only for \( N \to \infty \), consistent estimation of \( \Sigma \) would be infeasible. A convenient feature of the Shanken (1992) estimator is that it depends on \( \Sigma \) only through the average of the \( \sigma_i^2 \). Moreover, its asymptotic covariance matrix depends on the limits of \( \sum_{i,j=1}^{N} \sigma_{ij} / N \) and \( \sum_{i=1}^{N} \sigma_{ii}^2 / N \). Our large \( N \) asymptotic theory shows how these quantities can be estimated consistently. In contrast, the individual covariances \( \sigma_{ij} \) cannot be consistently estimated due to the fixed \( T \).

\(^{36}\)See Assumption 1.5 for the definition of \( \kappa_4 \) (the cross-sectional average of the fourth-order cumulants of the \( \epsilon_{it} \)) and \( \sigma_4 \) (the cross-sectional average of the \( \sigma_i^2 \)).
1.3. ASYMPTOTIC ANALYSIS UNDER CORRECTLY SPECIFIED MODELS

The condition $\kappa_4 = 0$ is required as a consequence of the small-$T$ and large-$N$ framework.\(^{31}\) However, $\kappa_4 = 0$ is not as restrictive as it may seem. A sufficiently large level of heterogeneity in the $\sigma_t^2$ generates a substantial level of volatility in the conditional distribution of assets’ returns by inducing a mixture-distribution effect.\(^{32}\)

1.3.2 Time-varying case

In this section, we study the behavior of the estimator $\hat{\Gamma}^*$ when the risk premia are allowed to be time-varying, again under the assumption of correct model specification. It turns out that $\hat{\Gamma}^*$ is suitable for time-varying risk premia estimation because it estimates accurately local averages (over the, possibly very short, time window of size $T > K + 1$) of the true time-varying risk premia, regardless of their form and degree of time variation. Noticeably, we are also able to derive a consistent estimator of the true $t$-th period risk premia and to characterize its asymptotic distribution.\(^{33}\)

Throughout this section, we substitute Assumption 1.1 with

$$E_{t-1}[R_{it}] = \gamma_0,t-1 + \beta_t'\gamma_{1,t-1}, \quad (1.48)$$

where $E_{t-1}[\cdot]$ denotes the conditional expectation with respect to all the available information up to time $t-1$. Importantly, our theory does not need to restrict the type of time variation in $\Gamma_{t-1} = [\gamma_0,t-1, \gamma_{1,t-1}]'$. To simplify the treatment of time variation in the premia, without altering the estimation procedure developed in this paper, we maintain the $\beta_t$ in Eq. (1.48) constant over time.\(^{34}\) Our results below easily extend to the case of $\beta_{t-1} = B_t z_{t-1}$, for some (vector of) predetermined state variables $z_{t-1}$ and a suitable matrix of loadings $B_t$.

Under Eq. (1.48), asset returns are now given by $R_{it} = [1, \beta_t'\Gamma_{t-1}^P + \epsilon_{it}]$, where $\Gamma_{t-1}^P$ are the $(t-1)$-th ex post risk premia:

$$\Gamma_{t-1}^P = \Gamma_{t-1} + f_t - E_{t-1}[f_t], \text{ with a sample average } \Gamma^P = \frac{1}{T} \sum_{t=1}^{T} \Gamma_{t-1}^P. \quad (1.49)$$

By construction, the ex post time-varying risk premia $\Gamma_{t-1}^P$ have a conditional mean that equals $\Gamma_{t-1}$, the ex ante time-varying risk premia.

To estimate the $(t-1)$-th risk premia, for $t = 1, \ldots, T$, we introduce the following novel estimator:

$$\hat{\Gamma}_{t-1} = \begin{bmatrix} \hat{\gamma}_{0,t-1}^- \\ \hat{\gamma}_{1,t-1}^- \end{bmatrix} = \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \hat{X}'R_t - \hat{\sigma}^2 \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left( \hat{\Sigma}' \hat{\Gamma} \right)^{-1} \hat{X}'I_{t,T} \right), \quad (1.50)$$

where, as before, $I_{t,T}$ denotes the $t$-th column, for $t = 1, \ldots, T$, of the identity matrix $I_T$.\(^{35}\) The next theorem derives the large-$N$ behavior of both $\hat{\Gamma}^*$ and $\hat{\Gamma}_{t-1}$.\(^{36}\)

**Theorem 1.3** Under Eq. (1.48) and Assumptions 1.2-1.6, as $N \to \infty$, we have

\( (i) \) $\hat{\Gamma}^*$ and $\sqrt{N}(\hat{\Gamma}^* - \Gamma^P)$ satisfy Theorem 1 with $\Gamma^P$ replaced by $\Gamma^P$.\(^{37}\)

\(^{31}\)As we show in detail in Lemma 1.6 of Appendix A, the limit of $\hat{\sigma}_4$ in Eq. (1.47) converges to a linear combination of $k_4$ and $\sigma_4$. These two parameters could be identified and consistently estimated only under the stronger assumption of independence across assets, since, in this case, $\sigma_4$ would reduce to $\sigma^4$ (which could be easily estimated using the square of $\hat{\sigma}^2$). In contrast, allowing for some arbitrary degree of cross-correlation implies that $k_4$ and $\sigma_4$ cannot be separately identified. This is the reason for setting $k_4 = 0$.

\(^{32}\)In our empirical applications our estimated $\sigma_4$ is about 10 times the estimate for $\sigma^4$.

\(^{33}\)Our new estimator for the time-varying risk premia appears useful also for traded factors, and not just for non-traded factors, particularly within our fixed-$T$ environment (see the Appendix 1.6 for further details), especially when $T$ is assumed to be very small.

\(^{34}\)See, e.g., Ferson and Harvey (1991) who argue that the time variation in expected returns is mainly due to time variation in the premia as opposed to time variation in the betas.

\(^{35}\)Note that $\hat{\Gamma}_{t-1}$ is a new estimator that successfully tackles the problem of estimating time-varying risk premia in a large-$N$ setting. It should not be confused with the Shanken (1992) formula in his Theorem 5.
(ii) \( \hat{\Gamma}_{t-1} - \Gamma_{t-1}^P = O_p \left( \frac{1}{\sqrt{N}} \right) \) and

\[
\sqrt{N} \left( \hat{\Gamma}_{t-1} - \Gamma_{t-1}^P \right) \rightarrow_d N \left( 0_{K+1}, V_{t-1} + \Sigma X^{-1} W_{t-1} \Sigma X^{-1} \right),
\]

where \( V_{t-1} = \sigma^2 Q_{t-1} Q_{t-1}^{-1}, \) \( W_{t-1} = \begin{bmatrix} 0 & 0 \end{bmatrix} Z_{t-1}^{-1} U, Z_{t-1} \), \( Q_{t-1} = \mu_t - P \gamma_{t-1}^{P}, \) and \( Z_{t-1} = (Q_{t-1} \otimes P) - \frac{\text{vec} (M)}{T-1} Q_{t-1}^{P} \), with \( U_t \) as in Theorem 1.

**Proof:** See Appendix B.

Theorem 1.3 states that, when Eq. (1.48) holds, \( \hat{\Gamma}^* \) consistently estimates the local average of the ex post time-varying risk premia over \( T \) periods, the only requirement being that \( T > K + 1 \). If one is interested in the ex post risk premia for a specific time period, \( \Gamma_{t-1}^P \), then asymptotically correct inference can be carried out by using \( \hat{\Gamma}_{t-1} \). Interestingly, \( \hat{\Gamma}^* \) is numerically identical to the sample mean of \( \hat{\Gamma}_{t-1} \), over \( t = 1, \ldots, T \), because the additive bias adjustment, on the right-hand side of Eq. (1.50), vanishes due to the identity \( \sum_{t=1}^{T} \tilde{F}^t u_t = \tilde{F}^{1T} = 0 \).

To better understand the importance of our large-\( N \) results, it is useful to consider the behavior of the OLS CSR estimator \( \hat{\Gamma} \) when Eq. (1.48) holds. In this case, we have

\[
\hat{\Gamma} \to_p \Gamma_{\infty} \quad \text{as} \ T \to \infty,
\]

where \( \Gamma_{\infty} = \lim_{T \to \infty} \frac{1}{T} \int_0^T \Gamma_t ds \) denotes the integrated risk premia, namely the long-run average over the entire timeline.\(^{36}\) Next, consider \( \hat{\Gamma}_{t-1} = (\tilde{X}' \tilde{X})^{-1} \tilde{X}' R_t \), which can be thought of as the OLS CSR estimator for the \( (t-1) \)-th risk premia.\(^{37}\) It follows that

\[
\hat{\Gamma}_{t-1} \to_p \Gamma_{t-1}^P + \begin{pmatrix} N & 1' N B' \end{pmatrix}^{-1} \begin{pmatrix} 1' N \end{pmatrix} \varepsilon_t \quad \text{as} \ T \to \infty.
\]

Hence, the limit of \( \hat{\Gamma}_{t-1} \) is the sum of two components, that is, the \( (t-1) \)-th ex post risk premia \( \Gamma_{t-1}^P \) and a random term that is a function of \( \varepsilon_t \). This last term cannot be consistently estimated, thus making \( \hat{\Gamma}_{t-1} \) an unreliable estimator of both \( \Gamma_{t-1} \) and \( \Gamma_{t-1}^P \), even when \( T \to \infty \). In contrast, in our large-\( N \) framework, \( (\tilde{X}' \tilde{X} - \tilde{\Lambda})^{-1} \tilde{X}' R_t \to_p \Gamma_{t-1}^P + \sigma^2 \Sigma X^{-1} \begin{pmatrix} 0 \end{pmatrix} (\tilde{F}^{1T})^{-1} \tilde{F}^t u_t,T \) as \( N \to \infty \), where the bias term \( \sigma^2 \Sigma X^{-1} \begin{pmatrix} 0 \end{pmatrix} (\tilde{F}^{1T})^{-1} \tilde{F}^t u_t,T \) can now be consistently estimated, leading to the bias-adjusted estimator \( \hat{\Gamma}_{t-1} \) in Eq. (1.50). Finally, a consistent estimator of the asymptotic covariance matrix of \( \hat{\Gamma}_{t-1} \) in Eq. (1.51) can be easily obtained. (See Theorem 1.2 and its proof.)

### 1.4. Asymptotic Analysis under Potentially Misspecified Models

In this section, we explore the implications of model misspecification for model and parameter testing. Under the full rank assumption on the \( X \) matrix, the focus of the analysis is on the fixed (global) type of misspecification considered in Shanken and Zhou (2007) and several follow-up papers. A beta-pricing model is misspecified if there exists no value of the risk premia \( \Gamma \) for which the associated vector of pricing errors is zero. This misspecification might be due, for example, to the omission of some relevant risk factor, imperfect measurement of the factors, or failure to incorporate some relevant aspect of the economic environment

---

\(^{36}\)If one assumes, as in Ang and Kristensen (2012), that \( \Gamma_t = \Gamma(t/T) \), \( 1 \leq t \leq T \), for a smooth function \( \Gamma(\cdot) \), then the integrated risk premia \( \Gamma_{\infty} \) become \( \int_0^T \Gamma_t ds \).

\(^{37}\)The quantity \( \hat{\Gamma}_{t-1} \) is well-known in empirical finance because its sample variance is routinely used to compute the Fama and MacBeth (1973) standard errors of \( \hat{\Gamma} \).
- taxes, transaction costs, irrational investors, and the like. Thus, misspecification of some sort seems inevitable, given the inherent limitations of beta-pricing models.

This section is organized as follows. In Section 1.4.1, we propose a new specification test that is appropriately designed to detect model misspecification of unknown form. Section 1.4.2 deals with risk premia estimation and provides standard errors that are valid under potential model misspecification. Finally, Section 1.4.3 explores the situation in which the beta-pricing model is misspecified due to priced firm characteristics.

### 1.4.1 Testing for model misspecification

When a beta-pricing model is correctly specified (see Assumption 1),

\[ H_0 : e_i = 0 \quad \text{for every } i = 1, 2, \ldots, \]  

(1.54)

where \( e_i = E[R_t] - \gamma_0 - \beta_i' \gamma_1 \) is the population (ex ante) pricing error associated with asset \( i \). Denoting the vector of sample ex post pricing errors by

\[ \hat{\epsilon} = (\hat{\epsilon}_1^P, \ldots, \hat{\epsilon}_N^P)' = \hat{R} - \hat{X}\hat{\gamma}, \]

(1.55)

we have

\[ \hat{\epsilon}_i^P = \hat{R}_i - \hat{X}_i\hat{\gamma} = e_i + Q' \epsilon_i - \hat{X}_i \left( \hat{\gamma} - \Gamma P \right). \]

(1.56)

Theorem 1.1(i) implies that, for every \( i \),

\[ \hat{\epsilon}_i^P \rightarrow_p e_i + Q' \epsilon_i \equiv e_i^P. \]  

(1.57)

Eq. (1.57) shows that even when the ex ante pricing errors, \( e_i \), are zero, \( \hat{\epsilon}_i^P \) will not converge in probability to zero because \( T \) is fixed. Nonetheless, a test of \( H_0 \) with correct size and good power can be developed. Define the sum of the sample squared ex post pricing errors as

\[ \hat{Q} = \frac{1}{N} \sum_{i=1}^{N} (\hat{\epsilon}_i^P)^2. \]

(1.58)

Consider the centered statistic

\[ S = \sqrt{N} \left( \hat{Q} - \frac{\hat{Q}^2}{N} \right) \left( 1 + \frac{1}{N} \frac{(P'\hat{P}/T)^{-1} \hat{\gamma}^*}{(P'\hat{P}/T)^{-1} \hat{\gamma}^*} \right). \]

(1.59)

The centering is needed because of Eq. (1.57). To see this, from the population ex post pricing errors, \( e_i^P \), we have

\[ \frac{1}{N} \sum_{i=1}^{N} (\epsilon_i^P)^2 = \frac{1}{N} \sum_{i=1}^{N} e_i^2 + Q' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_i e_i \right) Q + o_p(1) = \frac{1}{N} \sum_{i=1}^{N} e_i^2 + \sigma^2 Q'Q + o_p(1). \]

(1.60)

Therefore, even under \( H_0 : e_i = 0 \) for all \( i \), the average of the population squared ex post pricing errors will not converge to zero but rather to \( \sigma^2 Q'Q = \sigma^2 (1 + \gamma_1^* (P'\hat{P}/T)^{-1} \hat{\gamma}^*) \). This is the quantity whose consistent estimate we need to demean our test statistic by in order to obtain its limiting distribution. The following theorem provides the limiting distribution of \( S \) under \( H_0 : e_i = 0 \) for every \( i \).

**Theorem 1.4** Under Eq. (1.54) and Assumptions 1.1-1.6, as \( N \rightarrow \infty \), we have

\[ S \rightarrow_d \mathcal{N}(0, \mathcal{V}), \]

where \( \mathcal{V} = \mathcal{V}_N \mathcal{Q} \) and \( \mathcal{Q} = (Q \otimes Q) - \frac{\text{vec}(\mathcal{M})}{\text{vec}(\mathcal{M})^T} Q'Q. \)

(1.61)
Proof: See Appendix B.

The asymptotic variance of the test in Eq. (1.61) can be consistently estimated by replacing $Q$ with $\hat{Q}$ and $U_i$ with $\hat{U}_i$. Specifically, using Theorem 1.2 and Lemma 1.6 in Appendix A, we have

$$\hat{V} = \hat{Z}_Q \hat{U}_i \hat{Z}_Q \rightarrow_p Z'_Q U_i Z_Q,$$

where

$$\hat{Z}_Q = \left( \hat{Q} \otimes \hat{Q} \right) - \frac{\text{vec}(M)}{T-K-1} \hat{Q}' \hat{Q}.$$  \hfill (1.63)

Then, under $H_0$, it follows that

$$S^* = \frac{S}{V^*_T} \rightarrow_d N(0,1).$$ \hfill (1.64)

It turns out that our test statistic $S^*$ has power when $c_i^2$ is greater than zero for the majority of the test assets.\(^{38}\) Moreover, it is straightforward to show that the distribution of our test under the null hypothesis is invariant to asset repackaging.

1.4.2 Estimation under potential model misspecification

If the null hypothesis of correct model specification, for the beta-pricing model under consideration, is rejected, one has two options. The first possibility is to conclude that the model is wrong, and to modify the model accordingly before proceeding with risk premia estimation. If one still wishes to conduct inference on risk premia with the same beta-pricing model, then the standard errors of the risk premia estimates need to be robustified against potential model misspecification. This is the approach we propose in this section.

Suppose that Assumption 1.1 is violated and assume that

$$E[R_t] = 1_N \bar{\gamma}_0 + B \bar{\gamma}_1 + e_t,$$ \hfill (1.65)

where, following Shanken and Zhou (2007), the (pseudo)-true values $\bar{\Gamma} = [\bar{\gamma}_0, \bar{\gamma}_1]'$ are given by

$$\bar{\Gamma} = \arg\min_C \frac{(E[R_t] - X \Gamma)'(E[R_t] - X \Gamma)}{N},$$ \hfill (1.66)

for an arbitrary $(K + 1)$-vector $C$.

When the model is correctly specified, $\bar{\Gamma} = \Gamma$, the vector of ex ante risk premia.\(^{39}\)

We now introduce an additional assumption that governs the behavior of the population pricing errors in terms of cross-sectional moments with the returns’ innovations.

**Assumption 1.7** As $N \to \infty$, we have

(i) \[ \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i' \rightarrow_p 0. \] \hfill (1.67)

(ii) \[ \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i' \rightarrow_p \tau_0 I_T. \] \hfill (1.68)

---

\(^{38}\)Specifically, our test will reject $H_0$ when the pricing errors $\epsilon_i$ are zero for only a number $N_0$ of assets, such that $N_0/N \to 0$ as $N \to \infty$. This condition allows $N_0$ to diverge, although not too fast. A formal power analysis can be developed by using the notion of local alternatives as in Gagliardini et al. (2016). In Section 1.6, we present a Monte Carlo simulation experiment calibrated to real data that demonstrates the desirable size and power properties of our test.

\(^{39}\)Under the i.i.d. normality assumption and Eq. (1.65), Shanken and Zhou (2007) establish the asymptotic distribution of the OLS and GLS CSR estimators of $\Gamma$ as $T \to \infty$. (See also Hou and Kimmel (2006).) Kan et al. (2013) generalize their results to the case of temporally dependent and nonnormal test asset returns and factors, and derive the large-$T$ distribution of the OLS and GLS CSR $R^2$. 
1.4. ASYMPTOTIC ANALYSIS UNDER POTENTIALLY MISSPECIFIED MODELS

(iii) \[
\frac{1}{N} \sum_{i=1}^{N} \epsilon_i c_i e_i \rightarrow_p \tau_\Phi I_T. \tag{1.69}
\]

(iv) \[
\sum_{i,j=1}^{N} \sigma_{ij} c_i e_j \mid 1_{(i \neq j)} = o(N), \tag{1.70}
\]

for some constants \(\tau_\Omega = \plim \frac{1}{N} \sum_{i=1}^{N} \epsilon_i^2 \) and \(\tau_\Phi = \plim \frac{1}{N} \sum_{i=1}^{N} \epsilon_i c_i.

Assumption 1.7(i) implies that the \(\epsilon_{it}\) and the pricing errors are cross-sectionally uncorrelated, although, by Assumption 1.7(ii) and 1.7(iii), they could be cross-sectionally dependent in terms of second moments of the \(\epsilon_{it}\). Assumption 1.7(iv) implies that the pricing errors are not altering the degree of cross-sectional dependence of the \(\epsilon_{it}\).

Let \(\bar{\Gamma}^P = \bar{\Gamma} + \bar{f} - E[f_t]\). The following theorem extends Theorems 1.1 and 1.2 to the case of globally misspecified beta-pricing models.

**Theorem 1.5** As \(N \to \infty\), we have

(i) Under Assumptions 1.2-1.5, Assumption 1.7, and Eq. (1.65),

\[
\bar{\Gamma}^* - \bar{\Gamma}^P = O_p \left( \frac{1}{\sqrt{N}} \right). \tag{1.71}
\]

(ii) Under Assumptions 1.2-1.7 and Eq. (1.65),

\[
\sqrt{N} \left( \bar{\Gamma}^* - \bar{\Gamma}^P \right) \rightarrow_d N \left( 0_{K+1}, V + \Sigma_X^{-1} (W + \Omega + \Phi + \Phi') \Sigma_X^{-1} \right), \tag{1.72}
\]

where \(V\) and \(W\) are defined in Theorem 1.1 by replacing \(\gamma_{it}^P\) with \(\hat{\gamma}_t^P\),

\[
\Omega = \begin{bmatrix} 0 & 0_{K} \\ 0_{K} & \tau_\Omega P'P \end{bmatrix} \quad \text{and} \quad \Phi = \begin{bmatrix} 0 & \tau_\Phi Q'P \\ 0_{K} & \tau_\Phi (Q' \otimes \mu_\beta) P \end{bmatrix}. \tag{1.73}
\]

(iii) Under Assumptions 1.2-1.5, Assumption 1.7, Eq. (1.65), and \(\kappa_4 = 0\),

\[
\hat{V} + \left( \Sigma_X - \hat{\Lambda} \right)^{-1} (\hat{W} + \hat{\Omega} + \hat{\Phi} + \hat{\Phi}') \left( \Sigma_X - \hat{\Lambda} \right)^{-1} \rightarrow_p V + \Sigma_X^{-1} (W + \Omega + \Phi + \Phi') \Sigma_X^{-1}, \tag{1.74}
\]

where \(\hat{V}\) and \(\hat{W}\) are defined in Theorem 1.2,

\[
\hat{\Omega} = \begin{bmatrix} 0 & 0_{K} \\ 0_{K} & \hat{\tau}_{\Omega} P'P \end{bmatrix} \quad \text{and} \quad \hat{\Phi} = \begin{bmatrix} 0 & \hat{\tau}_\Phi Q'P \\ 0_{K} & \hat{\tau}_\Phi (Q' \otimes \hat{\mu}_\beta) P \end{bmatrix}, \tag{1.75}
\]

and \(\hat{\tau}_\Phi\) and \(\hat{\tau}_{\Omega}\) are defined in Lemmas 1.8 and 1.9 in Appendix A, respectively.

**Proof:** See Appendix B.

Similar to the expressions in Shanken and Zhou (2007) and Kan et al. (2013), the asymptotic covariance of \(\bar{\Gamma}^*\) contains three additional terms, \(\Omega\), \(\Phi\), and \(\Phi'\). The contribution of the pricing errors to the overall asymptotic covariance increases when the variability of the residuals \(\epsilon_{it}\) increases or, alternatively, when the variability of the pricing errors \(c_i\) increases, leading to a larger \(\tau_{\Omega}\).
Notice that under model misspecification \( \hat{\Gamma} \) changes with \( N \) and, as a consequence, one can define the limit risk premia \( \Gamma_\infty = \lim_{N \to \infty} \hat{\Gamma} \). Theorem 3 of Ingersoll (1984) provides the conditions for the existence and the uniqueness of \( \hat{\Gamma}_\infty \).\(^{40}\) It follows that, by Theorem 1.5, \( \hat{\Gamma}^* \) also converges to \( \hat{\Gamma}_\infty^* = \left[ \hat{\gamma}_0^* \right] = \hat{\Gamma}_\infty + \tilde{f} - E[f_t] \). Moreover, if \( \hat{\Gamma}_\infty^* \) is \( o \left( 1/\sqrt{N} \right) \), then the asymptotic distribution of \( \hat{\Gamma}^* \) around \( \hat{\Gamma}_\infty^* \) is the same as the one in Eq. (1.72).\(^{41}\) Interestingly, even under model misspecification, there is no loss of speed of convergence. This differs from Gagliardini et al. (2016), who obtain a slower rate of convergence, \( O \left( \sqrt{N} \right) \) instead of \( O \left( \sqrt{N^T} \right) \), of their estimator to the true ex ante risk premia, \( \Gamma_\infty \), when the model is misspecified.

### 1.4.3 Misspecification due to priced characteristics

We follow Section 3.3 of Shanken (1992) and allow for Assumption 1.1 to be potentially violated because the cross-section of expected returns now satisfies

\[
E[R_{it}] = \gamma_0 + \gamma_1 \beta_i + \delta c_i,
\]

where \( c_i \) denotes a \( K \)-vector of time-invariant firm characteristics and \( \delta \) denotes the corresponding vector of characteristic premia. Our theory requires characteristics and loadings to be sufficiently heterogenous across assets although we allow them to be (almost) arbitrarily cross-sectionally correlated.\(^{42}\) Since characteristics exhibit only modest changes over short time windows, Eq. (1.76) would be a good approximation to the true data generating process also in a time-varying setting with a small \( T \).\(^{43}\)

Imposing Eq. (1.76), averaging (1.2) over time, and replacing \( X \) with \( \hat{X} \), we obtain

\[
\tilde{R} = \tilde{X} \Gamma^P + C \delta + \eta^P,
\]

where \( C = [c_1, \ldots, c_N]' \) and \( \eta^P = \left( \tilde{\epsilon} - (\tilde{X} - X) \Gamma^P \right) \). The estimates of \( \Gamma^P \) and \( \delta \) are given by

\[
\begin{bmatrix}
\hat{\Gamma}^* \\
\hat{\delta}^*
\end{bmatrix} = \begin{bmatrix}
\tilde{X}' \tilde{X} - N \hat{\Lambda} & \tilde{X}'C' \\
C' \tilde{X} & C'C
\end{bmatrix}^{-1} \begin{bmatrix}
\tilde{X}' \tilde{R} \\
C' \tilde{R}
\end{bmatrix},
\]

where \( \hat{\Lambda} \) is the bias adjustment from Theorem 1.1. In line with the discussion around Theorem 1.3, \( \hat{\Gamma}^* \) and \( \hat{\delta}^* \) will also estimate (consistently) the local averages of the risk and characteristic premia if these are allowed to be time-varying.

In this setting with characteristics, we need to make the following additional assumption. Let \( z_i = \varepsilon_i \otimes c_i \) and \( \Sigma_{z,z} = \text{Cov}(z_i, z_j') = \sigma_{ij} \left[ I_T \otimes c_i c_j' \right] \).

**Assumption 1.8** As \( N \to \infty \),

\(^{40}\)In particular, asymptotic no-arbitrage (see Ingersoll (1984), Eq. (7)), our Assumption 1.2, and boundedness of the maximum eigenvalue of \( \Sigma \) imply Ingersoll’s result.

\(^{41}\)It can be shown that (deterministic) convergence of \( \tilde{\Gamma} \) to \( \Gamma_\infty \) occurs at most at rate \( O \left( 1/\sqrt{N} \right) \), which equals \( O \left( 1/\sqrt{N} \right) \) by Assumption 1.2, although any faster rate is allowed for in principle. Notice that if \( \hat{\Gamma} - \Gamma_\infty \) is exactly \( O \left( 1/\sqrt{N} \right) \), then we need to modify our sampling scheme and select an arbitrary, slightly smaller, set of assets \( n \) such that \( n/N \to 0 \) as \( N \) diverges. When estimating \( \hat{\Gamma}^* \) using these \( n \) assets, then the slower \( O \left( \sqrt{N} \right) \) rate of convergence to \( \Gamma_\infty^* \) is obtained.

\(^{42}\)The case for (linear or nonlinear) dependence, whereby \( \beta_i = \beta(c_i) \), has been forcefully made by both the empirical (see Connor et al. (2012), Chordis et al. (2015), and Kelly et al. (2018), among others) and theoretical literature (see the survey in Kogan and Papamikou (2013)) in order to resolve the debate on systematic risk versus characteristic-based stories of expected returns that was spurred from the influential empirical findings of Daniel and Titman (1997).

\(^{43}\)Chordis et al. (2015) highlight the challenges that arise when estimating time-varying characteristic premia and propose a bootstrap procedure to perform correct inference in this setting.
(i)
\[
\hat{\mu}_C = \frac{C^t 1_N}{N} \rightarrow_p \mu_C = [\mu_{c1}, \ldots, \mu_{cK}]', \text{ a finite } K_c\text{-vector},
\]
\[\text{(1.79)}\]
\[
\hat{\Sigma}_{CC} = \frac{C^t C}{N} \rightarrow_p \Sigma_{CC}, \text{ a finite positive-definite } (K_c \times K_c) \text{ matrix},
\]
\[\text{(1.80)}\]
\[
\hat{\Sigma}_{CB} = \frac{C^t B}{N} \rightarrow_p \Sigma_{CB}, \text{ a finite } (K_c \times K) \text{ matrix},
\]
\[\text{(1.81)}\]
with positive-definite matrices
\[
\begin{bmatrix}
\Sigma_{CC} & \Sigma_{CB} \\
\Sigma_{CB}' & \Sigma_{\beta}
\end{bmatrix}
- 
\begin{bmatrix}
\mu_C \\
\mu_{\beta}
\end{bmatrix}
\begin{bmatrix}
\mu_C' \\
\mu_{\beta}'
\end{bmatrix}.
\]
\[\text{(1.82)}\]
(iii)
\[
\frac{C' c}{N} \rightarrow_p 0_{(K_c \times T)}.
\]
\[\text{(1.83)}\]
(iii)
\[
\frac{1}{N} \sum_{i=1}^{N} \Sigma_{zz,ii} \rightarrow \sigma^2 (I_T \otimes \Sigma_{CC}) \text{ and } \sum_{i,j=1}^{N} \Sigma_{zz,ij} 1_{(i \neq j)} = o(N).
\]
\[\text{(1.84)}\]
(iv)
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} z_i \rightarrow_d N(0_{K_c T}, \sigma^2 (I_T \otimes \Sigma_{CC})).
\]
\[\text{(1.85)}\]
Since \[
\begin{bmatrix}
\Sigma_{CC} & \Sigma_{CB} \\
\Sigma_{CB}' & \Sigma_{\beta}
\end{bmatrix}
- 
\begin{bmatrix}
\mu_C \\
\mu_{\beta}
\end{bmatrix}
\begin{bmatrix}
\mu_C' \\
\mu_{\beta}'
\end{bmatrix}
\]
in Assumption 1.8(i) is positive-definite, then \[
\begin{bmatrix}
\Sigma_{CC} & \Sigma_{CB} \\
\Sigma_{CB}' & \Sigma_{\beta}
\end{bmatrix}
\]
is also positive-definite, and this implies that the \(\beta_i\) and the \(c_i\) cannot be proportional.

In the next two theorems, we characterize the asymptotic properties of the estimators \(\hat{\Gamma}^*\) and \(\hat{\delta}^*\).

**Theorem 1.6** As \(N \rightarrow \infty\), we have

(i) Under Assumptions 1.2-1.5 and 1.8, and Eq. (1.76),
\[
\hat{\Gamma}^* - \Gamma^P = O_p \left( \frac{1}{\sqrt{N}} \right), \hat{\delta}^* - \delta = O_p \left( \frac{1}{\sqrt{N}} \right).
\]
\[\text{(1.86)}\]
(ii) Under Assumptions 1.2-1.6 and 1.8, and Eq. (1.76),
\[
\sqrt{N} \left[ \hat{\Gamma}^* - \Gamma^P \right] 
\rightarrow_d N(0_{K_c + K_c + 1}, \sigma^2 (Q'Q) L^{-1} + L^{-1} O L^{-1}),
\]
\[\text{(1.87)}\]
with
\[
L = \begin{bmatrix}
\Sigma_X \\
\mu_{\mu}'
\end{bmatrix},
L = \begin{bmatrix}
0_{K_c} & 0_{K_c}' & 0_{K_c \times (K+1)} & 0_{K_c \times K_c}
\end{bmatrix}.
\]
\[\text{(1.88)}\]
where \(Q, Z,\) and \(U,\) are defined in Theorem 1.1.

**Proof:** See Appendix B.

A consistent estimator of the asymptotic covariance matrix of \(\hat{\Gamma}^*\) and \(\hat{\delta}^*\) is provided in the next theorem.

\[\text{[44]}\] The proof of Theorem 1.7 follows the same steps of the proof of Theorem 1.2 and is therefore omitted.
Theorem 1.7 Under Assumptions 1.2-1.5 and 1.8, Eq. (1.76), and the identification condition \( \kappa_4 = 0 \), as \( N \to \infty \), we have
\[
\hat{\sigma}^2(\hat{Q}'\hat{Q})\hat{L}^{-1} + \hat{L}^{-1}\hat{O}\hat{L}^{-1} \to_p \sigma^2(Q'Q)L^{-1} + L^{-1}OL^{-1},
\]
(1.89)
with
\[
\hat{L} = \begin{bmatrix}
\hat{\Sigma}_{XX} - \hat{\Lambda}
\hat{\Sigma}_{XG}
\end{bmatrix}
\begin{bmatrix}
\hat{\mu}_G
\hat{\Sigma}_{GB}
\end{bmatrix},
\quad \hat{O} = \begin{bmatrix}
0
0_K
\hat{Z}'\hat{U}_c\hat{Z}
0_{K \times (K+1)}
0_{K \times K_c}
\end{bmatrix},
\]
(1.90)
where \( \hat{\sigma}^2 \) is defined in Eq. (1.11), and \( \hat{Q}, \hat{Z}, \) and \( \hat{U}_c \) are defined in Theorem 1.2.

1.5. Unbalanced Panels

In this section, we extend our methodology to the case of an unbalanced panel, focusing for simplicity on the base case of correctly specified models with constant risk premia. Following Gagliardini et al. (2016), we assume a missing at random design (see, for example, Rubin (1976)), that is, independence between unobservability and return generating process. This allows us to keep the factor structure linear. In the following analysis, we explicitly account for the randomness of \( T_i \), the time-series sample size for asset \( i \).

Define the following \( T \times T \) matrix
\[
J_i = \text{diag}(J_{i1}, \ldots, J_{iT}) \quad i = 1, \ldots, N,
\]
(1.91)
where \( \sum_{t=1}^{T} J_{it} = T_i \) and \( J_{it} = 1 \) if the return on asset \( i \) is observed by the econometrician at date \( t \), and zero otherwise. In addition, let \( R_{i,u} = J_i R_i, F_{i,u} = J_i F_i, \) and \( \epsilon_{i,u} = J_i \epsilon_i \), and assume that asset returns are governed by the multifactor model
\[
J_i R_i = J_i \alpha_i + J_i \beta_i' \epsilon_i + J_i \epsilon_{it},
\]
(1.92)
that is, the same data generating process of Section 1.2 pre-multiplied by \( J_i \). Let \( \bar{R}_{i,u} = \frac{1}{T_i} \sum_{t=1}^{T} J_i R_{it}, \)
\( \bar{f}_{i,u} = \frac{1}{T_i} \sum_{t=1}^{T} J_i f_{it}, \) and \( \bar{\epsilon}_{i,u} = \frac{1}{T_i} \sum_{t=1}^{T} J_i \epsilon_{it}. \)
Averaging Eq. (1.92) over time, imposing the asset-pricing restriction, and noting that \( E[R_{it}] = \alpha_i + \beta_i' E[f_i] \) yields
\[
\bar{R}_{i,u} = \gamma_0 + J_i \beta_i' \bar{f}_{i,u} + \eta_{i,u},
\]
(1.93)
where \( \gamma_{1,i,u} = \gamma_1 + J_i f_i - E(f_i), \)
\( \eta_{1,u} = \bar{\epsilon}_{i,u} - (\hat{\beta}_i - \beta_i) \gamma_{1,i,u}, \)
\( \hat{\beta}_i = \beta_i + \hat{P}_{i,u} \epsilon_i, \)
\( \hat{P}_{i,u} = \bar{F}_{i,u} (\hat{F}_{i,u}' \bar{F}_{i,u})^{-1}, \) and
\( \bar{F}_{i,u} = F_{i,u} - J_i \bar{f}_{i,u}. \) Since the panel is unbalanced, there is now a sequence of ex post risk premia, one for each asset \( i \).

In matrix form, we have
\[
\bar{R}_u = \gamma_0 1_N + \begin{bmatrix}
\hat{\beta}_{1,u}' \\
\vdots \\
\hat{\beta}_{N,u}'
\end{bmatrix}
\begin{bmatrix}
\gamma_{11,u} \\
\vdots \\
\gamma_{1N,u}
\end{bmatrix}
\begin{bmatrix}
\eta_{1,u}' \\
\vdots \\
\eta_{N,u}'
\end{bmatrix},
\]
(1.94)
where \( \bar{R}_u = (\bar{R}_{1,u}, \ldots, \bar{R}_{N,u})' \). Define the \( N \times K \) matrix \( \bar{X}_u = [1_N, \bar{B}_u], \) where \( \bar{B}_u = (\hat{\beta}_{1,u}, \ldots, \hat{\beta}_{N,u})' \).

Denote by \( \hat{\epsilon}_{i,u} \) the \( T \)-vector of residuals from the first-pass (unbalanced) OLS regressions in
\[
R_{i,u} = \alpha_i J_i 1_T + F_{i,u} \beta_i + \epsilon_{i,u}, \quad i = 1, \ldots, N.
\]
(1.95)
The modified estimator of the ex post risk premia in the unbalanced panel case is
\[
\hat{\gamma}_u = \begin{bmatrix}
\hat{\gamma}_{0,u}' \\
\hat{\gamma}_{1,u}'
\end{bmatrix} = (\hat{\Sigma}_{XX} - \hat{\Lambda}_u)^{-1} \hat{X}_u' \bar{R}_u / N,
\]
(1.96)
where $\hat{\Sigma}_{X,u} = \frac{\hat{\Sigma}^r + \hat{\Sigma}^s}{N}$, $\hat{\Lambda}_u = \begin{bmatrix} 0 & 0' \\ 0' & \hat{\Sigma}^2_u \end{bmatrix}$ with $\hat{\Sigma}^2_u = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T_i - K - 1} \text{tr} \left( \hat{\epsilon}_{i,u} \hat{\epsilon}_{i,u}' \right) \right)$ and $\hat{F}_u = \frac{1}{N} \sum_{i=1}^{N} \left( \hat{F}_{i,u} \hat{F}_{i,u}' \right)^{-1}$.

The estimator $\hat{\Gamma}^*_u$ in Eq. (1.96) generalizes the bias-adjusted estimator of Shanken (1992) to the unbalanced panel case and coincides with Shanken’s estimator when the panel is balanced. Let $\hat{\Sigma}_{X,i} = \begin{bmatrix} 1 & \beta_i' \\ \beta_i & \beta_i' \end{bmatrix}$, $\Sigma_{F,i} = \text{plim} \frac{1}{N} \sum_{i=1}^{N} \beta_i' F_i F_i' \beta_i$, $\Sigma_{X,i}$, $\hat{F}_i = \text{plim} \frac{1}{N} \sum_{i=1}^{N} P_{i,u} P_{i,u}$, and $Q_{i,u} = \frac{1}{T_i - K - 1} - P_{i,u} P'_{i,u}$.

Finally, define $Z_{i,u} = \left( Q_{i,u} \otimes P_{i,u} \right) + \frac{\text{vec} \left( M_{i,u} \right)}{T_i - K - 1} \gamma_i P_{i,u} P'_{i,u}$ and $M_{i,u} = \left[ I_{T_i} - J_i D (D' J_i D)^{-1} D' J_i \right] J_i$.

The following additional assumptions are required for the asymptotic analysis in the unbalanced panel case.

**Assumption 1.9**

$$\sup_i ||\beta_i|| \leq C < \infty,$$  \hspace{1cm} (1.97)

where $C$ is a generic constant.

**Assumption 1.10** $J_{it}$ is i.i.d. across $i$ and $t$. Let $C$ be a generic constant, and assume that $T_i > K + 1$, for every $i = 1, \ldots, N$. Then, we have

(i) $$\tau = E \left[ \frac{1}{T_i} \right] \leq C < \infty \; \text{and} \; \theta = \text{Var} \left( \frac{J_{it}}{T_i} \right) \leq C < \infty.$$ \hspace{1cm} (1.98)

(ii) $$\left| E \left[ \frac{J_{it}}{T_i} \right] \right| \leq C < \infty.$$ \hspace{1cm} (1.99)

(iii) $$E \left[ \frac{J_{it}}{(T_i - K - 1)^2} \right] \leq C < \infty.$$ \hspace{1cm} (1.100)

(iv) Let $d_t$ be the $t$-th row of $D$ and denote $p_{it,u} = J_{it} d_t (D' J_i D)^{-1} D' J_i$. Then, we have

$$E \left[ \frac{p_{it,u} P_{it,u}'}{(T_i - K - 1)^2} \right] \leq C < \infty.$$ \hspace{1cm} (1.101)

(v) Let $\hat{F}_{i,u} = F_{i,u} - J_i \frac{1_{T_i} - J_i F}{T_i}$, where $F_{i,u} = J_i F$. Then,

$$E \left[ (\hat{F}_{i,u}' \hat{F}_{i,u})^{-1} \right] \leq C < \infty,$$ \hspace{1cm} (1.102)

and

$$E \left[ \left| (\hat{F}_{i,u}' \hat{F}_{i,u})^{-1} \right|^4 \right] \leq C < \infty.$$ \hspace{1cm} (1.103)

(vi) Let $m_{i,s}^u$ be the $(t,s)$-th element of the matrix $M_{i,u} = (I_{T_i} - J_i D (D' J_i D)^{-1} D' J_i) J_i$, and define $M_{i,u}^{(2)} = M_{i,u} \odot M_{i,u}$, where $\odot$ denotes the Hadamard product operator. Then,

$$\sup_i E \left[ |m_{i,s}^u|^8 \right] \leq C < \infty,$$ \hspace{1cm} (1.104)

and

$$\frac{1}{N} \sum_{i=1}^{N} \text{tr} \left( M_{i,u}^{(2)} \right) \rightarrow_p C > 0.$$ \hspace{1cm} (1.105)
Assumption 1.9 is a boundedness assumption. In Assumption 1.10, we assume a missing at random design, that is, independence between unobservability and return generating process. Assumptions 1.10(i)-(iv) are ruling out that the distribution of the $T_i$ is too concentrated around zero. Assumption 1.10(v) is essentially extending the non-singularity of the covariance matrix of the factors to the missing-at-random design, and Assumption 1.10(vi) is technical in nature. The consistency and asymptotic normality of the proposed estimator are provided in the following theorem.

**Theorem 1.8** Under Assumptions 1.1–1.6 and 1.9–1.10, as $N \to \infty$, we have

(i)

$$\hat{\Gamma}_u - \Gamma^p = O_p \left( \frac{1}{\sqrt{N}} \right). \tag{1.106}$$

(ii)

$$\sqrt{N} \left( \hat{\Gamma}_u - \Gamma^p \right) \overset{d}{\to} \left( 0_{K+1}, V_u + \Sigma_X^{-1} (W_u + \Theta) \Sigma_X^{-1} \right), \tag{1.107}$$

where

$$V_u = \sigma^2 \left( \tau + \gamma' F \sigma X^{-1} \right), \quad W_u = \begin{bmatrix} 0 & 0' \cr 0_K & \text{plim} \frac{1}{N} \sum_{i=1}^N Z_i' U_i Z_i, \end{bmatrix}, \quad \Theta = \sigma \Sigma \beta^2 - \sigma^2 \Psi,$$

with $\Psi = \begin{bmatrix} 0_{K} & \gamma' F \sigma X^{-1} \cr F' \sigma X^{-1} F & \text{plim} \frac{1}{N} \sum_{i=1}^N P_i' u_i P_i u_i (f_i - \bar{f}) \beta, \quad F' \sigma X^{-1} F \end{bmatrix}$, where $\gamma = \text{plim} \frac{1}{N} \sum_{i=1}^N P_i' u_i P_i u_i (f_i - \bar{f}) \beta$, and $F = \text{plim} \frac{1}{N} \sum_{i=1}^N (\beta_i - \bar{\beta}) (f_i - \bar{f}) (f_i - \bar{f})' (f_i - \bar{f}) P_i' u_i P_i u_i$.

It should be noted that the asymptotic covariance matrix in Theorem 1.8 is similar to the one for the balanced panel case provided in Theorem 1.1. The additional terms in part (ii) of Theorem 1.8 account for the randomness of the sample size $T_i$. When the panel is balanced, Theorem 1.8 reduces to Theorem 1.1 since $T_i = T$, $J_i = J$, $f_i = \bar{f}$, which implies that $\tau = 1/T$, $\theta = 0$, $\Psi = 0_{(K+1) \times (K+1)}$, and all the relevant quantities do not depend on $i$ anymore.

For statistical inference, we need a consistent estimator of the asymptotic covariance matrix of $\hat{\Gamma}_u$ as illustrated in the next theorem. Let $\hat{\gamma} = \frac{1}{N} \sum_{i=1}^N \frac{1}{T_i} \gamma_{i,T_i}$, with $\gamma_{i,T_i} = \text{plim} \frac{1}{N} \sum_{i=1}^N P_i' u_i P_i u_i (f_i - \bar{f}) \beta$, and $A_i = P_i' u_i P_i u_i F' F$. Also, let $\hat{U}_i = \sum_{t=1}^{T_i} (P_i' u_i \otimes f_i') \hat{U}_t (P_i u_i \otimes P_i u_i f_i)$, where $\hat{U}_t$ (as in the balanced panel case) is a plug-in estimator of $U_t$ that depends only on $\hat{\gamma}_{i,T_i} = \frac{1}{T_i} \sum_{t=1}^{T_i} \frac{1}{T_i} \gamma_{i,T_i}$, $\hat{\beta}_{i,u} = \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{i,u}^2 \hat{\beta}_{i,u} - \hat{\beta}_{i,u}^2 P_i' u_i P_i u_i$, and $b_i = \text{tr}(F' F \Sigma_{X,i}^{-1})$, and $A_i = P_i' u_i P_i u_i F' F$. Finally, let $\hat{\Sigma}_{F\beta} = \frac{1}{N} \sum_{i=1}^N b_i \Sigma_{X,i} - \hat{\gamma}$, where $\hat{\gamma} = \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{2} \hat{\beta}_{i,u}^2 \hat{\beta}_{i,u} + \hat{\gamma}_{i,T_i} A_i + \hat{\beta}_{i,u} A_i' \right]$, with $\hat{\beta}_{i,u} = \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{i,u}^2 \hat{\beta}_{i,u} - \hat{\beta}_{i,u}^2 P_i' u_i P_i u_i$, and $\hat{\Sigma}_{X,i} = \left[ \hat{\Sigma}_{X,i} + \hat{\gamma}_{i,T_i} P_i u_i P_i u_i \right]$, where $\hat{\Sigma}_{X,i} = \hat{\gamma}_{i,T_i} P_i u_i P_i u_i$.

**Theorem 1.9** Under Assumptions 1.2–1.6 and 1.9–1.10, setting $\kappa = 0$, as $N \to \infty$, we have

$$\hat{\gamma}_u + \left( \Sigma_{X,u} - \hat{\Lambda}_u \right)^{-1} (\hat{W}_u + \hat{\Theta}) \left( \Sigma_{X,u} - \hat{\Lambda}_u \right)^{-1} \to_p V_u + \Sigma_X^{-1} (W_u + \Theta) \Sigma_X^{-1}, \tag{1.108}$$

where

$$\hat{V}_u = \left[ \hat{\sigma}_u^2 \left( \hat{\gamma}_u + \gamma' F \sigma X^{-1} \right) \left( \Sigma_{X,u} - \hat{\Lambda}_u \right)^{-1} \right] \left( \Sigma_{X,u} - \hat{\Lambda}_u \right)^{-1}, \quad \hat{W}_u = \begin{bmatrix} 0 & 0' \cr 0_K & \text{plim} \frac{1}{N} \sum_{i=1}^N Z_i' U_i Z_i, \end{bmatrix}, \quad \hat{\Theta} = \hat{\sigma} \Sigma \beta^2 - \hat{\sigma}^2 \Psi,$$
with $\hat{\Psi} = \begin{bmatrix} 0 & \hat{\gamma}_{1,u}^{T} \hat{\beta}_{1,u}^\gamma \\ \hat{\gamma}_{1,u}^{T} & \hat{\beta}_{1,u}^\gamma \end{bmatrix}$, 
$\hat{\beta}_{1,u}^\gamma = \frac{1}{N} \sum_{i=1}^{N} P_{1,i}^{u} P_{1,i}(\hat{f}_{i,u} - \bar{f})^{\gamma}_{1,u}$ and 
$\hat{\beta}_{1,u}^\beta = \frac{1}{N} \sum_{i=1}^{N} \hat{\Sigma}_{1,i}^{u} (\hat{f}_{i,u} - \bar{f})^\beta_{1,u}^\beta P_{1,i}^{u} P_{1,i} + 
\frac{1}{N} \sum_{i=1}^{N} P_{1,i}^{u} P_{1,i} \hat{\gamma}_{1,u}^{T} \hat{\beta}_{1,u}^\gamma (\hat{f}_{i,u} - \bar{f})^\gamma \hat{\Sigma}_{1,i}^{u} (\hat{f}_{i,u} - \bar{f})^\beta_{1,u}^\beta P_{1,i}^{u} P_{1,i}.$ 

Turning to specification testing, let 
\[ \hat{e}_{u}^{P} = \hat{R}_{u} - \hat{X}_{u} \hat{\gamma}_{u}^* \] (1.109)
be the $N$-vector of ex post sample pricing errors. Define $\hat{Q}_{u} = \frac{\hat{e}_{u}^{P} \hat{e}_{u}^{P}}{N}$ as the sum of squared ex post sample 
pricing errors and denote by $\hat{\Sigma}_{u} = \left( \frac{\hat{\beta}_{u}^\beta}{\hat{\beta}_{u}^\beta} - \hat{\beta}_{u}^\beta \right)$, 
$\hat{\beta} = \text{tr}(F'F\hat{\Sigma}_{u})$, $\omega_{u} = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \frac{\hat{\mu}_{i}^{u}}{\hat{\mu}_{i}^{u}} - \frac{1}{T} \right)^{2} \text{tr}(P_{i,u}f_{i,t}f_{i,t}^{P_{i,u}})$, 
and $Z_{Q_{1,u}} = \left[ (Q_{1,u} \otimes Q_{1,u}) - \frac{Q_{1,u}^{u} \text{vec}(M_{1,u}^{u})}{T_{i} - K - 1} \right]'$. Finally, consider the centered statistic 
\[ S_{u} = \sqrt{N} \left( \hat{Q}_{u} - \hat{\beta}^2 \left( \hat{\gamma}_{u}^{T} + \hat{\gamma}_{u}^{T} \hat{\beta}_{u}^\gamma \right) - \hat{\beta} \hat{\beta} \right). \] (1.110)

\textbf{Theorem 1.10} \textit{Under Assumptions 1.2–1.6 and 1.9–1.10, as $N \to \infty$, we have} 
\[ S_{u} \to_{d} (0, \mathcal{V}_{u} + \mathcal{W}_{u}), \] (1.111)

where $\mathcal{V}_{u} = \text{plim} \frac{1}{N} \sum_{i=1}^{N} \hat{Z}_{Q_{1,u}} U_{i}$, $\hat{Z}_{Q_{1,u}}$ and $\mathcal{W}_{u} = 4\sigma^{2} \text{plim} \frac{1}{N} \sum_{i=1}^{N} W_{i} W_{i}$, with 
\[ \hat{Z}_{Q_{1,u}} = Z_{Q_{1,u}} + \left( \omega_{u} \left( \text{vec}(M_{1,u}^{u}) \right) - \sum_{t=1}^{T} \left( \frac{J_{t}^{u}}{T_{i} - 1} - \frac{1}{T} \right)^{2} \text{vec}(P_{i,u}f_{i,t}f_{i,t}^{P_{i,u}}) \right) \]
and 
\[ W_{i} = \left( (\gamma_{1,u}^{P} - \gamma_{1,u}^{P})^{T} \beta_{1,u}^\gamma - \sum_{t=1}^{T} \left( \frac{J_{t}^{u}}{T_{i} - 1} - \frac{1}{T} \right)^{2} \beta_{1,u}^{P} f_{i,t} f_{i,t}^{P_{i,u}} \right)' \].

Note that when the panel is balanced, Theorem 1.10 reduces to Theorem 4 since $\frac{J_{t}^{u}}{T_{i}} = \frac{1}{T}$ and $\hat{f}_{i,u} = \bar{f}$, 
which implies that $\mathcal{W}_{u} = 0$, $Q_{1,u} = Q$, and $\hat{Z}_{Q_{1,u}} = Z_{Q_{1,u}} = Z_{Q}$. This variance can be consistently estimated.

Let $\hat{Z}_{Q_{1,u}} = \left[ (Q_{1,u} \otimes Q_{1,u}) - \frac{Q_{1,u}^{u} \text{vec}(M_{1,u}^{u})}{T_{i} - K - 1} \right]'$ and $\hat{Z}_{Q_{1,u}} = \hat{Z}_{Q_{1,u}} + \left( \omega_{u} \left( \text{vec}(M_{1,u}^{u}) \right) - \sum_{t=1}^{T} \left( \frac{J_{t}^{u}}{T_{i}} - \frac{1}{T} \right)^{2} \text{vec}(P_{i,u}f_{i,t}f_{i,t}^{P_{i,u}}) \right)$. 
Then, the estimators of $\mathcal{V}_{u}$ and $\mathcal{W}_{u}$ are given by 
\[ \hat{V}_{u} = \frac{1}{N} \sum_{i=1}^{N} \hat{Z}_{Q_{1,u}} \hat{U}_{i,u} \hat{Z}_{Q_{1,u}} \] (1.112)
and 
\[ \hat{W}_{u} = 4\sigma^{2} \frac{1}{N} \sum_{i=1}^{N} \left( \hat{Q}_{1,u}^{u} \hat{Q}_{1,u}(\hat{f}_{i,u} - \bar{f})^{\gamma} \hat{\Sigma}_{1,u}^{u} (\hat{f}_{i,u} - \bar{f}) \right) \]
\[ + 4\sigma^{2} \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{t=1}^{T} \left( \frac{J_{t}}{T_{i}} - \frac{1}{T} \right)^{4} \text{tr}(f_{i,t} f_{i,t}^{P_{i,u}} P_{i,u} f_{i,t} f_{i,t}^{P_{i,u}}) \right) \]
\[ - 2\hat{Q}_{1,u} P_{1,u} \sum_{t=1}^{T} \left( \frac{J_{t}}{T_{i}} - \frac{1}{T} \right)^{2} f_{i,t} f_{i,t}^{P_{i,u}} (\hat{f}_{i,u} - \bar{f}) \right), \] (1.113)
1.6. Monte Carlo simulations

In this section, we undertake a Monte Carlo simulation experiment to study the empirical rejection rates of the specification test and $t$-ratios of the bias-adjusted estimator of Shanken (1992). The return-generating process under the null of a correctly specified asset-pricing model is given by

$$ R_t = \gamma_0 1_T + B (\gamma_1 + f_t - E[f_t]) + \epsilon_t, $$  \hspace{1cm} (1.114)

where $\epsilon_t \sim \mathcal{N}(0, \Sigma)$. To study the power of the specification test, we generate the returns on the test assets as in Eq. (1.2), that is, we do not impose the asset-pricing restriction.

In all of our simulation experiments, we consider balanced panels with a time-series dimension of $T = 36$ and $T = 72$ observations. Specifically, $f_t$ in Eq. (1.114) is the excess market return (from Kenneth French’s website) from January 2008 to December 2010 for $T = 36$, and the excess market return from January 2008 to December 2013 for $T = 72$. In our simulation designs, the factor realizations are taken as given and kept fixed throughout. This is consistent with the fact that our analysis of the ex post risk premia is conditional on the realizations of the factors. In addition, $E[f_t]$ in Eq. (1.114) is set equal to the time-series mean of $f_t$ over the 2008–2010 sample when performing the analysis for $T = 36$ and to the time-series mean of $f_t$ over the 2008–2013 sample when performing the analysis for $T = 72$. To obtain representative values for the parameters $\gamma_0$, $\gamma_1$, $B$, and $\Sigma$ in Eq. (1.114) and Eq. (1.2), we employ a cross-section of 3,000 stocks from CRSP in addition to the excess market return. Based on this balanced panel of 3,000 stock returns and the excess market return, for each time-series sample size, we compute the OLS estimates of $B$, $\gamma_0$, and $\gamma_1$. Then, we set the $B$, $\gamma_0$, and $\gamma_1$ parameters in Eq. (1.114) and in Eq. (1.2) equal to these OLS estimates. The calibration of $\Sigma$ is a more delicate task and is described below. In the simulations, we consider cross-sections of $N = 100, 500, 1,000$, and 3,000 stocks. All results are based on 10,000 Monte Carlo replications. Our econometric approach, designed for large $N$ and fixed $T$, should be able to handle this large number of assets over relative short time spans. The rejection rates of the various tests are computed using our asymptotic results in the paper.

1.6.1 Percentage errors and root mean squared errors of the estimates

We start from the case in which $\Sigma$ is a spherical matrix, that is, $\Sigma = \sigma^2 I_T$. In the simulations, we set $\sigma^2$ equal to the cross-sectional average (over the 3000 stocks) of the $\sigma^2_t$ estimated from the data. Table 1.1 reports the percentage error (Bias) and root mean squared error (RMSE), all in percent, of the OLS estimator and of the bias-adjusted estimator of Shanken (1992). Panels A and B are for $T = 36$ and $T = 72$, respectively.

| Table 1.1 about here |

Panel A shows that the bias of the OLS estimator is substantial. For $\gamma_0$, the bias ranges from 28.8% for $N = 100$ to 22.9% for $N = 3000$, while for $\gamma_1$ the bias ranges from $-24.8\%$ for $N = 100$ to $-17.8\%$ for $N = 3000$. For $\Gamma^s$, the bias is small for $N = 100$ ($-2.3\%$ for $\gamma_0^s$ and $1.8\%$ for $\gamma_1^s$) and becomes negligible for $N \geq 500$. As for the RMSE, the typical bias-variance trade-off emerges up to $N = 500$, with the OLS estimator exhibiting a smaller RMSE than the OLS bias-adjusted estimator. When $N > 500$, the RMSE of the bias-adjusted estimator of Shanken (1992) becomes substantially smaller than the one of the OLS estimator. Panel B for $T = 72$ conveys a similar message. As expected from the theoretical analysis, the larger time-series dimension helps in reducing the bias and RMSE associated with the OLS estimator. However, the bias for the OLS estimator is still substantial and ranges from $-18.5\%$ for $N = 100$ to $-11.7\%$ for $N = 3000$. For the bias-adjusted estimator, the bias becomes negligible, even for $N = 100$ when $T = 72$.

Next, we consider the case in which the $\Sigma$ matrix is either diagonal or full. As emphasized above, our theoretical results hinge upon the assumption that the model disturbances are weakly cross-sectionally correlated. In order to generate shocks under a weak factor structure, we consider the following data-generating process (DGP). Define

$$ \epsilon^{(1)} = \eta \left( \frac{\sqrt{\theta}}{\sqrt{N}} \right) \epsilon' + \sqrt{1-\theta} Z, $$  \hspace{1cm} (1.115)
where $\eta$ and $c$ are $T$ and $N$-vectors of i.i.d. standard normal random variables, respectively, $Z$ is a $T \times N$ matrix of i.i.d. standard normal random variables, $0 \leq \theta \leq 1$ is a shrinkage parameter that controls the weight assigned to the diagonal and extra-diagonal elements of $\Sigma$, and $\delta$ is a parameter that controls the strength of the cross-sectional dependence of the shocks (the bigger $\delta$ is, the weaker the dependence). Our $T \times N$ matrix of shocks is then generated as

$$c = \epsilon^{(1)} \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \cdots & \cdots & \sigma_N^2 \\ \sigma_1^2 & \sigma_2^2 & \cdots & \cdots & \sigma_N^2 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \sigma_1^2 & \sigma_2^2 & \cdots & \cdots & \sigma_N^2 \end{bmatrix}^{0.5} \begin{bmatrix} \frac{\theta}{N\pi} \sigma_1^2 + (1 - \theta) \\ \frac{\theta}{N\pi} \sigma_2^2 + (1 - \theta) \\ \vdots \\ \frac{\theta}{N\pi} \sigma_N^2 + (1 - \theta) \end{bmatrix}^{-0.5}$$

(1.116)

where $c_i$ is the $i$-th element of $c$. Given this specification for the shocks, for our theoretical results to hold, we require $\delta > 0$.

As discussed in the paper, the factor structure in Eqs. (1.115)-(1.116) induces a substantial degree of cross-correlation between the $\epsilon_{it}$. We demonstrate this by means of a simple numerical example. For each Monte Carlo replication, we compute the following quantity for the data generating process above:

$$A(\delta, N) = \frac{1}{N} \sum_{j=1}^{N} \left( \frac{\sum_{i=1}^{N} |\sigma_{ij}|}{\sigma_j^2 + \sum_{i=1}^{N} |\sigma_{ij}|} \right).$$

(1.117)

For the case of zero cross-correlation, that is, $\sigma_{ij} = 0$ when $i \neq j$, $A(\delta, N) = 0$. In contrast, when the cross-correlations become big, $A(\delta, N)$ approaches 1. As we vary $\delta$ and $N$ (average across 1000 Monte Carlo iterations), we obtain

<table>
<thead>
<tr>
<th>$A(\delta, N)$</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Lambda(0.25, 100)$</td>
<td>0.8585</td>
</tr>
<tr>
<td>$\Lambda(0.25, 500)$</td>
<td>0.9268</td>
</tr>
<tr>
<td>$\Lambda(0.25, 3000)$</td>
<td>0.9628</td>
</tr>
<tr>
<td>$\Lambda(0.255, 100)$</td>
<td>0.8560</td>
</tr>
<tr>
<td>$\Lambda(0.255, 500)$</td>
<td>0.9235</td>
</tr>
<tr>
<td>$\Lambda(0.255, 3000)$</td>
<td>0.9606</td>
</tr>
<tr>
<td>$\Lambda(0.5, 100)$</td>
<td>0.5418</td>
</tr>
<tr>
<td>$\Lambda(0.5, 500)$</td>
<td>0.5680</td>
</tr>
<tr>
<td>$\Lambda(0.5, 3000)$</td>
<td>0.5612</td>
</tr>
</tbody>
</table>

This simple numerical example shows that for every value of $\delta$ (which measures the degree of dependence), the sum of the cross-covariances tends to increase with $N$ in relative terms.

In Table 1.2, we report results for the diagonal case, that is, we set $\theta = 0$ in the above data-generating process. To obtain representative values of the shock variances, while accounting for the fact that $\tilde{\Sigma}$ is ill conditioned when $T$ is small and $N$ is large, we first estimate the residual variances from the historical data. Then, at each Monte Carlo iteration, we generate a string of Beta$(p, q)$-distributed random variables with the $p$ and $q$ parameters calibrated to the cross-sectional mean and variance of the $\tilde{\sigma}_t^2$. This resampling procedure is used to minimize the impact of an ill-conditioned $\tilde{\Sigma}$ on the simulation results.

| Table 1.2 about here |

Overall, we find that the OLS estimator exhibits a slightly higher bias compared to the spherical $\Sigma$ case. The bias-adjusted estimator of Shanken (1992) continues to perform very well in terms of bias for all the
time-series and cross-sectional dimensions considered. The RMSEs of both estimators are now a bit higher than in the spherical case, and the bias-adjusted estimator still outperforms the OLS estimator for \( N \geq 500 \).

Finally, in Tables 1.3 and 1.4, we allow for weak cross-sectional dependence of the model disturbances by setting \( \theta = 0.5 \) in the above DGP.

| Tables 1.3 and 1.4 about here |

In Table 1.3, we consider the situation in which \( \delta \), the parameter that regulates the strength of the cross-sectional dependence, is equal to 0.5. Consistent with our theoretical results, the bias-adjusted estimator continues to perform very well in this scenario. Setting \( \delta = 0.25 \) in Table 1.4 has only a modest effect on the bias and RMSEs of the two estimators. Overall, the first four tables reveal a superiority of the bias-adjusted estimator of Shanken (1992) over the OLS estimator, not only in terms of bias, but also in terms of RMSE when \( N > 500 \). Furthermore, the bias-adjusted estimator shows little sensitivity to changes in the length of the time-series, consistent with the idea that this estimator should perform well for any given \( T \).

### 1.6.2 Rejection rates of the \( t \)-tests

In Tables 1.5 through 1.8, we consider the empirical rejection rates of the centered \( t \)-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10\%, 5\%, and 1\%) and for different values of the number of time-series and cross-sectional observations using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with covariance matrix calibrated as in Tables 1.1 through 1.4. The \( t \)-statistics are compared with the critical values from a standard normal distribution. We consider three \( t \)-statistics. For the OLS estimator of the cross section premium, the first \( t \)-statistic is the one that uses the traditional Fama-MacBeth standard error \((t_{FM})\), while the second \( t \)-statistic \((t_{EIV})\) is the one that uses the EIV-adjusted standard error in Theorem 1(ii) of Shanken (1992). Both of these \( t \)-statistics were developed in a large \( T \) and fixed \( N \) framework. We report them here to determine how misleading inference can be when using these \( t \)-statistics in a large \( N \) and fixed \( T \) setup. Finally, the third \( t \)-statistic is the one associated with the bias-adjusted estimator of Shanken (1992) and is based on the asymptotic distribution in part (ii) of our Theorem 1.

| Table 1.5 about here |

Starting from the spherical \( \Sigma \) case, Table 1.5 shows that the \( t \)-statistics associated with the OLS estimator only slightly overreject the null hypothesis for \( N = 100 \). However, as \( N \) increases, the performance of these \( t \)-statistics substantially deteriorates. For example, when \( N = 3000 \), the rejections rate of the Fama-MacBeth \( t \)-statistic associated with \( \hat{\gamma}_1 \) is either 41.6\% for \( T = 36 \) or 33.3\% for \( T = 72 \) at the 5\% nominal level. The strong size distortions of the Fama-MacBeth \( t \)-test do not show any improvement when accounting for the EIV bias, due to the estimation of the betas in the first stage. In contrast, our proposed \( t \)-statistic, based on Theorems 1 and 2, performs extremely well for all \( T \) and \( N \). A similar picture emerges in the \( \Sigma \) full case (Tables 1.6 and 1.7), with the rejection rates of our proposed \( t \)-test being always aligned with the critical values from a standard normal distribution.

| Tables 1.6 and 1.7 about here |

In Table 1.8, we increase the strength of the cross-sectional dependence of the residuals by setting \( \delta = 0.25 \).

| Table 1.8 about here |

In this situation, we start to notice some slight over-rejections for the \( t \)-test associated with the bias-adjusted estimator of Shanken (1992). For example, when \( T = 36 \) and \( N = 3000 \), the rejection rate for the \( t \)-test associated with \( \hat{\gamma}_1 \) is 6.8\% at the 5\% level, and when \( T = 72 \) and \( N = 3000 \), the rejection rate for the \( t \)-test associated with \( \hat{\gamma}_1 \) is 5.8\% at the 5\% level. Overall, these results suggest that our proposed
1.6. MONTE CARLO SIMULATIONS

t-test is relatively well behaved even when moving toward a fairly strong factor structure in the residuals. Furthermore, using the standard tools that were developed in a large \( T \) and fixed \( N \) framework can lead to strong over-rejections of the null hypothesis, with the likely consequence that a factor will be found to be priced even when it does not help in explaining the cross-sectional variation in individual stock returns.

1.6.3 Rejection rates of the specification test

In Tables 1.9 and 1.10, we investigate the size and power properties of our specification test \( S^* \) based on the results in Theorem 4. Table 1.9 refers to \( T = 36 \), while Table 1.10 is for \( T = 72 \).

Since our test statistic \( S^* \) has a standard normal distribution, we consider two-sided \( p \)-values in the computation of the rejection rates. The results in the two tables suggest that the rejection rates of our test under the null that the model is correctly specified are excellent for the spherical and diagonal cases. When simulating with \( \Sigma \) full, the specification test is very well sized when \( \delta = 0.5 \) but it over-rejects a bit too much when \( \delta = 0.25 \). The power properties of our specification test are fairly good when \( N = 100 \) and excellent when \( N \geq 500 \). As expected, power increases when the number of assets becomes large and the rejection rates are similar across time-series sample sizes. Overall, these simulation results suggest that our test \( S^* \) should be fairly reliable for the time-series and cross-sectional dimensions encountered in our empirical work.
Table 1.1
Bias and root mean squared error of the OLS and bias-adjusted Shanken (1992) estimators in a one-factor model ($\Sigma$ spherical)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\gamma}_0$</td>
<td>28.8%</td>
<td>26.2%</td>
<td>24.6%</td>
<td>22.9%</td>
</tr>
<tr>
<td>$\hat{\gamma}'_0$</td>
<td>-2.3%</td>
<td>-0.3%</td>
<td>0.3%</td>
<td>-0.2%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_0$)</td>
<td>0.3675</td>
<td>0.1875</td>
<td>0.1427</td>
<td>0.1066</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}'_0$)</td>
<td>0.4509</td>
<td>0.1892</td>
<td>0.1255</td>
<td>0.0699</td>
</tr>
<tr>
<td>$\hat{\gamma}_1$</td>
<td>-24.8%</td>
<td>-20.0%</td>
<td>-18.8%</td>
<td>-17.8%</td>
</tr>
<tr>
<td>$\hat{\gamma}'_1$</td>
<td>1.8%</td>
<td>0.1%</td>
<td>-0.2%</td>
<td>0.2%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_1$)</td>
<td>0.3539</td>
<td>0.1642</td>
<td>0.1277</td>
<td>0.1000</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}'_1$)</td>
<td>0.4529</td>
<td>0.1655</td>
<td>0.1098</td>
<td>0.0609</td>
</tr>
</tbody>
</table>

Panel A: $T = 36$

Panel B: $T = 72$

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets, for the OLS estimator $\hat{\gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]'$ and the bias-adjusted estimator $\hat{\gamma}^* = [\hat{\gamma}_0^*, \hat{\gamma}_1^*]'$. The model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12.
### Table 1.2
Bias and root mean squared error of the OLS and bias-adjusted Shanken (1992) estimators in a one-factor model ($\Sigma$ diagonal)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias($\hat{\gamma}_0$)</td>
<td>30.1%</td>
<td>25.8%</td>
<td>24.8%</td>
<td>23.0%</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_b$)</td>
<td>-0.7%</td>
<td>-0.8%</td>
<td>0.4%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_0$)</td>
<td>0.4047</td>
<td>0.1976</td>
<td>0.1495</td>
<td>0.1100</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_b$)</td>
<td>0.5027</td>
<td>0.2054</td>
<td>0.1364</td>
<td>0.0763</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_1$)</td>
<td>-25.5%</td>
<td>-19.6%</td>
<td>-18.7%</td>
<td>-17.9%</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_t$)</td>
<td>0.9%</td>
<td>0.6%</td>
<td>-0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_1$)</td>
<td>0.3949</td>
<td>0.1733</td>
<td>0.1339</td>
<td>0.1033</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_t$)</td>
<td>0.5104</td>
<td>0.1815</td>
<td>0.1208</td>
<td>0.0681</td>
</tr>
</tbody>
</table>

**Panel A: $T = 36$**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias($\hat{\gamma}_0$)</td>
<td>11.2%</td>
<td>10.0%</td>
<td>8.6%</td>
<td>8.0%</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_b$)</td>
<td>-1.2%</td>
<td>0.2%</td>
<td>-0.1%</td>
<td>0.0%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_0$)</td>
<td>0.2673</td>
<td>0.1246</td>
<td>0.0899</td>
<td>0.0643</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_b$)</td>
<td>0.3116</td>
<td>0.1223</td>
<td>0.0804</td>
<td>0.0446</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_1$)</td>
<td>-18.1%</td>
<td>-14.3%</td>
<td>-12.3%</td>
<td>-11.8%</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_t$)</td>
<td>1.5%</td>
<td>-0.3%</td>
<td>0.3%</td>
<td>-0.0%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_1$)</td>
<td>0.2621</td>
<td>0.1112</td>
<td>0.0809</td>
<td>0.0612</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_t$)</td>
<td>0.3120</td>
<td>0.1087</td>
<td>0.0711</td>
<td>0.0400</td>
</tr>
</tbody>
</table>

**Panel B: $T = 72$**

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets, for the OLS estimator $\hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]$ and the bias-adjusted estimator $\hat{\Gamma}^* = [\hat{\gamma}_0, \hat{\gamma}_1]^*$. The model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12.
### Table 1.3
Bias and root mean squared error of the OLS and bias-adjusted Shanken (1992) estimators in a one-factor model ($\Sigma$ full, $\delta = 0.5$)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: $T = 36$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias((\hat{\gamma}_0))</td>
<td>28.8%</td>
<td>26.0%</td>
<td>24.6%</td>
<td>22.7%</td>
</tr>
<tr>
<td>Bias((\hat{\gamma}_0^*))</td>
<td>-2.6%</td>
<td>-0.6%</td>
<td>0.3%</td>
<td>-0.4%</td>
</tr>
<tr>
<td>RMSE((\hat{\gamma}_0))</td>
<td>0.4065</td>
<td>0.1960</td>
<td>0.1506</td>
<td>0.1089</td>
</tr>
<tr>
<td>RMSE((\hat{\gamma}_0^*))</td>
<td>0.5081</td>
<td>0.2031</td>
<td>0.1385</td>
<td>0.0760</td>
</tr>
<tr>
<td>Bias((\hat{\gamma}_1))</td>
<td>-24.2%</td>
<td>-19.6%</td>
<td>-18.9%</td>
<td>-17.7%</td>
</tr>
<tr>
<td>Bias((\hat{\gamma}_1^*))</td>
<td>2.7%</td>
<td>0.7%</td>
<td>-0.3%</td>
<td>0.3%</td>
</tr>
<tr>
<td>RMSE((\hat{\gamma}_1))</td>
<td>0.3963</td>
<td>0.1727</td>
<td>0.1352</td>
<td>0.1028</td>
</tr>
<tr>
<td>RMSE((\hat{\gamma}_1^*))</td>
<td>0.5159</td>
<td>0.1806</td>
<td>0.1220</td>
<td>0.0681</td>
</tr>
</tbody>
</table>

Panel B: $T = 72$

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias((\hat{\gamma}_0))</td>
<td>11.8%</td>
<td>9.4%</td>
<td>8.6%</td>
<td>8.0%</td>
</tr>
<tr>
<td>Bias((\hat{\gamma}_0^*))</td>
<td>-0.5%</td>
<td>-0.5%</td>
<td>-0.1%</td>
<td>-0.0%</td>
</tr>
<tr>
<td>RMSE((\hat{\gamma}_0))</td>
<td>0.2671</td>
<td>0.1227</td>
<td>0.0910</td>
<td>0.0642</td>
</tr>
<tr>
<td>RMSE((\hat{\gamma}_0^*))</td>
<td>0.3099</td>
<td>0.1225</td>
<td>0.0820</td>
<td>0.0447</td>
</tr>
<tr>
<td>Bias((\hat{\gamma}_1))</td>
<td>-19.0%</td>
<td>-13.6%</td>
<td>-12.4%</td>
<td>-11.7%</td>
</tr>
<tr>
<td>Bias((\hat{\gamma}_1^*))</td>
<td>0.5%</td>
<td>0.6%</td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>RMSE((\hat{\gamma}_1))</td>
<td>0.2614</td>
<td>0.1104</td>
<td>0.0819</td>
<td>0.0611</td>
</tr>
<tr>
<td>RMSE((\hat{\gamma}_1^*))</td>
<td>0.3096</td>
<td>0.1100</td>
<td>0.0720</td>
<td>0.0405</td>
</tr>
</tbody>
</table>

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets, for the OLS estimator $\hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]^\prime$ and the bias-adjusted estimator $\hat{\Gamma}^* = [\hat{\gamma}_0^*, \hat{\gamma}_1^*]^\prime$. The model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12.
Table 1.4
Bias and root mean squared error of the OLS and bias-adjusted Shanken (1992) estimators in a one-factor model (\( \Sigma \) full, \( \delta = 0.25 \))

<table>
<thead>
<tr>
<th>Statistics</th>
<th>( N = 100 )</th>
<th>( N = 500 )</th>
<th>( N = 1000 )</th>
<th>( N = 3000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: ( T = 36 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias(( \hat{\gamma}_0 ))</td>
<td>28.8%</td>
<td>26.6%</td>
<td>24.2%</td>
<td>23.5%</td>
</tr>
<tr>
<td>Bias(( \hat{\gamma}_0^* ))</td>
<td>−2.5%</td>
<td>0.1%</td>
<td>−0.3%</td>
<td>0.5%</td>
</tr>
<tr>
<td>RMSE(( \hat{\gamma}_0 ))</td>
<td>0.4191</td>
<td>0.2053</td>
<td>0.1536</td>
<td>0.1135</td>
</tr>
<tr>
<td>RMSE(( \hat{\gamma}_0^* ))</td>
<td>0.5254</td>
<td>0.2152</td>
<td>0.1450</td>
<td>0.0809</td>
</tr>
<tr>
<td>Bias(( \hat{\gamma}_1 ))</td>
<td>−24.8%</td>
<td>−19.9%</td>
<td>−18.5%</td>
<td>−18.3%</td>
</tr>
<tr>
<td>Bias(( \hat{\gamma}_1^* ))</td>
<td>2.0%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>−0.4%</td>
</tr>
<tr>
<td>RMSE(( \hat{\gamma}_1 ))</td>
<td>0.4116</td>
<td>0.1824</td>
<td>0.1380</td>
<td>0.1072</td>
</tr>
<tr>
<td>RMSE(( \hat{\gamma}_1^* ))</td>
<td>0.5355</td>
<td>0.1935</td>
<td>0.1288</td>
<td>0.0731</td>
</tr>
<tr>
<td>Panel B: ( T = 72 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias(( \hat{\gamma}_0 ))</td>
<td>12.2%</td>
<td>9.7%</td>
<td>8.8%</td>
<td>7.9%</td>
</tr>
<tr>
<td>Bias(( \hat{\gamma}_0^* ))</td>
<td>−0.1%</td>
<td>−0.2%</td>
<td>0.1%</td>
<td>−0.1%</td>
</tr>
<tr>
<td>RMSE(( \hat{\gamma}_0 ))</td>
<td>0.2795</td>
<td>0.1287</td>
<td>0.0939</td>
<td>0.0645</td>
</tr>
<tr>
<td>RMSE(( \hat{\gamma}_0^* ))</td>
<td>0.3252</td>
<td>0.1292</td>
<td>0.0853</td>
<td>0.0459</td>
</tr>
<tr>
<td>Bias(( \hat{\gamma}_1 ))</td>
<td>−19.3%</td>
<td>−13.9%</td>
<td>−12.6%</td>
<td>−11.7%</td>
</tr>
<tr>
<td>Bias(( \hat{\gamma}_1^* ))</td>
<td>0.0%</td>
<td>0.2%</td>
<td>−0.1%</td>
<td>0.2%</td>
</tr>
<tr>
<td>RMSE(( \hat{\gamma}_1 ))</td>
<td>0.2761</td>
<td>0.1155</td>
<td>0.0854</td>
<td>0.0615</td>
</tr>
<tr>
<td>RMSE(( \hat{\gamma}_1^* ))</td>
<td>0.3279</td>
<td>0.1158</td>
<td>0.0763</td>
<td>0.0416</td>
</tr>
</tbody>
</table>

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets, for the OLS estimator \( \hat{\Gamma} = [\hat{\gamma}_0', \hat{\gamma}_1'] \) and the bias-adjusted estimator \( \hat{\Gamma}^* = [\hat{\gamma}_0^*, \hat{\gamma}_1^*]' \). The model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1−2013:12.
Table 1.5  
Size of t-tests in a one-factor model (Σ spherical)

Panel A: $T = 36$

<table>
<thead>
<tr>
<th></th>
<th>$t_{FM}(\hat{\gamma}_0)$</th>
<th>$t_{FM}(\hat{\gamma}_1)$</th>
<th>$t_{EIV}(\hat{\gamma}_0)$</th>
<th>$t_{EIV}(\hat{\gamma}_1)$</th>
<th>$t(\hat{\gamma}_0^*)$</th>
<th>$t(\hat{\gamma}_1^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>0.10</td>
<td>0.05</td>
<td>0.01</td>
<td>0.10</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>100</td>
<td>0.128</td>
<td>0.074</td>
<td>0.021</td>
<td>0.141</td>
<td>0.078</td>
<td>0.022</td>
</tr>
<tr>
<td>500</td>
<td>0.186</td>
<td>0.113</td>
<td>0.040</td>
<td>0.213</td>
<td>0.132</td>
<td>0.047</td>
</tr>
<tr>
<td>1000</td>
<td>0.243</td>
<td>0.156</td>
<td>0.059</td>
<td>0.289</td>
<td>0.197</td>
<td>0.075</td>
</tr>
<tr>
<td>3000</td>
<td>0.438</td>
<td>0.324</td>
<td>0.153</td>
<td>0.538</td>
<td>0.416</td>
<td>0.219</td>
</tr>
</tbody>
</table>
Table 1.5  (Continued)
Size of $t$-tests in a one-factor model ($\Sigma$ spherical)

Panel B: $T = 72$

<table>
<thead>
<tr>
<th>N</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t_{FM}(\hat{\gamma}_0)$</td>
<td>$t_{FM}(\hat{\gamma}_1)$</td>
<td>$t_{EIV}(\hat{\gamma}_0)$</td>
<td>$t_{EIV}(\hat{\gamma}_1)$</td>
<td>$t(\hat{\gamma}_0)$</td>
<td>$t(\hat{\gamma}_1)$</td>
</tr>
<tr>
<td>100</td>
<td>0.123</td>
<td>0.063</td>
<td>0.016</td>
<td>0.124</td>
<td>0.066</td>
<td>0.016</td>
</tr>
<tr>
<td>500</td>
<td>0.167</td>
<td>0.099</td>
<td>0.030</td>
<td>0.181</td>
<td>0.109</td>
<td>0.033</td>
</tr>
<tr>
<td>1000</td>
<td>0.211</td>
<td>0.133</td>
<td>0.041</td>
<td>0.237</td>
<td>0.154</td>
<td>0.053</td>
</tr>
<tr>
<td>3000</td>
<td>0.378</td>
<td>0.263</td>
<td>0.109</td>
<td>0.449</td>
<td>0.333</td>
<td>0.150</td>
</tr>
<tr>
<td>100</td>
<td>0.122</td>
<td>0.063</td>
<td>0.015</td>
<td>0.123</td>
<td>0.065</td>
<td>0.016</td>
</tr>
<tr>
<td>500</td>
<td>0.166</td>
<td>0.099</td>
<td>0.030</td>
<td>0.181</td>
<td>0.108</td>
<td>0.033</td>
</tr>
<tr>
<td>1000</td>
<td>0.210</td>
<td>0.132</td>
<td>0.040</td>
<td>0.236</td>
<td>0.153</td>
<td>0.052</td>
</tr>
<tr>
<td>3000</td>
<td>0.377</td>
<td>0.261</td>
<td>0.108</td>
<td>0.448</td>
<td>0.331</td>
<td>0.149</td>
</tr>
<tr>
<td>100</td>
<td>0.096</td>
<td>0.047</td>
<td>0.009</td>
<td>0.100</td>
<td>0.048</td>
<td>0.009</td>
</tr>
<tr>
<td>500</td>
<td>0.097</td>
<td>0.049</td>
<td>0.010</td>
<td>0.098</td>
<td>0.049</td>
<td>0.010</td>
</tr>
<tr>
<td>1000</td>
<td>0.100</td>
<td>0.047</td>
<td>0.009</td>
<td>0.103</td>
<td>0.048</td>
<td>0.009</td>
</tr>
<tr>
<td>3000</td>
<td>0.103</td>
<td>0.054</td>
<td>0.010</td>
<td>0.106</td>
<td>0.054</td>
<td>0.010</td>
</tr>
</tbody>
</table>

The table presents the size properties of $t$-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks ($N$) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12. $t_{FM}(\cdot)$ denotes the $t$-statistic associated with the OLS estimator $\hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]'$, which uses the traditional Fama-MacBeth standard error, $t_{EIV}(\cdot)$ denotes the $t$-statistic associated with the OLS estimator $\hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]'$, which uses the EIV-adjusted standard error in Theorem 1(ii) of Shanken (1992) and $t(\cdot)$ denotes the $t$-statistic associated with the Shanken estimator $\hat{\Gamma}^* = [\hat{\gamma}_0^*, \hat{\gamma}_1^*]'$, which uses the standard error formulae of Theorem 2. Finally, the rejection rates for the $t$-test associated with the bias-adjusted estimator of Shanken (1992) are based on the asymptotic distribution in part (ii) of Theorem 1. The $t$-statistics are compared with the critical values from a standard normal distribution.
Table 1.6
Size of $t$-tests in a one-factor model ($\Sigma$ diagonal)

Panel A: $T = 36$

<table>
<thead>
<tr>
<th>$N$</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t_{FM}(\hat{\gamma}_0)$</td>
<td></td>
<td></td>
<td>$t_{FM}(\hat{\gamma}_1)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.122</td>
<td>0.066</td>
<td>0.019</td>
<td>0.125</td>
<td>0.072</td>
<td>0.018</td>
</tr>
<tr>
<td>500</td>
<td>0.163</td>
<td>0.104</td>
<td>0.033</td>
<td>0.179</td>
<td>0.112</td>
<td>0.036</td>
</tr>
<tr>
<td>1000</td>
<td>0.226</td>
<td>0.141</td>
<td>0.050</td>
<td>0.248</td>
<td>0.166</td>
<td>0.060</td>
</tr>
<tr>
<td>3000</td>
<td>0.398</td>
<td>0.292</td>
<td>0.128</td>
<td>0.474</td>
<td>0.362</td>
<td>0.174</td>
</tr>
<tr>
<td></td>
<td>$t_{EIV}(\hat{\gamma}_0)$</td>
<td></td>
<td></td>
<td>$t_{EIV}(\hat{\gamma}_1)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.120</td>
<td>0.065</td>
<td>0.018</td>
<td>0.124</td>
<td>0.070</td>
<td>0.017</td>
</tr>
<tr>
<td>500</td>
<td>0.163</td>
<td>0.103</td>
<td>0.033</td>
<td>0.179</td>
<td>0.111</td>
<td>0.036</td>
</tr>
<tr>
<td>1000</td>
<td>0.225</td>
<td>0.141</td>
<td>0.050</td>
<td>0.247</td>
<td>0.165</td>
<td>0.060</td>
</tr>
<tr>
<td>3000</td>
<td>0.397</td>
<td>0.291</td>
<td>0.127</td>
<td>0.473</td>
<td>0.362</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>$t(\hat{\gamma}_0^*)$</td>
<td></td>
<td></td>
<td>$t(\hat{\gamma}_1^*)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.093</td>
<td>0.045</td>
<td>0.011</td>
<td>0.091</td>
<td>0.044</td>
<td>0.010</td>
</tr>
<tr>
<td>500</td>
<td>0.102</td>
<td>0.051</td>
<td>0.010</td>
<td>0.096</td>
<td>0.049</td>
<td>0.011</td>
</tr>
<tr>
<td>1000</td>
<td>0.099</td>
<td>0.048</td>
<td>0.009</td>
<td>0.101</td>
<td>0.051</td>
<td>0.009</td>
</tr>
<tr>
<td>3000</td>
<td>0.099</td>
<td>0.053</td>
<td>0.012</td>
<td>0.099</td>
<td>0.051</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Table 1.6 (Continued)
Size of \( t \)-tests in a one-Factor model (\( \Sigma \) diagonal)

Panel B: \( T = 72 \)

<table>
<thead>
<tr>
<th>( N )</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t_{FM}(\hat{\gamma}_0) )</td>
<td>( t_{FM}(\hat{\gamma}_1) )</td>
<td>( t_{EIV}(\hat{\gamma}_0) )</td>
<td>( t_{EIV}(\hat{\gamma}_1) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.115</td>
<td>0.060</td>
<td>0.015</td>
<td>0.121</td>
<td>0.064</td>
<td>0.015</td>
</tr>
<tr>
<td>500</td>
<td>0.157</td>
<td>0.089</td>
<td>0.027</td>
<td>0.165</td>
<td>0.096</td>
<td>0.030</td>
</tr>
<tr>
<td>1000</td>
<td>0.199</td>
<td>0.121</td>
<td>0.036</td>
<td>0.219</td>
<td>0.137</td>
<td>0.044</td>
</tr>
<tr>
<td>3000</td>
<td>0.353</td>
<td>0.250</td>
<td>0.103</td>
<td>0.416</td>
<td>0.302</td>
<td>0.134</td>
</tr>
</tbody>
</table>

The table presents the size properties of \( t \)-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (\( N \)) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008.1–2013.12. \( t_{FM}(\cdot) \) denotes the \( t \)-statistic associated with the OLS estimator \( \hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]' \), which uses the traditional Fama-MacBeth standard error, \( t_{EIV}(\cdot) \) denotes the \( t \)-statistic associated with the OLS estimator \( \hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]' \), which uses the EIV-adjusted standard error in Theorem 1(ii) of Shanken (1992) and \( t(\cdot) \) denotes the \( t \)-statistic associated with the Shanken estimator \( \hat{\Gamma}^* = [\hat{\gamma}_0, \hat{\gamma}_1]' \), which uses the standard error formulae of Theorem 2. Finally, the rejection rates for the \( t \)-test associated with the bias-adjusted estimator of Shanken (1992) are based on the asymptotic distribution in part (ii) of Theorem 1. The \( t \)-statistics are compared with the critical values from a standard normal distribution.
Table 1.7
Size of t-tests in a one-factor model (Σ full, δ = 0.5)

Panel A: T = 36

<table>
<thead>
<tr>
<th>N</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(t_{FM}(\hat{\gamma}_0))</td>
<td>(t_{FM}(\hat{\gamma}_1))</td>
<td></td>
<td>(t_{EIV}(\hat{\gamma}_0))</td>
<td>(t_{EIV}(\hat{\gamma}_1))</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.126</td>
<td>0.069</td>
<td>0.020</td>
<td>0.125</td>
<td>0.070</td>
<td>0.021</td>
</tr>
<tr>
<td>500</td>
<td>0.166</td>
<td>0.097</td>
<td>0.030</td>
<td>0.181</td>
<td>0.109</td>
<td>0.034</td>
</tr>
<tr>
<td>1000</td>
<td>0.227</td>
<td>0.143</td>
<td>0.049</td>
<td>0.258</td>
<td>0.170</td>
<td>0.063</td>
</tr>
<tr>
<td>3000</td>
<td>0.393</td>
<td>0.282</td>
<td>0.123</td>
<td>0.472</td>
<td>0.354</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>(t(\hat{\gamma}_0^*))</td>
<td>(t(\hat{\gamma}_1^*))</td>
<td></td>
<td>(t(\hat{\gamma}_0^*))</td>
<td>(t(\hat{\gamma}_1^*))</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.097</td>
<td>0.045</td>
<td>0.012</td>
<td>0.094</td>
<td>0.046</td>
<td>0.011</td>
</tr>
<tr>
<td>500</td>
<td>0.094</td>
<td>0.045</td>
<td>0.009</td>
<td>0.095</td>
<td>0.045</td>
<td>0.010</td>
</tr>
<tr>
<td>1000</td>
<td>0.106</td>
<td>0.051</td>
<td>0.011</td>
<td>0.102</td>
<td>0.050</td>
<td>0.010</td>
</tr>
<tr>
<td>3000</td>
<td>0.100</td>
<td>0.051</td>
<td>0.011</td>
<td>0.100</td>
<td>0.053</td>
<td>0.011</td>
</tr>
</tbody>
</table>
Table 1.7  (Continued)
Size of $t$-tests in a one-factor model ($\Sigma$ full, $\delta = 0.5$)

Panel B: $T = 72$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$t_{FM}(\hat{\gamma}_0)$</th>
<th>$t_{FM}(\hat{\gamma}_1)$</th>
<th>$t_{EIV}(\hat{\gamma}_0)$</th>
<th>$t_{EIV}(\hat{\gamma}_1)$</th>
<th>$t(\hat{\gamma}_0)$</th>
<th>$t(\hat{\gamma}_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.113 0.062 0.014</td>
<td>0.119 0.061 0.014</td>
<td>0.112 0.062 0.014</td>
<td>0.117 0.060 0.014</td>
<td>0.094 0.046 0.010</td>
<td>0.091 0.044 0.009</td>
</tr>
<tr>
<td>500</td>
<td>0.150 0.086 0.025</td>
<td>0.165 0.096 0.029</td>
<td>0.149 0.085 0.025</td>
<td>0.164 0.096 0.029</td>
<td>0.095 0.047 0.010</td>
<td>0.094 0.050 0.011</td>
</tr>
<tr>
<td>1000</td>
<td>0.202 0.127 0.041</td>
<td>0.228 0.141 0.047</td>
<td>0.201 0.126 0.041</td>
<td>0.227 0.141 0.047</td>
<td>0.105 0.052 0.011</td>
<td>0.102 0.052 0.010</td>
</tr>
<tr>
<td>3000</td>
<td>0.353 0.246 0.102</td>
<td>0.417 0.302 0.137</td>
<td>0.352 0.244 0.100</td>
<td>0.415 0.301 0.136</td>
<td>0.102 0.052 0.012</td>
<td>0.102 0.053 0.013</td>
</tr>
</tbody>
</table>

The table presents the size properties of $t$-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks ($N$) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008.1–2013.12. $t_{FM}(\cdot)$ denotes the $t$-statistic associated with the OLS estimator $\hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]'$, which uses the traditional Fama-MacBeth standard error. $t_{EIV}(\cdot)$ denotes the $t$-statistic associated with the EIV estimator $\hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]'$, which uses the EIV-adjusted standard error in Theorem 1(ii) of Shanken (1992) and $t(\cdot)$ denotes the $t$-statistic associated with the Shanken estimator $\hat{\Gamma}^* = [\hat{\gamma}_0, \hat{\gamma}_1]'$, which uses the standard error formulae of Theorem 2. Finally, the rejection rates for the $t$-test associated with the bias-adjusted estimator of Shanken (1992) are based on the asymptotic distribution in part (ii) of Theorem 1. The $t$-statistics are compared with the critical values from a standard normal distribution.
Table 1.8
Size of t-tests in a one-factor model ($\Sigma$ full, $\delta = 0.25$)

Panel A: $T = 36$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$t_{FM}(\hat{\gamma}_0)$</th>
<th>$t_{FM}(\hat{\gamma}_1)$</th>
<th>$t_{EIV}(\hat{\gamma}_0)$</th>
<th>$t_{EIV}(\hat{\gamma}_1)$</th>
<th>$t(\hat{\gamma}_0^*)$</th>
<th>$t(\hat{\gamma}_1^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.125 0.068 0.017</td>
<td>0.124 0.068 0.018</td>
<td>0.123 0.067 0.017</td>
<td>0.123 0.067 0.017</td>
<td>0.109 0.060 0.015</td>
<td>0.112 0.059 0.015</td>
</tr>
<tr>
<td>500</td>
<td>0.163 0.095 0.034</td>
<td>0.174 0.109 0.039</td>
<td>0.162 0.095 0.033</td>
<td>0.174 0.109 0.039</td>
<td>0.115 0.062 0.018</td>
<td>0.117 0.064 0.019</td>
</tr>
<tr>
<td>1000</td>
<td>0.215 0.131 0.046</td>
<td>0.241 0.155 0.057</td>
<td>0.214 0.130 0.046</td>
<td>0.240 0.155 0.057</td>
<td>0.122 0.065 0.016</td>
<td>0.119 0.066 0.017</td>
</tr>
<tr>
<td>3000</td>
<td>0.389 0.280 0.125</td>
<td>0.459 0.343 0.164</td>
<td>0.388 0.278 0.124</td>
<td>0.458 0.341 0.163</td>
<td>0.121 0.069 0.018</td>
<td>0.124 0.068 0.018</td>
</tr>
</tbody>
</table>
### Table 1.8 (Continued)
Size of \( t \)-tests in a one-factor model (\( \Sigma \) full, \( \delta = 0.25 \))

**Panel B: \( T = 72 \)**

<table>
<thead>
<tr>
<th>( N )</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t_{FM}(\hat{\gamma}_0) )</td>
<td>( t_{FM}(\hat{\gamma}_1) )</td>
<td>( t_{EIV}(\hat{\gamma}_0) )</td>
<td>( t_{EIV}(\hat{\gamma}_1) )</td>
<td>( t(\hat{\gamma}_0) )</td>
<td>( t(\hat{\gamma}_1) )</td>
</tr>
<tr>
<td>100</td>
<td>0.119</td>
<td>0.060</td>
<td>0.014</td>
<td>0.123</td>
<td>0.066</td>
<td>0.015</td>
</tr>
<tr>
<td>500</td>
<td>0.155</td>
<td>0.091</td>
<td>0.025</td>
<td>0.163</td>
<td>0.098</td>
<td>0.030</td>
</tr>
<tr>
<td>1000</td>
<td>0.199</td>
<td>0.126</td>
<td>0.042</td>
<td>0.222</td>
<td>0.138</td>
<td>0.050</td>
</tr>
<tr>
<td>3000</td>
<td>0.334</td>
<td>0.229</td>
<td>0.092</td>
<td>0.390</td>
<td>0.280</td>
<td>0.124</td>
</tr>
<tr>
<td>100</td>
<td>0.117</td>
<td>0.059</td>
<td>0.014</td>
<td>0.122</td>
<td>0.065</td>
<td>0.015</td>
</tr>
<tr>
<td>500</td>
<td>0.155</td>
<td>0.090</td>
<td>0.025</td>
<td>0.162</td>
<td>0.098</td>
<td>0.030</td>
</tr>
<tr>
<td>1000</td>
<td>0.198</td>
<td>0.125</td>
<td>0.042</td>
<td>0.222</td>
<td>0.138</td>
<td>0.049</td>
</tr>
<tr>
<td>3000</td>
<td>0.333</td>
<td>0.228</td>
<td>0.091</td>
<td>0.388</td>
<td>0.278</td>
<td>0.123</td>
</tr>
<tr>
<td>100</td>
<td>0.108</td>
<td>0.057</td>
<td>0.012</td>
<td>0.110</td>
<td>0.059</td>
<td>0.015</td>
</tr>
<tr>
<td>500</td>
<td>0.114</td>
<td>0.062</td>
<td>0.015</td>
<td>0.119</td>
<td>0.065</td>
<td>0.015</td>
</tr>
<tr>
<td>1000</td>
<td>0.121</td>
<td>0.063</td>
<td>0.015</td>
<td>0.122</td>
<td>0.067</td>
<td>0.016</td>
</tr>
<tr>
<td>3000</td>
<td>0.111</td>
<td>0.057</td>
<td>0.012</td>
<td>0.114</td>
<td>0.058</td>
<td>0.014</td>
</tr>
</tbody>
</table>

The table presents the size properties of \( t \)-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (\( N \)) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008.1–2013.12. \( t_{FM}(\cdot) \) denotes the \( t \)-statistic associated with the OLS estimator \( \hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]' \), which uses the traditional Fama-MacBeth standard error, \( t_{EIV}(\cdot) \) denotes the \( t \)-statistic associated with the OLS estimator \( \hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]' \), which uses the EIV-adjusted standard error in Theorem 1(ii) of Shanken (1992) and \( t(\cdot) \) denotes the \( t \)-statistic associated with the Shanken estimator \( \hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_1]' \), which uses the standard error formulae of Theorem 2. Finally, the rejection rates for the \( t \)-test associated with the bias-adjusted estimator of Shanken (1992) are based on the asymptotic distribution in part (ii) of Theorem 1. The \( t \)-statistics are compared with the critical values from a standard normal distribution.
### Table 1.9
Rejection rates of the specification test in a one-factor model \((T = 36)\)

<table>
<thead>
<tr>
<th>(N)</th>
<th>(10%)</th>
<th>(5%)</th>
<th>(1%)</th>
<th>(10%)</th>
<th>(5%)</th>
<th>(1%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: (\Sigma) spherical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.103</td>
<td>0.049</td>
<td>0.009</td>
<td>0.882</td>
<td>0.823</td>
<td>0.675</td>
</tr>
<tr>
<td>500</td>
<td>0.098</td>
<td>0.050</td>
<td>0.009</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
</tr>
<tr>
<td>1000</td>
<td>0.101</td>
<td>0.052</td>
<td>0.011</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>3000</td>
<td>0.101</td>
<td>0.050</td>
<td>0.009</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Panel B: (\Sigma) diagonal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.085</td>
<td>0.037</td>
<td>0.010</td>
<td>0.634</td>
<td>0.529</td>
<td>0.340</td>
</tr>
<tr>
<td>500</td>
<td>0.093</td>
<td>0.046</td>
<td>0.010</td>
<td>0.983</td>
<td>0.967</td>
<td>0.894</td>
</tr>
<tr>
<td>1000</td>
<td>0.099</td>
<td>0.050</td>
<td>0.009</td>
<td>1.000</td>
<td>1.000</td>
<td>0.996</td>
</tr>
<tr>
<td>3000</td>
<td>0.097</td>
<td>0.046</td>
<td>0.011</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Panel C: (\Sigma) full ((\delta = 0.5))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.084</td>
<td>0.040</td>
<td>0.011</td>
<td>0.639</td>
<td>0.534</td>
<td>0.332</td>
</tr>
<tr>
<td>500</td>
<td>0.101</td>
<td>0.050</td>
<td>0.012</td>
<td>0.982</td>
<td>0.965</td>
<td>0.887</td>
</tr>
<tr>
<td>1000</td>
<td>0.095</td>
<td>0.049</td>
<td>0.011</td>
<td>1.000</td>
<td>1.000</td>
<td>0.997</td>
</tr>
<tr>
<td>3000</td>
<td>0.108</td>
<td>0.056</td>
<td>0.011</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Panel D: (\Sigma) full ((\delta = 0.25))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.110</td>
<td>0.060</td>
<td>0.021</td>
<td>0.621</td>
<td>0.522</td>
<td>0.336</td>
</tr>
<tr>
<td>500</td>
<td>0.145</td>
<td>0.084</td>
<td>0.029</td>
<td>0.977</td>
<td>0.956</td>
<td>0.874</td>
</tr>
<tr>
<td>1000</td>
<td>0.145</td>
<td>0.088</td>
<td>0.029</td>
<td>1.000</td>
<td>0.999</td>
<td>0.993</td>
</tr>
<tr>
<td>3000</td>
<td>0.146</td>
<td>0.087</td>
<td>0.030</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The table presents the size and power properties of our test \(S^*\) of correct model specification. The null hypothesis is that the model is correctly specified. The alternative hypothesis is that the model is misspecified. The results are reported for different levels of significance (10\%, 5\%, and 1\%) and for different values of the number of stocks \((N)\) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2010:12 \((T = 36)\). Finally, the rejection rates for the specification test are based on the asymptotic distribution in Theorem 4. The rejection rates of the test are based on two-sided \(p\)-values.
Table 1.10
Rejection rates of the specification test in a one-factor model \((T = 72)\)

<table>
<thead>
<tr>
<th></th>
<th>Size</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>(N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.095</td>
<td>0.045</td>
<td>0.009</td>
<td>0.929</td>
<td>0.891</td>
<td>0.781</td>
</tr>
<tr>
<td>500</td>
<td>0.101</td>
<td>0.047</td>
<td>0.009</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>1000</td>
<td>0.104</td>
<td>0.055</td>
<td>0.010</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>3000</td>
<td>0.099</td>
<td>0.048</td>
<td>0.010</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Panel A: (\Sigma) spherical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.085</td>
<td>0.041</td>
<td>0.010</td>
<td>0.771</td>
<td>0.676</td>
<td>0.480</td>
</tr>
<tr>
<td>500</td>
<td>0.098</td>
<td>0.046</td>
<td>0.010</td>
<td>1.000</td>
<td>1.000</td>
<td>0.997</td>
</tr>
<tr>
<td>1000</td>
<td>0.101</td>
<td>0.049</td>
<td>0.012</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>3000</td>
<td>0.102</td>
<td>0.051</td>
<td>0.011</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Panel B: (\Sigma) diagonal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.085</td>
<td>0.039</td>
<td>0.011</td>
<td>0.770</td>
<td>0.681</td>
<td>0.482</td>
</tr>
<tr>
<td>500</td>
<td>0.092</td>
<td>0.046</td>
<td>0.009</td>
<td>1.000</td>
<td>0.999</td>
<td>0.996</td>
</tr>
<tr>
<td>1000</td>
<td>0.094</td>
<td>0.049</td>
<td>0.010</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>3000</td>
<td>0.097</td>
<td>0.047</td>
<td>0.010</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Panel C: (\Sigma) full ((\delta = 0.5))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.120</td>
<td>0.063</td>
<td>0.023</td>
<td>0.749</td>
<td>0.660</td>
<td>0.470</td>
</tr>
<tr>
<td>500</td>
<td>0.140</td>
<td>0.083</td>
<td>0.029</td>
<td>1.000</td>
<td>0.999</td>
<td>0.994</td>
</tr>
<tr>
<td>1000</td>
<td>0.149</td>
<td>0.086</td>
<td>0.030</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>3000</td>
<td>0.153</td>
<td>0.093</td>
<td>0.034</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Panel D: (\Sigma) full ((\delta = 0.25))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table presents the size and power properties of our test \(S^*\) of correct model specification presented in Theorem 3. The null hypothesis is that the model is correctly specified. The alternative hypothesis is that the model is misspecified. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks \((N)\) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008.1–2013.12 \((T = 72)\). Finally, the rejection rates for the specification test are based on the asymptotic distribution in Theorem 4. The rejection rates of the test are based on two-sided \(p\)-values.
1.7. Empirical Analysis

In this section, we show empirically that the results obtained with our fixed-$T$ and large-$N$ methodology can differ substantially from the results obtained with traditional large-$T$ and fixed-$N$ methods. Using a large number of individual equity returns from CRSP, we estimate and test FF5 and an extension of this model that includes the non-traded liquidity factor of Pástor and Stambaugh (2003). The demonstrated empirical success of FF5 in explaining the cross-sectional variation in expected equity returns is what motivates our interest in this model. In the second part of this section, we analyze the extent to which firm characteristics contribute to explaining the cross-section of expected equity returns.

The risk and characteristic premia estimators, their confidence intervals, and the various test statistics employed are based on our theoretical analysis in Sections 1.3 and 1.4.

1.7.1 Data

The monthly data on the traded factors of FF5 is available from Kenneth French's website and the non-traded liquidity factor of Pástor and Stambaugh (2003) is taken from Lubos Pástor's website. As for the test assets, we download monthly stock returns (from January 1966 to December 2013) from CRSP and apply two filters in the selection of stocks. First, we require that a stock has a Standard Industry Classification (SIC) code. (We adopt the 49 industry classifications listed on Kenneth French's website.) Second, we keep a stock in our sample only for the months in which its price is at least three dollars. The resulting dataset consists of 3,435 individual stocks. We perform the empirical analysis using balanced panels over fixed-time windows of three and 10 years (that is, $T = 36$ and 120), respectively. We obtain time series of estimated risk premia and test statistics by shifting the time window month by month over the 1966-2013 period. After filtering the data, we obtain an average number (over the overlapping time windows) of approximately 2,800 stocks when $T = 36$ and 1,200 stocks when $T = 120$.

1.7.2 Specification testing

For the analysis with traded factors only, we report the p-values of our specification test, $S^*$, as well as the p-values of two alternative tests, the Gibbons et al. (1989a) (GRS) and Gungor and Luger (2016) (GL) tests. It should be noted that GRS requires $N$ to be fixed, while the Gungor and Luger (2016) test is valid for any $N$ and $T$. All three tests are tests of the same null hypothesis; that is, $H_0 : c_i = 0$, for every $i = 1, 2, \ldots$.

(i) $S^*$ test

We first assess the performance of FF5 using $S^*$.

The black line in Figure 1.1 denotes the time series of p-values associated with our test statistic $S^*$ for time windows of three years (top panel) and 10 years (bottom panel), respectively. When the black line is below the 5% significance level (dotted red line), we reject FF5. Figure 1.1 shows that based on our test, we reject the validity of FF5 about 60% of the times when $T = 36$. As expected, the rejection of FF5 happens more frequently when we increase the time window from $T = 36$ to $T = 120$. The rejection of FF5 occurs in about 95% of the cases when the latter scenario is considered. Given the availability of a time series of p-values, one could cast the analysis in a multiple testing framework, as suggested by Barras et al. (2010). Applying

\footnote{Section 1.7.5 reports further empirical results for FF5, as well as results for CAPM and FF3.}

\footnote{Several studies (see Konak et al. (2018), Kelly et al. (2018), and Huang et al. (2018), among others) have shown that these five factors are highly correlated with appropriately constructed latent factors such as the first five principal components, and variations of, from the data.}

\footnote{The five traded factors of FF5 are the market excess return ($mkt$), the return difference between portfolios of stocks with small and large market capitalizations ($smb$), the return difference between portfolios of stocks with high and low book-to-market ratios ($hml$), the average return on two robust operating profitability portfolios minus the average return on two weak operating profitability portfolios ($rmw$), and the average return on two conservative investment portfolios minus the average return on two aggressive investment portfolios ($cma$).}
their methodology to $S^*$, we reject the null of correct model specification in 61% and 95% of the cases for $T = 36$ and $T = 120$, respectively. In Figure 1.2, we perform the same analysis for the liquidity-augmented FF5.

This variant of FF5 turns out to be strongly rejected, even when $T = 36$. The rejection frequencies are approximately equal to 82% and 92% for $T = 36$ and 120, respectively. Overall, the frequent and strong rejections of FF5 justify our use of confidence intervals that are robust to model misspecification in the subsequent analysis.

(ii) GRS and GL tests

Figure 1.3 reports the GRS $p$-values (blue line) as well as the GL $p$-values (green line).

Unlike ours, these two tests are only applicable to beta-pricing models with traded factors. As a consequence, we consider only FF5 here. Since GRS is a GLS-based test, effectively, it is implementable only when $N$ is substantially smaller than $T$. Therefore, we construct 25 equally weighted portfolio returns from our individual stock returns and analyze the performance of these two tests, using this smaller asset set.\footnote{The results in Figure 1.3 are obtained by randomly assigning the various stocks to 25 portfolios. For instance, when $T = 36$, each of the 25 portfolios contains approximately 110 randomly selected stocks. We also experimented with 25 portfolios formed on CAPM betas. The results of the analysis are qualitatively similar to those in Figure 1.3.}

Differently from our large-$N$ test, we are much less likely to reject FF5 based on the GRS test. When considering time windows of $T = 36$, the average rejection rate for FF5 is only about 30%. In addition, FF5 is rejected almost always when $T = 120$. We obtain similar results when using the GL test, although it is harder to quantify the rejection rates in this case because the GL test often leads to an inconclusive outcome. Based on the GL test, FF5 is not rejected in about 70% of the cases when $T = 36$, but the test is inconclusive about 29% of the time. Moreover, FF5 is not rejected in only about 18% of the cases when $T = 120$, but the test is inconclusive about 76% of the time. The main message here is that using our test can lead to qualitatively different conclusions relative to existing methods.

1.7.3 Risk premia estimates

Since our test, $S^*$, points to serious misspecification of the risk-return relation, in this section we perform parameter testing by means of standard errors that are robust to model misspecification. Specifically, we use the large-$N$ standard errors derived in Theorems 1.5. To highlight the differences between our approach and standard large-$T$ methods, we also consider the OLS CSR estimator and the corresponding large-$T$ standard errors from Theorem 1(ii) in Shanken (1992). For traded factors, we also report the rolling sample mean of the factor returns, which is a valid risk premium estimator when $T$ is large. In contrast, when considering non-traded factors such as liquidity, we consider the rolling sample mean of the corresponding mimicking portfolio return. (See footnote 16 above.)

(i) FF5

Based on a time window of three years, the top panel of Figure 1.4 presents the rolling-window estimates of the risk premium on the market factor and the corresponding 95% confidence intervals. (The results for the other four factors are reported in Section 1.7.5.)

In the figure, the bold black line and the dotted red line refer to the Shanken (1992) and OLS CSR estimators, respectively. The grey band represents the large-$N$ 95% confidence intervals that are robust to model misspecification, whereas the striped orange band is for the large-$T$ confidence intervals. Finally, the dashed black line displays the rolling factor sample mean. Noticeably, the large-$T$ confidence intervals include the zero value in about 60% of the cases. In contrast, our large-$N$ confidence intervals include the zero value...
only about 30% of the time. Not surprisingly, the bottom panel of Figure 1.4 (T = 120 case) shows that the risk premia estimates are smoother than in the T = 36 scenario. However, the large-T confidence intervals are still larger than the corresponding large-N confidence intervals, and they indicate that the OLS CSR and the Shanken (1992) estimates are statistically significant 30% and 80% of the time, respectively. The large-N estimates appear to be systematically larger than the corresponding large-T estimates for most dates, especially for the longer time window. This is the result of the systematic (negative) bias that affects the OLS CSR estimator when N is large. The relationship between the large-N and the rolling sample mean estimates (the latter are based on windows of T = 36 and T = 120 monthly data, respectively) is less stable. The two sets of estimates exhibit a correlation of about 0.5 when T = 36 and 0.7 when T = 120.

Figure 1.4 shows that the large-T approach supports the hypothesis of constant risk premia, whereas our large-N results point toward a significant time variation in risk premia. Therefore, it seems plausible to interpret $\hat{\Gamma}^*$ as the estimator of the local average, over T periods, of the (time-varying) risk premia, $\tilde{\Gamma}$, as explained in Section 1.3.2.

The top panel of Figure 1.5 reports the Shanken (1992) large-N estimates, expressed in terms of a single line (black line) and in terms of local averages (horizontal bars of length T = 36, blue lines), with the corresponding 95% confidence intervals for these local averages based on the large-N standard errors of Theorem 1.5 (grey band).

Figure 1.5 about here

The local average estimates appear to be significantly different from each other in most cases, which is a clear symptom of time variation in risk premia. In the same panel, we also report the rolling sample mean (over fixed windows of six months of daily data) of the market excess return (dashed dotted red line) and the corresponding 95% confidence interval (orange band). As our results indicate, although the latter is a suitable (nonparametric) estimator of the time-varying risk premium, it requires a large number of observations (over a short time window) to produce sufficiently narrow confidence intervals. The correlation between the Shanken (1992) large-N estimator and the six-month rolling sample mean based on daily data is positive but small (the sample correlation coefficient is 0.14). In addition, differently from the Shanken (1992) large-N estimator, the six-month rolling sample mean based on daily data appears to be very noisy.

Given the pronounced time variation in risk premia, the bottom panel of Figure 1.5 reports our novel estimator $\hat{\gamma}_{t-1}$ (black line), formally defined in Eq. (1.50), and the corresponding 95% confidence interval (grey band). Although noisier than $\hat{\Gamma}^*$, the $\hat{\gamma}_{t-1}$ estimates are still statistically significant about 50% of the time. As the figure indicates, there is a sharp increase in risk premia volatility in correspondence and in the aftermath of major economic and financial crises and episodes such as the Black Monday of October 1987 and the US savings and loan crisis of the 80s and 90s. Our empirical findings on risk premia counter-cyclicity confirm the results in Gagliardini et al. (2016) and corroborate the predictions of many theoretical models. (See the discussion in Section 4.3 of Gagliardini et al. (2016).)

(ii) Liquidity-augmented FF5

As for the liquidity-augmented FF5, Figure 1.6 presents the estimated liquidity risk premium in the time-invariant setting.

Figure 1.6 about here

The estimated liquidity risk premia in Figure 1.6 are positive 55% and 37% of the time for T = 36 and T = 120, respectively. However, the risk premia estimates are statistically significant at the 5% level only in the 21% and 32% of the cases, for T = 36 and T = 120, respectively. In the same figure, we also report the OLS CSR estimator and the corresponding mimicking portfolio rolling sample mean (based on windows of T = 36 and T = 120 monthly data). The OLS CSR estimates in this case are not too far from the Shanken (1992) estimates. In contrast, the rolling mimicking portfolio sample means are now only mildly positively correlated with the $\hat{\Gamma}^*$ estimates. (The correlation coefficients are 0.15 and 0.27 for T = 36 and T = 120, respectively.)

As in the traded factor case, Figure 1.7 indicates that the time variation in risk premia is pronounced.
Based on the top panel of Figure 1.7, the correlation between the mimicking portfolio six-month rolling sample mean and the Shanken (1992) large-\(N\) estimates is about 0.19. Similar to the FF5 case, the large-\(N\) estimator seems to exhibit a higher precision. Looking at the bottom panel of Figure 1.7, the risk premia counter-cyclicity emerges again, especially around major economic and financial downturns.

Finally, Table 1.11 reports the percentage difference (averaged over rolling time windows of size \(T = 36\) and \(T = 120\), respectively) between the Shanken (1992) estimator, \(\hat{\Gamma}^*\), and the OLS CSR estimator, \(\hat{\Gamma}\), for the various risk premia in CAPM, FF3, and FF5.

Panel A shows that the percentage difference between estimators is quite large (about 64% when \(T = 36\) and 27% when \(T = 120\)). As for FF3 in Panel B, the discrepancy between the two estimators is sizeable for \(hml\), ranging from 31% to 52%, and less pronounced for \(mkt\) and \(smb\). Moreover, relative to FF5, Panel C indicates that the percentage difference between the two estimators is relatively large for \(cma\), ranging from 33% to about 43%. Finally, sizeable differences between the two estimators exist for \(iq\), especially in Panel A.

In summary, we often find significant differences between the results based on our large-\(N\) approach and the results based on conventional large-\(T\) methods. The difference mainly stems from the smaller standard errors of the Shanken (1992) estimator relative to the OLS CSR estimator and the nontrivial bias correction induced by the Shanken (1992) estimator when \(N\) is large. These differences are even more pronounced when comparing the results based on the Shanken (1992) estimator with those based on the rolling sample mean estimator. Finally, the estimated risk premium on the (non-traded) liquidity factor of Pástor and Stambaugh (2003) is often found to be statistically insignificant.

### 1.7.4 Characteristics

In this section, for ease of comparison with Chordia et al. (2015), we use balanced panel data from January 1980 to December 2015.\(^{49}\) In the dataset we use, an average of 3,071 firms have return data in a particular month. Consistent with Daniel and Titman (1997) and Chordia et al. (2015), among others, we focus on five firm characteristics that have often been found to be related to the cross-section of expected returns: book-to-market ratio (\(B/M\)), asset growth (\(ASSGR\)), operating profitability (\(OPERPROF\)), market capitalization (\(MCAPIT\)), and six-month momentum (\(MOM6\)). As it is common in this literature, we cross-sectionally standardize the characteristics.

In the interest of space, we focus only on the \(T = 36\) case. For each time window, we compute the average of the characteristics. In the first pass, we obtain beta estimates for CAPM, FF3, and FF5. We then estimate the ex post risk and characteristic premia using our second-pass CSR estimator in Eq. (1.78). Figure 1.8 reports the time series of the characteristic premia estimates, \(\delta^*\), and the 95% confidence intervals for each model.

Although the confidence intervals tend to widen when moving from CAPM to FF5, averaging across the three models, the estimated \(B/M\) premium is positive about 59% of the time, but it is only statistically significant at the 5% level in about 3% of the cases. The estimated \(ASSGR\) premium is almost always negative (in 81% of the cases) and significantly so about 16% of the time, whereas the estimated \(OPERPROF\) premium is positive in about 32% of the cases and statistically significant only about 19% of the time. For \(MCAPIT\), the estimated premium is positive 32% of the time and statistically significant in about 12% of the cases, while the \(MOM6\) estimate is almost always positive (99.6% of the time) and significant in 86% of the cases.

We now analyze the joint importance of the five characteristics in explaining deviations from correct model specification; that is, we assess whether the expected returns on individual stocks represent a compensation

---

\(^{49}\)We thank Alberto Martín-Utrera for sharing his data with us and refer to DeMiguel et al. (2018) for data details.
for risk or firm characteristics. We consider two alternative approaches. First, we conduct formal tests of the two hypotheses, $H_0 : \gamma_l^p = 0_K$ and $H_0 : \delta = 0_K$, using the asymptotic distribution theory in Theorems 1.6 and 1.7. The results are in Panel A of Table 1.12. The $F$-tests indicate that the characteristic premia estimates are statistically significant at any conventional level, with the average $F$-test (over rolling windows of size $T = 36$) for the null hypothesis $H_0 : \delta = 0_K$ being equal to 1278.60, 1108.41, and 927.04 for CAPM, FF3, and FF5, respectively. In contrast, the average $F$-test for the null hypothesis $H_0 : \gamma_l^p = 0_K$ equals 12.45, 17.19, and 57.18 for CAPM, FF3, and FF5, respectively, with rejections rates, in the order, of 25.70%, 25.90%, and 37.90%.

Next, Panel B of Table 1.12 presents the cross-sectional variance contribution of betas and characteristics to the overall cross-sectional dispersion in the (sample) average returns, $\bar{R}$. Chordia et al. (2015) suggest to consider the ratios of the (cross-sectional) variance of the beta component (betas times the factor risk premia) and of the characteristics component (characteristics times the characteristic premia), with respect to the overall (cross-sectional) variance of average returns. However, since the beta and characteristics components are not orthogonal cross-sectionally, this can lead to a percentage of the cross-sectional variance explained by the betas and by the characteristics that is jointly greater than 100%. In addition, the estimated pricing errors based on our bias-adjusted estimator are not necessarily orthogonal to the regressors of the CSR, thus complicating the interpretation even further.

We modify the approach of Chordia et al. (2015) as follows. From the estimated CSR, we have $\bar{R} = \bar{X}\hat{\Gamma}^* + C\hat{\delta}^* + \hat{\eta}^p$, where $\hat{\eta}^p$ are the sample counterparts of $\eta^p$. Consider the orthogonalization of the estimated pricing errors, $\tilde{\eta}^p$,

$$\tilde{\bar{R}} = \tilde{X}\hat{\Gamma}^* + C\tilde{\delta}^* + P_{\tilde{Z}}\tilde{\eta}^p = \tilde{X}\hat{\Gamma}^* + C\tilde{\delta}^* + P_{\tilde{Z}}\tilde{\eta}^p,$$

where $P_{\tilde{Z}} = \hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'$ with $\hat{Z} = [\hat{X}, C]$, and $I_N$ denotes the identity matrix of order $N$. By construction, the orthogonalized estimated pricing errors, $\tilde{\eta}^p = (I_N - P_{\tilde{Z}})\tilde{\eta}^p$, satisfy $\tilde{Z}'\tilde{\eta}^p = 0_{K+K_c+1}$. Setting $P_C = C(C'C)^{-1}C'$, rewrite the estimated CSR as

$$\tilde{\bar{R}} = (\tilde{X}\hat{\Gamma}^* + P_{\tilde{Z}}\tilde{\eta}^p) + C\tilde{\delta}^* + \tilde{\eta}^p$$

$$= [(I_N - P_C)(\tilde{X}\hat{\Gamma}^* + P_{\tilde{Z}}\tilde{\eta}^p)] + [P_C(\tilde{X}\hat{\Gamma}^* + P_{\tilde{Z}}\tilde{\eta}^p)] + C\tilde{\delta}^* + \tilde{\eta}^p$$

$$\equiv \tilde{R}_{1,C} + \tilde{R}_C + \tilde{\eta}^p,$$

where $\tilde{R}_{1,C} = (I_N - P_C)(\tilde{X}\hat{\Gamma}^* + P_{\tilde{Z}}\tilde{\eta}^p)$ is the component of the average returns that is explained only by the estimated betas, and thus (perfectly) uncorrelated with $C$ in sample, and $\tilde{R}_C = P_C(\tilde{X}\hat{\Gamma}^* + P_{\tilde{Z}}\tilde{\eta}^p) + C\tilde{\delta}^*$ is the component of the average returns due to $C$ only. Since $\tilde{R}_{1,C}$ and $\tilde{R}_C$ are orthogonal to each other and to $\tilde{\eta}^p$, the sample variance of the average returns equals the sum of the sample variances of the betas component, of the characteristics component, and of the orthogonalized pricing errors, that is,

$$S^2_R = \frac{\tilde{R}'\tilde{R}}{N} - \left(\frac{1_N'\tilde{R}}{N}\right)^2$$

$$= \frac{\tilde{R}'_{1,C}\tilde{R}_{1,C}}{N} - \left(\frac{1_N'\tilde{R}_{1,C}}{N}\right)^2 + \frac{\tilde{R}'_C\tilde{R}_C}{N} - \left(\frac{1_N'\tilde{R}_C}{N}\right)^2 + \frac{\tilde{\eta}'^p\tilde{\eta}^p}{N}$$

$$\equiv S^2_{R_{1,C}} + S^2_R + S^2_{\tilde{\eta}^p}.$$ 

Panel B of Table 1.12 reports the average, over rolling windows of size $T = 36$, of the variance ratios $100 \times S^2_{R_{1,C}}/S^2_R$ and $100 \times S^2_{R_{1,C}}/S^2_R$.

Table 1.12 about here

---

50 This problem is acknowledged, although not solved, in Chordia et al. (2015).
The results are largely supportive of our findings based on the $F$-tests; that is, characteristics overwhelmingly dominate the cross-sectional variation in average individual stock returns. Averaging across the three beta-pricing models, the characteristic variance ratio, $100 \times \frac{S_{R,c}^2}{S_{R}^2}$, is about 76%, whereas the beta variance ratio, $100 \times \frac{S_{R,c}^2}{S_{R}^2}$, is about 2.8%. The rest (about 21.5%) represents the unexplained portion of the average return cross-sectional variance. Overall, our empirical findings support the conclusions of Chordia et al. (2015), who argue that regardless of the beta-pricing model and whether the premia are allowed to be time-varying, it is mainly the characteristics that contribute to the cross-sectional variation in expected stock returns.


This section contains several figures (Figures 1.10-1.34) for the CAPM and the Fama and French (1993) three-factor model (FF3). We also report further results for the Fama and French (2015) five-factor model (FF5). We first consider specification testing. Then, we present the risk premia estimates for these three beta-pricing models.

---

51 Confidence intervals for these variance ratios could be computed based on our asymptotic results. The details are available upon request.
Figure 1.1
Specification testing for the Fama and French (2015) five-factor model
The figure presents the time series of $p$-values (black line) of $S^*$ for FF5. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the test. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.2
Specification testing for the liquidity-augmented Fama and French (2015) five-factor model
The figure presents the time series of p-values (black line) of $S^*$ for the liquidity-augmented FF5. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the test. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s and Lubos Pástor’s websites from January 1966 to December 2013.
Figure 1.3
Specification testing for the Fama and French (2015) five-factor model using the Gibbons et al. (1989a) and Gungor and Luger (2016) tests

The figure presents the time series of $p$-values of the GRS (blue line) and GL (green line) tests for FF5. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the tests. The grey bars are for the periods in which the GL test is inconclusive. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.4
Estimates and confidence intervals for the market risk premium

The figure presents the estimates and the associated confidence intervals for the market risk premium from the Fama and French (2015) five-factor model. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-$N$ standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-$T$ standard errors. Finally, the dashed black line is for the rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French's website from January 1966 to December 2013.
Figure 1.5
Estimates and confidence intervals for the time-varying market risk premium

Figure 1.6
Estimates and confidence intervals for the liquidity risk premium

The figure presents the estimates and the associated confidence intervals for the liquidity risk premium from the liquidity-augmented Fama and French (2015) five-factor model. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-$N$ standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-$T$ standard errors. Finally, the dashed black line is for the mimicking portfolio rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s and Lubos Pastor’s websites from January 1966 to December 2013.
Figure 1.7
Estimates and confidence intervals for the time-varying liquidity risk premium

Figure 1.8
Estimates and confidence intervals for the characteristic premia
The figure presents estimates (blue line) of the characteristic premia on the book-to-market ratio ($B/M$), asset growth ($ASSGR$), operating profitability ($OPERPROF$), market capitalization ($MCAPIT$), and sixmonth momentum ($MOM6$), and the associated confidence intervals based on Theorem 7 (light blue band), for the Fama and French (2015) five-factor model. The data is from DeMiguel et al. (2018) and Kenneth Frenchs website (from January 1980 to December 2015).
Figure 1.9
Specification testing for CAPM
The figure presents the time series of p-values (black line) of $S^*$ for CAPM. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the test. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.10
Specification testing for the liquidity-augmented CAPM
The figure presents the time series of $p$-values (black line) of $S^*$ for the liquidity-augmented CAPM. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the test. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s and Lubos Pástor’s websites from January 1966 to December 2013.
Figure 1.11
Specification testing for CAPM using the Gibbons et al. (1989a) and Gungor and Luger (2016) tests

The figure presents the time series of $p$-values of the GRS (blue line) and GL (green line) tests for CAPM. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the tests. The grey bars are for the periods in which the GL test is inconclusive. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.12
Estimates and confidence intervals for the market risk premium (CAPM)
The figure presents the estimates and the associated confidence intervals for the market \( \text{mkt} \) risk premium from CAPM. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-\( N \) standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-\( T \) standard errors. Finally, the dashed black line is for the rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.13
Estimates and confidence intervals for the time-varying market risk premium (CAPM)

Figure 1.14
Estimates and confidence intervals for the liquidity risk premium (liquidity-augmented CAPM)

The figure presents the estimates and the associated confidence intervals for the liquidity risk premium from the liquidity-augmented CAPM. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-$N$ standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-$T$ standard errors. Finally, the dashed black line is for the mimicking portfolio rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s and Lubos Pástor’s websites from January 1966 to December 2013.
Figure 1.15
Estimates and confidence intervals for the time-varying liquidity risk premium (liquidity-augmented CAPM)

The figure presents the estimates and the associated confidence intervals for the time-varying liquidity risk premium from the liquidity-augmented CAPM based on our large-$N$ methodology. The top panel reports the Shanken (1992) large-$N$ estimates, expressed in terms of a single line (black line) and in terms of horizontal bars of length $T = 36$ observations (blue line), with the corresponding 95% confidence intervals based on the large-$N$ standard errors of Theorem 1.5 (grey band). We also report the rolling sample mean (using fixed rolling windows of six months) of the corresponding mimicking portfolio excess return (dashed dotted red line) and the corresponding 95% confidence interval (orange band). The bottom panel reports the modified Shanken (1992) estimator (black line) and the corresponding 95% confidence interval (grey band) based on the large-$N$ standard errors of part (ii) of Theorem 1.3.
Figure 1.16
Specification testing for the Fama and French (1993) three-factor model
The figure presents the time series of p-values (black line) of $S^*$ for FF3. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the test. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.17
Specification testing for the liquidity-augmented Fama and French (1993) three-factor model
The figure presents the time series of p-values (black line) of $S^*$ for the liquidity-augmented FF3. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the test. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s and Lubos Pástor’s websites from January 1966 to December 2013.
Figure 1.18
Specification testing for the Fama and French (1993) three-factor model using the Gibbons et al. (1989a) and Gungor and Luger (2016) tests

The figure presents the time series of p-values of the GRS (blue line) and GL (green line) tests FF3. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the tests. The grey bars are for the periods in which the GL test is inconclusive. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.19
Estimates and confidence intervals for the market risk premium (FF3)
The figure presents the estimates and the associated confidence intervals for the market \( (\text{mkt}) \) risk premium from FF3. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-\( N \) standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-\( T \) standard errors. Finally, the dashed black line is for the rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French's website from January 1966 to December 2013.
Figure 1.20
Estimates and confidence intervals for the size premium (FF3)
The figure presents the estimates and the associated confidence intervals for the small-minus-big ($\text{smb}$) premium from FF3. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-$N$ standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-$T$ standard errors. Finally, the dashed black line is for the rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.21
Estimates and confidence intervals for the value premium (FF3)

The figure presents the estimates and the associated confidence intervals for the high-minus-low (hml) premium from FF3. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-N standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-T standard errors. Finally, the dashed black line is for the rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.22
Estimates and confidence intervals for the liquidity risk premium (liquidity-augmented FF3)

The figure presents the estimates and the associated confidence intervals for the liquidity risk premium from the liquidity-augmented FF3. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-N standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-T standard errors. Finally, the dashed black line is for the mimicking portfolio rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s and Lubos Pastor’s websites from January 1966 to December 2013.
Estimates and confidence intervals for the time-varying market risk premium (FF3)

Estimates and confidence intervals for the time-varying size premium (FF3)

Figure 1.25
Estimates and confidence intervals for the time-varying value premium (FF3)

Figure 1.26
Estimates and confidence intervals for the time-varying liquidity risk premium (liquidity-augmented FF3)

The figure presents the estimates and the associated confidence intervals for the time-varying liquidity risk premium from the liquidity-augmented FF3 based on our large-N methodology. The top panel reports the Shanken (1992) large-N estimates, expressed in terms of a single line (black line) and in terms of horizontal bars of length \( T = 36 \) observations (blue line), with the corresponding 95% confidence intervals based on the large-N standard errors of Theorem 1.5 (grey band). We also report the rolling sample mean (using fixed rolling windows of six months) of the corresponding mimicking portfolio excess return (dashed dotted red line) and the corresponding 95% confidence interval (orange band). The bottom panel reports the modified Shanken (1992) estimator (black line) and the corresponding 95% confidence interval (grey band) based on the large-N standard errors of part (ii) of Theorem 1.3.

Figure 1.27
Estimates and confidence intervals for the size premium (FF5)
The figure presents the estimates and the associated confidence intervals for the small-minus-big ($smb$) premium from FF5. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-$N$ standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-$T$ standard errors. Finally, the dashed black line is for the rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.28
Estimates and confidence intervals for the value premium (FF5)
The figure presents the estimates and the associated confidence intervals for the high-minus-low (hml) premium from FF5. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-N standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-T standard errors. Finally, the dashed black line is for the rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.29
Estimates and confidence intervals for the profitability premium (FF5)

The figure presents the estimates and the associated confidence intervals for the robust-minus-weak (rmw) premium from FF5. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-$N$ standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-$T$ standard errors. Finally, the dashed black line is for the rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.30
Estimates and confidence intervals for the investment premium (FF5)
The figure presents the estimates and the associated confidence intervals for the conservative-minus-aggressive (cma) premium from FF5. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-N standard errors of Theorem 1.5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-T standard errors. Finally, the dashed black line is for the rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s website from January 1966 to December 2013.
Figure 1.31
Estimates and confidence intervals for the time-varying size premium (FF5)
Figure 1.32
Estimates and confidence intervals for the time-varying value premium (FF5)

Figure 1.33
Estimates and confidence intervals for the time-varying profitability premium (FF5)
Figure 1.34

Estimates and confidence intervals for the time-varying investment premium (FF5)

1.8. Conclusion

This paper is concerned with estimation of risk premia and testing of beta-pricing models when the data is available for a large cross-section of securities, \( N \), but only for a fixed number of time periods, \( T \). Since in this context the traditional OLS CSR estimator of the risk premia is asymptotically biased and inconsistent, we provide a new methodology built on the appealing bias-adjusted estimator of the ex post risk premia proposed by Shanken (1992). We establish its consistency and asymptotic normality for the baseline case of correctly specified beta-pricing models with constant risk premia, and then extend our setting to deal with time-varying risk premia. We also explore in detail the case of misspecified beta-pricing models. We derive a new specification test and its large-\( N \) properties, and we then show how to robustify the asymptotic standard errors of the risk premia estimator when the beta-pricing relation is violated. The important case of misspecification due to priced firm characteristics is considered. Finally, we analyze the case of unbalanced panels.

We apply our large-\( N \) methodology to empirically investigate the performance of some prominent beta-pricing specifications using individual stock return data, that is, the monthly returns (from CRSP) on about 3,500 individual stocks for the January 1966 – December 2013 period. We consider three beta-pricing models: the CAPM, the three-factor model of Fama and French (1993), and the five-factor model of Fama and French (2015). We also augment these models with the (non-traded) liquidity factor of Pástor and Stambaugh (2003).

Our large-\( N \) test often rejects the Fama and French (2015) model, with and without the liquidity factor, at conventional significance levels even for short time windows of three years. In contrast, when using a suitable aggregation of the same data, in most cases we are unable to reject the Fama and French (2015) model using the traditional large-\( T \) methodologies. Similar conclusions hold when testing the validity of the CAPM and the Fama and French (1993) three-factor model, with and without the liquidity factor. The empirical rejection of these models suggests that the misspecification-robust standard errors derived in this paper should be employed when performing inference on risk premia.

Turning to estimation, our results indicate that all the traded-factor risk premia estimates are statistically significant most of the time, even over short time windows of three years. In contrast, the (non-traded) liquidity factor is often not priced. We also provide evidence of significant time variation in risk premia for both traded and non-traded factors. Our overall evidence of pricing is at odds with the results obtained using the traditional approach based on the large-\( T \) Shanken (1992) standard errors.

Finally, allowing for characteristics in the risk-return relation, we find that the book-to-market ratio, asset growth, operating profitability, market capitalization, and six-month momentum explain most of the cross-sectional variation in estimated expected stock returns. Monte Carlo simulations corroborate our theoretical findings, both in terms of estimation and in terms of testing of the beta-pricing restriction.
1.A. Lemmas

Lemma 1.1 Under Assumptions 1.3-1.5,

\[ \hat{\sigma}^2 - \sigma^2 = \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right). \]  \hspace{1cm} (1.A.1)

Proof. Rewrite \( \hat{\sigma}^2 - \sigma^2 \) as

\[ \hat{\sigma}^2 - \sigma^2 = \left( \frac{\hat{\sigma}^2}{N} - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right) + \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 - \sigma^2 \right) \]
\[ = \left( \frac{\hat{\sigma}^2}{N} - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right) + o \left( \frac{1}{\sqrt{N}} \right) \]  \hspace{1cm} (1.A.2)

by Assumption 5(i). Moreover,

\[ \frac{\hat{\sigma}^2}{N} - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 = \frac{\text{tr} \left( M \epsilon' \epsilon \right)}{N(T-K-1)} - \frac{\text{tr} \left( M \epsilon \epsilon' \right)}{T-K-1} - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \]
\[ = \frac{\text{tr} \left( P \left( \sum_{i=1}^{N} \sigma_i^2 I_T - \epsilon' \epsilon \right) \right)}{N(T-K-1)} + \frac{\text{tr} \left( \epsilon' \epsilon \right) - T \sum_{i=1}^{N} \sigma_i^2}{N(T-K-1)}. \]  \hspace{1cm} (1.A.3)

As for the second term on the right-hand side of Eq. (1.A.3), we have

\[ \frac{\text{tr} \left( \epsilon' \epsilon \right) - T \sum_{i=1}^{N} \sigma_i^2}{N(T-K-1)} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left( \epsilon_{it}^2 - \sigma_i^2 \right)}{N(T-K-1)} \]
\[ = \mathcal{O}_p \left( \frac{1}{\sqrt{N(T-K-1)}} \right) = \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right). \]  \hspace{1cm} (1.A.4)

As for the first term on the right-hand side of Eq. (1.A.3), we have

\[ \frac{\text{tr} \left( P \left( \sum_{i=1}^{N} \sigma_i^2 I_T - \epsilon' \epsilon \right) \right)}{N(T-K-1)} = \frac{\sum_{t=1}^{T} d_t (D'T) \left( \sum_{i=1}^{N} \sigma_{i,t}^2 - \sum_{i=1}^{N} \sigma_{i,t} \right)}{N(T-K-1)} \]
\[ = \frac{\sum_{t=1}^{T} p_t \left( \sum_{i=1}^{N} \sigma_{i,t}^2 \right) - \sum_{i=1}^{N} \epsilon_{i,t} \epsilon_{i,t}}{N(T-K-1)}, \]  \hspace{1cm} (1.A.5)

where \( u_{i,t} \) is a \( T \)-vector with one in the \( t \)-th position and zeros elsewhere, \( d_t \) is the \( t \)-th row of \( D = [1_T, \; F] \), and \( p_t = d_t (D'T)^{-1} D' \). Since Eq. (1.A.5) has a zero mean, we only need to consider its variance to determine the rate of convergence. We have

\[ \text{Var} \left( \sum_{t=1}^{T} p_t \left( \sum_{i=1}^{N} \sigma_{i,t}^2 u_{i,T} - \sum_{i=1}^{N} \xi_{i,t} \right) \right) \]
\[ = \frac{1}{N^2(T-K-1)^2} \mathbb{E} \left[ \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} p_t \left( \sigma_{i,t}^2 u_{i,T} - \epsilon_{i,t} \right) \left( \sigma_{j,s}^2 u_{j,T} - \epsilon_{j,s} \right) \right] \]
\[ = \frac{1}{N^2(T-K-1)^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} p_t \mathbb{E} \left[ \left( \sigma_{i,t}^2 u_{i,T} - \epsilon_{i,t} \right) \left( \sigma_{j,s}^2 u_{j,T} - \epsilon_{j,s} \right) \right] \]  \hspace{1cm} (1.A.6)
Moreover, we have

\[
E \left[ \left( \sigma^2_{u,T} - \epsilon_i \epsilon_i \right) \left( \sigma^2_{v,T} - \epsilon_j \epsilon_j \right) \right] = E \left[ \sigma^2_{u,T} \right] \left( \sigma^2_{v,T} + \epsilon_i \epsilon_i \epsilon_j \epsilon_j - \sigma^2_{u,T} \epsilon_i \epsilon_i \right) \]

\[
= \begin{cases} 
\mu_{4i} \epsilon_i \epsilon_i \epsilon_i \epsilon_i + \sigma^4 \ (I_T - 2u,T) & \text{if } i = j, \ t = s \\
\kappa_{4,ij} \epsilon_i \epsilon_i \epsilon_i \epsilon_i + \sigma^4 \ (I_T + u,T) & \text{if } i \neq j, \ t = s \\
\sigma^4 \epsilon_i \epsilon_i \epsilon_i \epsilon_i & \text{if } i = j, \ t \neq s \\
\sigma^4 \epsilon_i \epsilon_i \epsilon_i \epsilon_i & \text{if } i \neq j, \ t \neq s.
\end{cases}
\]

(1.5.7)

It follows that

\[
\text{Var} \left( \frac{\sum_{l=1}^{T} p_l \left( \sum_{i=1}^{N} \sigma^2_{u,T} - \sum_{i=1}^{N} \epsilon_i \epsilon_i \right)}{N(T-K-1)} \right) = \frac{1}{N^2(T-K-1)^2} \sum_{t=1}^{T} \sum_{i=1}^{N} p_t \left( \mu_{4i} \epsilon_i \epsilon_i \epsilon_i \epsilon_i + \sigma^4 \ (I_T - 2u,T) \right) p_t'
\]

\[
+ \frac{1}{N^2(T-K-1)^2} \sum_{t=1}^{T} \sum_{i \neq j}^{N} p_t \left( \kappa_{4,ij} \epsilon_i \epsilon_i \epsilon_i \epsilon_i + \sigma^4 \ (I_T + u,T) \right) p_t'
\]

\[
+ \frac{1}{N^2(T-K-1)^2} \sum_{i=1}^{N} \sigma^4 \sum_{t \neq s} p_t \epsilon_i \epsilon_i \epsilon_i \epsilon_i p_t'
\]

\[
+ \frac{1}{N^2(T-K-1)^2} \sum_{i \neq j}^{N} \sigma^4 \sum_{t \neq s} p_t \epsilon_i \epsilon_i \epsilon_i \epsilon_i p_t'
\]

= \frac{1}{N} 
\]

(1.5.8)

by Assumptions 1.5(ii), 1.5(iii), 1.5(iv), and 1.5(viii), which implies that the first term on the right-hand side of Eq. (1.5.3) is \( O_p \left( \frac{1}{N} \right) \). Putting the pieces together concludes the proof.

\( \blacksquare \)

Lemma 1.2 Let

\[
\Lambda = \begin{bmatrix} 0 & 0_K \sigma^2 (P'P)^{-1} \end{bmatrix}.
\]

(1.5.9)

(i) Under Assumptions 1.2-1.5,

\[
\bar{X}' \bar{X} = O_p(N).
\]

(1.5.10)

In addition, under Assumption 1.6,

(ii) \( \hat{\Sigma}_X \rightarrow_p \Sigma X + \Lambda \),

(1.5.11)

and

(iii) \( \frac{\bar{X} - X}'(\bar{X} - X) \rightarrow_p \Lambda \).

(1.5.12)
Proof.

(i) Consider
\[ \hat{X}'\hat{X} = \begin{bmatrix} N & 1_N \hat{B} \\
\hat{B}'_1N & \hat{B}'\hat{B} \end{bmatrix}. \] (1.A.13)

Then,
\[ \hat{B}'_1N = \sum_{i=1}^{N} \hat{\beta}_i = \sum_{i=1}^{N} \beta_i + \mathcal{P}' \sum_{i=1}^{N} \epsilon_i. \] (1.A.14)

Under Assumptions 1.A-1.5,
\begin{align*}
\text{Var} \left( \sum_{t=1}^{T} \sum_{i=1}^{N} \epsilon_{it}(f_t - \bar{f}) \right) &= \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} (f_t - \bar{f})(f_s - \bar{f})' E[\epsilon_{it}\epsilon_{js}] \\
&\leq \sum_{t=1}^{T} \sum_{i,j=1}^{N} (f_t - \bar{f})(f_t - \bar{f})' \sigma_{ij} \\
&= O \left( N\sigma^2 \sum_{t=1}^{T} (f_t - \bar{f})(f_t - \bar{f})' \right) = O(NT). \quad (1.A.15)
\end{align*}

Using Assumption 1.2, we have
\[ \hat{B}'_1N = O_p \left( N + \left( \frac{N}{T} \right)^{\frac{3}{2}} \right) = O_p(N). \] (1.A.16)

Next, consider
\begin{align*}
\hat{B}'\hat{B} &= \sum_{i=1}^{N} \hat{\beta}_i\hat{\beta}_i' \\
&= \sum_{i=1}^{N} (\beta_i + \mathcal{P}'\epsilon_i) (\beta_i' + \epsilon_i'\mathcal{P}) \\
&= \sum_{i=1}^{N} \beta_i\beta_i' + \mathcal{P}' \left( \sum_{i=1}^{N} \epsilon_i\epsilon_i' \right) \mathcal{P} \\
&\quad + \mathcal{P}' \left( \sum_{i=1}^{N} \epsilon_i\beta_i' \right) + \left( \sum_{i=1}^{N} \beta_i\epsilon_i' \right) \mathcal{P} \text{.} \quad (1.A.17)
\end{align*}

By Assumption 1.2,
\[ \sum_{i=1}^{N} \beta_i\beta_i' = O(N). \] (1.A.18)

Using similar arguments as for Eq. (1.A.15),
\[ \mathcal{P}' \left( \sum_{i=1}^{N} \epsilon_i\beta_i' \right) = O_p \left( \left( \frac{N}{T} \right)^{\frac{3}{2}} \right) \] (1.A.19)

and
\[ \left( \sum_{i=1}^{N} \beta_i\epsilon_i' \right) \mathcal{P} = O_p \left( \left( \frac{N}{T} \right)^{\frac{3}{2}} \right) \text{.} \] (1.A.20)
For $\mathcal{P}' \left( \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) \mathcal{P}$, consider its central part and take the norm of its expectation. Using Assumptions 1.4-1.5,

$$
\left\| E \left[ \tilde{F}' \left( \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) \tilde{F} \right] \right\| \\
= \left\| E \left[ \sum_{t,s=1}^{T} \sum_{i=1}^{N} (f_t - \bar{f})(f_s - \bar{f})' \epsilon_{it} \epsilon_{is} \right] \right\| \\
\leq \sum_{t,s=1}^{T} \sum_{i=1}^{N} \left\| (f_t - \bar{f})(f_s - \bar{f})' \right\| E |\epsilon_{it} \epsilon_{is}| \\
= \sum_{t=1}^{T} \sum_{i=1}^{N} \left\| (f_t - \bar{f})(f_s - \bar{f})' \right\| \sigma_i^2 \\
= O \left( N \sigma_i^2 \sum_{t=1}^{T} \left\| (f_t - \bar{f})(f_s - \bar{f})' \right\| \right) = O(NT). \quad (1.A.21)
$$

Then, we have

$$
\mathcal{P}' \left( \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) \mathcal{P} = O_p \left( \frac{N}{T} \right) \quad (1.A.22)
$$

and

$$
\hat{B}' \hat{B} = O_p \left( N + \left( \frac{N}{T} \right)^{\frac{1}{2}} + \frac{N}{T} \right) = O_p(N). \quad (1.A.23)
$$

This concludes the proof of part (i).

(ii) Using part (i) and under Assumptions 1.3-1.6, we have

$$
N^{-1} \hat{B}' 1_N = \frac{1}{N} \sum_{i=1}^{N} \beta_i + O_p \left( \frac{1}{\sqrt{N}} \right) \quad (1.A.24)
$$

and

$$
N^{-1} \hat{B}' \hat{B} = \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) \mathcal{P} + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \beta_i' \right) + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \beta_i \epsilon_i' \right) \mathcal{P} \\
= \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i' - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 I_T \right) + \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 I_T - \sigma^2 \sigma_i^2 I_T + \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 I_T \mathcal{P} \\
+ \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \beta_i' \right) + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \beta_i \epsilon_i' \right) \mathcal{P} \\
= \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \left( \sigma_i^2 - \sigma^2 \right) I_T \right) + \frac{1}{N} \sum_{i=1}^{N} \left( \sigma_i^2 - \sigma^2 \right) \mathcal{P} \mathcal{P} + \sigma^2 \mathcal{P} \mathcal{P} \\
+ \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \beta_i' \right) + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \beta_i \epsilon_i' \right) \mathcal{P} \\
= \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' + \sigma^2 \mathcal{P} \mathcal{P} + O_p \left( \frac{1}{\sqrt{N}} \right) + o \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{N}} \right). \quad (1.A.25)
$$
Assumption 1.2 concludes the proof of part (ii).

(iii) Note that

\[
\frac{(\hat{X} - X)'(\hat{X} - X)}{N} = \frac{1}{N} \begin{bmatrix} 0' \n 0 \\ 0
\end{bmatrix} \left[ 0_N, (\hat{B} - B)' \right]
\]

\[
= \begin{bmatrix} 0' \\ 0 \\ 0
\end{bmatrix} \frac{0'_K}{N} \mathcal{P} \left. \begin{bmatrix} 0' \\ 0
\end{bmatrix} \right\}.
\]

(1.26)

As in part (ii) we can write

\[
\frac{\epsilon' \epsilon'}{N} = \frac{1}{N} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma_i^2 I_T) + \left( \frac{1}{N} \sum_{i=1}^{N} (\sigma_i^2 - \sigma^2) \right) I_T + \sigma^2 I_T.
\]

(1.27)

Assumptions 1.5(i) and 1.6(ii) conclude the proof since

\[
\mathcal{P}' \frac{\epsilon' \epsilon'}{N} \mathcal{P} = \sigma^2 \mathcal{P}' \mathcal{P} + O_p \left( \frac{1}{\sqrt{N}} \right) + o \left( \frac{1}{\sqrt{N}} \right).
\]

(1.28)

Lemma 1.3

Under Assumptions 1.2-1.5,

\[
X' \tau = O_p \left( \sqrt{N} \right).
\]

(1.29)

Proof. We have

\[
X' \tau = \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} 1' \\ B'
\end{bmatrix} \epsilon_t
\]

(1.30)

and

\[
\text{Var} \left( \frac{1}{T} \sum_{t=1}^{T} 1' \epsilon_t \right) = \frac{1}{T^2} \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} E[\epsilon_{it} \epsilon_{js}]
\]

\[
\leq \frac{1}{T^2} \sum_{t=1}^{T} \sum_{i,j=1}^{N} | \sigma_{ij} |
\]

\[
= O \left( \frac{NT}{T^2} \sigma^2 \right) = O(N).
\]

(1.31)

Moreover, using Assumptions 1.2 and 1.5(ii),

\[
\text{Var} \left( \frac{1}{T} \sum_{t=1}^{T} B' \epsilon_t \right) = \frac{1}{T^2} \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} E[\epsilon_{it} \epsilon_{js}] \beta_i \beta_j'
\]

\[
\leq \frac{1}{T^2} \sum_{t=1}^{T} \sum_{i,j=1}^{N} | \beta_i \beta_j' | | \sigma_{ij} |
\]

\[
= O \left( \frac{NT}{T^2} \sigma^2 \right) = O(N).
\]

(1.32)

Putting the pieces together, \(X' \tau = O_p \left( \sqrt{N} \right)\).
Lemma 1.4

Under Assumptions 1.3-1.5,
\[
(\bar{X} - X)^\prime X\Gamma_0 = O_p\left(\sqrt{N}\right).
\]  
(1.A.33)

Proof. We have
\[
(\bar{X} - X)^\prime X\Gamma_0 = \begin{bmatrix} 0 & \Gamma_0 \end{bmatrix} X\Gamma_0.
\]  
(1.A.34)

Using similar arguments to Eq. (1.A.15) concludes the proof. ■

Lemma 1.5

Under Assumptions 1.3-1.5,
\[
(\bar{X} - X)^\prime \epsilon = O_p\left(\sqrt{N}\right).
\]  
(1.A.35)

Proof.
\[
(\bar{X} - X)^\prime \epsilon = \begin{bmatrix} 0 \\ \mathcal{P}^\prime \epsilon \end{bmatrix} = \begin{bmatrix} 0 \\ \mathcal{P}^\prime \epsilon \right\mathcal{P}^{\prime \prime} \end{bmatrix}
\]
\[
= \mathcal{P}^\prime \left[ \left( \epsilon \epsilon' - \sum_{i=1}^{N} \sigma_i^2 I_T \right) + \left( \sum_{i=1}^{N} \sigma_i^2 - N \sigma^2 \right) I_T \right] \frac{1}{T} = O_p(\sqrt{N})
\]  
(1.A.36)

by Assumption 1.5. ■

Lemma 1.6 Under Assumption 1.5 and the identification assumption \( \kappa_4 = 0 \), we have
\[
\delta_4 \rightarrow p \sigma_4.
\]  
(1.A.37)

Proof. We need to show that (i) \( E(\hat{\sigma}_4) \rightarrow \sigma_4 \) and (ii) \( \text{Var}(\hat{\sigma}_4) = O\left(\frac{1}{N}\right) \).

(i) By Assumptions 1.5(iv), 1.5(vi), and 1.5(vii), we have
\[
E \left[ \frac{1}{N} \sum_{i=1}^{T} \sum_{t=1}^{N} \epsilon_{it}^4 \right] = \frac{1}{N} \sum_{i=1}^{T} \sum_{t=1}^{N} E \left[ \epsilon_{it}^4 \right]
\]
\[
= \frac{1}{N} \sum_{i=1}^{T} \sum_{t=1}^{N} \sum_{s_1, s_2, s_3, s_4 = 1}^{T} m_{i_1 i_2 m_{i_3 i_4} m_{i_4 i_4}} E \left[ \xi_{i_1 i_2 i_3 i_4} \right]
\]
\[
= \frac{1}{N} \sum_{i=1}^{T} \sum_{t=1}^{N} \kappa_4 \sum_{s=1}^{T} m_{i_4 i_4} + 3 \frac{1}{N} \sum_{i=1}^{T} \sum_{s=1}^{T} \sigma_i^2 \left( \sum_{s=1}^{T} m_{i_4 i_4}^2 \right)^2
\]
\[
\rightarrow \kappa_4 \sum_{i=1}^{T} \sum_{t=1}^{N} m_{i_4 i_4} + 3 \sigma_4 \sum_{i=1}^{T} \left( \sum_{s=1}^{T} m_{i_4 i_4}^2 \right)^2,
\]  
(1.A.38)

where \( \epsilon_{it} = \epsilon_{i,T} M \epsilon_i \) and \( M = [m_{it}] \) for \( t, s = 1, \ldots, T \). Note that
\[
\sum_{s=1}^{T} m_{i_4 i_4}^2 = ||m_{i_4}||^2
\]
\[
= \epsilon_{i,T} M \epsilon_i
\]
\[
= \epsilon_{i} \left( I_T - D(D'D)^{-1}D' \right) \epsilon_{i}
\]
\[
= 1 - \text{tr} \left( D(D'D)^{-1}D' \epsilon_{i} \right)
\]
\[
= 1 - \text{tr} \left( P \epsilon_{i} \right)
\]
\[
= 1 - p_{it}
\]  
(1.A.39)
where \( p_{tt} \) is the \((t, t)\)-element of \( P \). Then, we have

\[
\sum_{t=1}^{T} \left( \sum_{s=1}^{T} m_{ts} \right)^2 = \sum_{t=1}^{T} m_{tt}^2 = \text{tr} \left( M^{(2)} \right).
\]  

(1.40)

By setting \( \kappa_4 = 0 \), it follows that

\[
E \left[ \hat{\sigma}_4 \right] \to \sigma_4.
\]

(1.41)

This concludes the proof of part (i).

(ii) As for the variance of \( \hat{\sigma}_4 \), we have

\[
\text{Var} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{e}_{it}^2 \right) = \frac{1}{N^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \text{Cov} \left( \tilde{e}_{it}, \tilde{e}_{js} \right)
\]

\[
= \frac{1}{N^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \sum_{u_1, u_2, u_3, u_4 = 1}^{T} \sum_{v_1, v_2, v_3, v_4 = 1}^{T} \sum_{u_1, u_2, u_3, u_4 = 1}^{T} \sum_{v_1, v_2, v_3, v_4 = 1}^{T}
\]

\[
\times \text{Cov} \left( \epsilon_{u_1, u_2, u_3, u_4}, \epsilon_{j_1, j_2, j_3, j_4} \right)
\]

\[
= \frac{1}{N^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \sum_{u_1, u_2, u_3, u_4 = 1}^{T} \sum_{v_1, v_2, v_3, v_4 = 1}^{T}
\]

\[
\times \left( \kappa_8 \left( \epsilon_{u_1, u_2, u_3, u_4}, \epsilon_{j_1, j_2, j_3, j_4} \right)
\]

\[
+ \sum_{j} \kappa_6 \left( \epsilon_{u_1, u_2, u_3, u_4}, \epsilon_{j_1, j_2, j_3, j_4} \right) \text{Cov} \left( \epsilon_{j_1, j_2, j_3, j_4} \right)
\]

\[
+ \sum_{j} \kappa_4 \left( \epsilon_{u_1, u_2, u_3, u_4}, \epsilon_{j_1, j_2, j_3, j_4} \right) \kappa_4 \left( \epsilon_{u_1, u_2, u_3, u_4}, \epsilon_{j_1, j_2, j_3, j_4} \right)
\]

\[
+ \sum_{j} \kappa_4 \left( \epsilon_{u_1, u_2, u_3, u_4}, \epsilon_{j_1, j_2, j_3, j_4} \right) \text{Cov} \left( \epsilon_{u_1, u_2, u_3, u_4} \right) \text{Cov} \left( \epsilon_{j_1, j_2, j_3, j_4} \right)
\]

\[
+ \sum \text{Cov} \left( \epsilon_{u_1, u_2, u_3, u_4} \right) \text{Cov} \left( \epsilon_{u_1, u_2, u_3, u_4} \right) \text{Cov} \left( \epsilon_{j_1, j_2, j_3, j_4} \right)
\]

\[(1.42)\]

where \( \kappa_4 \left( \cdot \right), \kappa_6 \left( \cdot \right), \) and \( \kappa_8 \left( \cdot \right) \) denote the fourth-, sixth-, and eighth-order mixed cumulants, respectively. By \( \sum_{\nu_1, \nu_2, \ldots, \nu_k} \) we denote the sum over all possible partitions of a group of \( K \) random variables into \( k \) subgroups of size \( \nu_1, \nu_2, \ldots, \nu_k \), respectively. As an example, consider \( \sum_{(6,2)} \). \( \sum_{(6,2)} \) defines the sum over all possible partitions of the group of eight random variables \( \{ \epsilon_{u_1, u_2, u_3, u_4}, \epsilon_{j_1, j_2, j_3, j_4} \} \) into two subgroups of size six and two, respectively. Moreover, since \( E \left[ \epsilon_{it} \right] = \text{tr} \left( M^{(2)} \right) = 0 \), we do not need to consider further partitions in the relation above.\(^{52}\) Then, under Assumptions 1.5(i), 1.5(ii), 1.5(vii), and 1.5(viii), it follows that

\[
\text{Var} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{e}_{it}^2 \right) = O \left( \frac{1}{N} \right)
\]

\[(1.43)\]

\(^{52}\) According to the theory on cumulants (Brillinger (2001)), evaluation of \( \text{Cov} \left( \epsilon_{u_1, u_2, u_3, u_4}, \epsilon_{j_1, j_2, j_3, j_4} \right) \) requires considering the indecomposable partitions of the two sets, \( \{ \epsilon_{u_1, u_2, u_3, u_4} \} \) and \( \{ \epsilon_{j_1, j_2, j_3, j_4} \} \), meaning that there must be at least one subset that includes an element of both sets.
and \( \text{Var}(\hat{\sigma}_t) = O\left( \frac{1}{N} \right) \). This concludes the proof of part (ii). ■

**Lemma 1.7** Let \( w = [w_1, \ldots, w_T]' \) and \( s = [s_1, \ldots, s_T]' \) be two arbitrary \( T \)-vectors. Then, under Eq. (1.65) and Assumptions 1.2-1.7,

\[
\frac{1}{N(T-K)} \sum_{i=1}^{N} \sum_{k=1}^{T} w_k \epsilon_{ki} \sum_{r=1}^{T} s_r \epsilon_{ri} \rightarrow_p \frac{\text{tr}(M(S_1 + S_2))}{(T-K)},
\]

(1.A.44)

where \( S_1 = \text{diag}\left( s_1 w_1 \mu_4 + \sigma^4 \sum_{k \neq 1}^{T} w_k s_k, \ldots, s_T w_T \mu_4 + \sigma^4 \sum_{k \neq T}^{T} w_k s_k \right) \) and \( S_2 = \sigma^4 (ws' + sw' - 2 \text{diag}(w_1 s_1, \ldots, w_T s_T)) \).

**Proof.** Note that

\[
\frac{1}{N(T-K)} \sum_{i=1}^{N} \sum_{k=1}^{T} w_k \epsilon_{ki} \sum_{r=1}^{T} s_r \epsilon_{ri} =
\]

\[
= \frac{1}{N(T-K)} \text{tr}\left( M \left( \sum_{i=1}^{N} \epsilon_i' \epsilon_i \sum_{k=1}^{T} w_k s_k \epsilon_{ki}^2 + \sum_{r>k}^{T} w_k s_r \epsilon_{r} \epsilon_{ir} + \sum_{r<k}^{T} w_k s_r \epsilon_{r} \epsilon_{ir} \right) \right).
\]

(1.A.45)

For the first term of Eq. (1.A.45),

\[
\frac{1}{N(T-K)} \text{tr}\left( M \left( \sum_{i=1}^{N} \epsilon_i' \epsilon_i \sum_{k=1}^{T} w_k s_k \epsilon_{ki}^2 \right) \right) = \frac{1}{(T-K)} \text{tr}\left( M \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{T} \epsilon_i' w_k s_k \epsilon_{ki}^2 \right) \right) \rightarrow_p \frac{1}{(T-K)} \text{tr}(MS_1),
\]

(1.A.46)

where

\[
S_1 = \text{plim} \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{T} \epsilon_i' w_k s_k \epsilon_{ki}^2 = \text{diag}\left( s_1 w_1 \mu_4 + \sigma^4 \sum_{k \neq 1}^{T} w_k s_k, \ldots, s_T w_T \mu_4 + \sigma^4 \sum_{k \neq T}^{T} w_k s_k \right).
\]

(1.A.47)

For the second and third terms of Eq. (1.A.45), we obtain

\[
\frac{1}{N(T-K)} \text{tr}\left( M \left( \sum_{i=1}^{N} \epsilon_i' \left( \sum_{r>k}^{T} w_k s_r \epsilon_{r} \epsilon_{ir} + \sum_{r<k}^{T} w_k s_r \epsilon_{r} \epsilon_{ir} \right) \right) \right) = \frac{1}{(T-K)} \text{tr}\left( M \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{r>k}^{T} \epsilon_i' w_k s_r \epsilon_{r} \epsilon_{ir} + \frac{1}{N} \sum_{i=1}^{N} \sum_{r<k}^{T} \epsilon_i' w_k s_r \epsilon_{r} \epsilon_{ir} \right) \right) \rightarrow_p \frac{1}{(T-K)} \text{tr}(MS_2),
\]

(1.A.48)

where

\[
S_2 = \text{plim} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{r>k}^{T} \epsilon_i' w_k s_r \epsilon_{r} \epsilon_{ir} + \frac{1}{N} \sum_{i=1}^{N} \sum_{r<k}^{T} \epsilon_i' w_k s_r \epsilon_{r} \epsilon_{ir} \right) = \sigma^4 (ws' + sw' - 2 \text{diag}(w_1 s_1, \ldots, w_T s_T)).
\]

(1.A.49)
Lemma 1.8 Let \( \tilde{\tau}_\Phi = \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i \hat{\epsilon}_i^P \). Then, under Eq. (1.65) and Assumptions 1.2-1.7,

\[
\tilde{\tau}_\Phi \rightarrow_p \tau_\Phi.
\] (1.A.50)

Proof. Given

\[
\hat{\epsilon}_i^P = \hat{R}_i - \hat{X}_i^* \hat{\Gamma}^*
= X'_i \hat{\Gamma}^P + \epsilon_i - \hat{X}_i \hat{\Gamma}^*
= \epsilon_i + \epsilon_i - (\hat{X}_i - X_i)' \hat{\Gamma}^P - \hat{X}_i' (\hat{\Gamma}^* - \hat{\Gamma}^P),
\] (1.A.51)

using the fact that \( \hat{\epsilon}_i = M \epsilon_i \) and Eq. (1.51), we can write

\[
\begin{align*}
\tilde{\tau}_\Phi &= \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i \hat{\epsilon}_i^P = \frac{1}{N(T-K)} \sum_{i=1}^{N} \epsilon_i^T M \epsilon_i \hat{\epsilon}_i \\
&= \frac{1}{N(T-K)} \sum_{i=1}^{N} \text{tr}(M \epsilon_i^T \epsilon_i) (\epsilon_i + \epsilon_i - (\hat{X}_i - X_i)' \hat{\Gamma}^P - \hat{X}_i' (\hat{\Gamma}^* - \hat{\Gamma}^P)) \\
&= \frac{1}{N(T-K)} \sum_{i=1}^{N} \text{tr}(M \epsilon_i^T \epsilon_i) + o_p(1) \rightarrow_p \frac{1}{(T-K)} \text{tr}(M \tau_\Phi) = \tau_\Phi. 
\end{align*}
\] (1.A.52)

Lemma 1.9 Let

\[
\tilde{\tau}_\Omega = \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i \hat{\epsilon}_i^P - \frac{\sigma^4}{T} \left( 1 + \frac{2 \text{tr}(M \Pi^T \Pi)}{T(T-K)} \right) - \frac{\text{tr}(MS_P)}{(T-K)} + \frac{2 \text{tr}(MC_P)}{T(T-K)}.
\] (1.A.53)

where

\[
S_P = \sigma^4 \begin{bmatrix}
A' \left( 3 \tilde{f}_1 \tilde{f}_1^T + \sum_{i\neq 1}^{T} \tilde{f}_i \tilde{f}_i^T \right) A & 2A' \tilde{f}_1 \tilde{f}_2^T A & \cdots & 2A' \tilde{f}_1 \tilde{f}_T^T A \\
2A' \tilde{f}_2 \tilde{f}_1^T A & A' \left( 3 \tilde{f}_2 \tilde{f}_2^T + \sum_{i\neq 2}^{T} \tilde{f}_i \tilde{f}_i^T \right) & \cdots & 2A' \tilde{f}_2 \tilde{f}_T^T A \\
\vdots & \vdots & \ddots & \vdots \\
2A' \tilde{f}_T \tilde{f}_1^T A & 2A' \tilde{f}_T \tilde{f}_2^T A & \cdots & A' \left( 3 \tilde{f}_T \tilde{f}_T^T + \sum_{i\neq T}^{T} \tilde{f}_i \tilde{f}_i^T \right) A
\end{bmatrix}
\] (1.A.54)

and

\[
C_P = \sigma^4 \begin{bmatrix}
3 \tilde{f}_1^T A + \sum_{i\neq 1}^{T} \tilde{f}_i^T A & (\tilde{f}_1 + \tilde{f}_2)' A & \cdots & (\tilde{f}_1 + \tilde{f}_T)' A \\
(\tilde{f}_2 + \tilde{f}_1)' A & 3 \tilde{f}_2^T A + \sum_{i\neq 2}^{T} \tilde{f}_i^T A & \cdots & (\tilde{f}_2 + \tilde{f}_T)' A \\
\vdots & \vdots & \ddots & \vdots \\
(\tilde{f}_T + \tilde{f}_1)' A & (\tilde{f}_T + \tilde{f}_2)' A & \cdots & 3 \tilde{f}_T^T A + \sum_{i\neq T}^{T} \tilde{f}_i^T A
\end{bmatrix},
\] (1.A.55)

with \( A = (\hat{F}' \hat{F})^{-1} \hat{f}_1 \). Then, under Eq. (1.65) and Assumptions 1.2-1.7,

\[
\tilde{\tau}_\Omega \rightarrow_p \tau_\Omega.
\] (1.A.56)
**Proof:** By Eq. (1.51), we have

\[
(\tilde{e}_i^p)^2 = e_i^2 + \gamma_i^2 \left( (\hat{\beta}_i - \beta_i)^\prime \gamma_i^p \right)^2 + \left( [1, \hat{\beta}_i'] (\tilde{\Gamma}^* - \tilde{\Gamma}^p) \right)^2 + 2 e_i \left( \tilde{\xi}_i - (\hat{\beta}_i - \beta_i)^\prime \gamma_i^p [1, \hat{\beta}_i'] (\tilde{\Gamma}^* - \tilde{\Gamma}^p) \right) + 2 \tilde{\xi}_i \left( - (\hat{\beta}_i - \beta_i)^\prime \gamma_i^p [1, \hat{\beta}_i'] (\tilde{\Gamma}^* - \tilde{\Gamma}^p) \right) + 2 (\hat{\beta}_i - \beta_i)^\prime \gamma_i^p [1, \hat{\beta}_i'] (\tilde{\Gamma}^* - \tilde{\Gamma}^p).
\]

Then,

\[
\hat{\tau}_1 = \frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i (\tilde{e}_i^p)^2
= \frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i e_i^2 + \frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i e_i^2 + \frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i \left( (\hat{\beta}_i - \beta_i)^\prime \gamma_i^p \right)^2
- \frac{2}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i e_i \left( \hat{\beta}_i - \beta_i \right)^\prime \gamma_i^p + o_p(1),
\]

(1.58)

where all terms involving \((\tilde{\Gamma}^* - \tilde{\Gamma}^p)\) are condensed into the \(o_p(1)\) term. By Assumption 1.7, the first term in Eq. (1.58) satisfies

\[
\frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i e_i^2 = \frac{1}{(T-K)} \text{tr} \left( M \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i^\prime \right) \rightarrow_p \frac{1}{(T-K)} \text{tr}(M \tau_1) = \tau_1.
\]

(1.59)

For the second term in Eq. (1.58), we have

\[
\frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i e_i^2 = \frac{1}{T-K} \frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i \tilde{\xi}_i^\prime \rightarrow_p \sum_{s=1}^{T} \sum_{t=1}^{T} \epsilon_{ts}.
\]

(1.60)

Then, applying Lemma 1.7 with \(w = s = [1, \ldots, 1]^\prime\), we have

\[
\frac{1}{T^2} \frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i \tilde{\xi}_i^\prime \rightarrow_p \frac{\sigma^4}{T} \left( 1 + \frac{2 \text{tr}(M \tau_1^2 \tilde{\gamma}^p)}{T(T-K)} \right).
\]

(1.61)

For the third term in Eq. (1.58), we have

\[
\frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i \left( (\hat{\beta}_i - \beta_i)^\prime \gamma_i^p \right)^2 = \frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i \sum_{t=1}^{T} \gamma_i^p (\tilde{\Gamma}^* \tilde{\Gamma}^t)^{-1} f_t \epsilon_{it} \sum_{s=1}^{T} \tilde{\gamma}_t^p (\tilde{\Gamma}^t \tilde{\Gamma}^p)^{-1} f_s \epsilon_{is},
\]

(1.62)

and by Lemma 1.7 with \(w = s = [\gamma_1^p (\tilde{\Gamma}^* \tilde{\Gamma}^t)^{-1} f_t, \ldots, \tilde{\gamma}_1^p (\tilde{\Gamma}^t \tilde{\Gamma}^p)^{-1} f_T]^{\prime}\), one obtains

\[
\frac{1}{N(T-K)} \sum_{i=1}^{N} \tilde{\xi}_i \left( (\hat{\beta}_i - \beta_i)^\prime \gamma_i^p \right)^2 \rightarrow_p \frac{\text{tr}(M S_F)}{(T-K)}.
\]

(1.63)
Finally, for the fourth term in Eq. (1.58), rewriting it as

\[-2 \frac{1}{N(T-K)} \sum_{i=1}^{N} \frac{\epsilon'_i \gamma_i}{(\hat{\beta}_i - \beta_i)'} \gamma_i' = -2 \frac{1}{NT(T-K)} \sum_{i=1}^{N} \epsilon'_i \epsilon_i \sum_{s=1}^{T} \epsilon_{is} F'_s (\hat{F}'_s \hat{F})^{-1} \gamma_i',\]

(1.64)

and applying again Lemma 1.7 with \( w = [1, \ldots, 1]' \) and \( s = [A'_1 \hat{f}_1, \ldots, A'_T \hat{f}_T]' \), we obtain

\[-2 \frac{1}{N(T-K)} \sum_{i=1}^{N} \epsilon'_i \epsilon_i (\hat{\beta}_i - \beta_i)' \gamma_i' \rightarrow_p -2 \frac{\text{tr}(MC_P)}{T(T-K)}.\]

(1.65)

1.B. Proofs of Propositions and Theorems

**Proof of Proposition 1.1.** Consider the class of additive bias-adjusted estimators \( \hat{\Gamma}^{bias-adj} \) for \( \Gamma^P \):

\[ \hat{\Gamma}^{bias-adj} = \hat{\Gamma} + \frac{\hat{X}' \hat{X}}{N} \left( \sum_{i=1}^{m} \epsilon'_i \epsilon_i (\hat{\beta}_i - \beta_i)' \gamma_i' \right) \rightarrow_p \hat{\Gamma}^{prelim}, \]

(1.B.1)

where \( \hat{\Gamma}^{prelim} \) denotes any preliminary \( \sqrt{N} \)-consistent estimator of \( \Gamma^P \). Setting \( \hat{\Gamma}^{bias-adj} = \hat{\Gamma}^{prelim} \) and rearranging terms, we obtain

\[ \left[ I_{K+1} - \left( \frac{\hat{X}' \hat{X}}{N} \right)^{-1} \right] \left( \begin{array}{cc} 0 & 0 \cr 0 & \hat{\lambda} \end{array} \right) \left( \begin{array}{c} \hat{\lambda} \cr \hat{\lambda}' \end{array} \right) = \hat{\Gamma}^{bias-adj} \]

(1.B.2)

which implies that

\[ \hat{\Gamma}^{bias-adj} = \left( \hat{\sum}_X - \hat{\lambda} \right)^{-1} \frac{\hat{X}' \hat{R}}{N} = \hat{\lambda} \lambda. \]

(1.B.3)

**Proof of Proposition 1.2.** By means of simple calculations, \( \Sigma = \lambda \lambda' + \sigma^2 N \). Thus, \( \sum_{i=1}^{N} \sigma_i^2 / N = \sigma^2 N \), because \( \sum_{i=1}^{N} \lambda_i^2 \leq (\sum_{i=1}^{N} |\lambda_i|)^2 = O(N^{2\delta}) = o(N) \). Therefore, setting \( \sigma^2 = \sigma^2 \), one obtains \( \sum_{i=1}^{N} (\sigma_i^2 - \sigma^2) / N = \sum_{i=1}^{N} \lambda_i^2 / N = \lambda^2 / N = \sigma^2 / N = O(N^{\delta-1} + N^{2\delta-1}) = o(\sqrt{N}) \). It follows that Assumption 1.5(ii) is satisfied.

Next, given that \( \sigma_{ij} = \lambda_i \lambda_j \) for \( i \neq j \), we obtain \( \sum_{i \neq j} \sigma_{ij} \leq (\sum_{i=1}^{N} |\lambda_i|)^2 = O(N^{2\delta}) = o(N) \), thus satisfying Assumption 1.5(ii).

The maximum eigenvalue of \( \Sigma \) is bounded from below by the maximum eigenvalue of \( \lambda \lambda' \), which equals \( \lambda \lambda' \) (all the other \( N-1 \) eigenvalues of \( \lambda \lambda' \) are zero), where \( \lambda^2 + \cdots + \lambda^2 \leq \lambda \lambda' = O(N^{2\delta}) \). Therefore, the maximum eigenvalue diverges at least at rate \( o(\sqrt{N}) \).

**Proof of Proposition 1.3.** The Fama and MacBeth (1973) standard errors with the Shanken (1992) correction are given by

\[ SE_k^{FM} = \left( 1 + c \right) W_k - \mathbb{I}_{(k>0)} \gamma_k^2 / T \]

and \( SE_k^{FM,P} = \left( 1 + c \right) (W_k - \mathbb{I}_{(k>0)} \gamma_k^2 / T) \),

(1.B.4)

for \( k = 0, \ldots, K \), where \( W_k = V_{k+1,k+1} \sum_{t=1}^{T} (F_t - \hat{F}_t)'(F_t - \hat{F}_t)' / (T-1) \), \( \hat{F}_t = \left( \hat{X}' \hat{X} \right)^{-1} \hat{X}' \hat{R}_t \) with sample mean \( \hat{F} \), \( \gamma_j \) denotes the \( j \)-th column, for \( j = 1, \ldots, J \), of the identity matrix \( I_J \), \( \hat{\lambda} \) is the indicator function, and \( \gamma_k^2 / T \) denotes the \( (k,k) \)-th element of \( \hat{F}' \hat{F} / T \).

Consider the numerator of the \( t \)-ratios first. By Lemma 1.2(ii) and Lemmas 1.4 and 1.5, we obtain

\[ \hat{\Gamma} = \left[ \gamma_0, \gamma_1 \right]' = \left( \sum_{i=1}^{m} \gamma_i \right)^{-1} \Sigma X \Gamma^P + O_p \left( \frac{1}{\sqrt{N}} \right). \]

By the blockwise formula of the inverse of a matrix (Magnus
and Neudecker (2007), Section 1-11),

\[
(\Sigma_X + \Lambda)^{-1} \Sigma_X \Gamma^P = \left[ \begin{array}{cc} 1 & \mu'_\beta \\ \mu_\beta & \Sigma_\beta + C \end{array} \right]^{-1} \left[ \begin{array}{cc} 1 & \mu'_\beta \\ \mu_\beta & \Sigma_\beta \end{array} \right] \Gamma^P \\
= \left[ \begin{array}{cc} 1 + \mu'_\beta A^{-1} \mu_\beta & -\mu'_\beta A^{-1} \\ -A^{-1} \mu_\beta & A^{-1} \end{array} \right] \Gamma^P \\
= \left[ \begin{array}{cc} 1 & \mu'_\beta - \mu'_\beta A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta) \\ 0 & A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta) \end{array} \right] \Gamma^P.
\tag{1.1.5}
\]

Then,

\[
(\Sigma_X + \Lambda)^{-1} \Sigma_X \Gamma^P - \Gamma = \left[ \begin{array}{cc} 1 & \mu'_\beta - \mu'_\beta A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta) \\ 0 & A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta) \end{array} \right] \Gamma^P - \Gamma \\
= \left[ \begin{array}{cc} 0 & \mu'_\beta(I_K - A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta)) \\ 0 & -(I_K - A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta)) \end{array} \right] \Gamma \\
+ \left[ \begin{array}{cc} 1 & \mu'_\beta(I_K - A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta)) \\ 0 & A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta) \end{array} \right]\left[ \begin{array}{c} f - E[f_t] \\ f - E[f_t] \end{array} \right].
\tag{1.1.6}
\]

Hence, \( \hat{\gamma}_0 - \gamma_0 = \mu'_\beta(I_K - A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta)) \gamma^0 = \mu'_\beta A^{-1} C \gamma^0 \) and, for every \( j = 1, \ldots, K \), \( \hat{\gamma}_{1j} - \gamma_{1j} = -\gamma_{1j} I_K - A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta) \gamma_{1j} + \gamma_{1j} I_K - A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta)(f - E[f_t]) \) and \( \hat{\gamma}_{1j} - \gamma_{1j} = -\gamma_{1j} I_K - A^{-1}(\Sigma_\beta - \mu_\beta \mu'_\beta) \gamma_{1j}^0 \). Consider now the behavior of the denominator of the t-ratios. It is easy to see that

\[
\hat{W} = \frac{1}{T-1} \sum_{t=1}^T \left( \hat{r}_t - \hat{r} \right)^2 = \hat{W}_a + \hat{W}_b + \hat{W}_c,
\]

where

\[
\hat{W}_a = \hat{X}'\hat{X}^{-1} \hat{X} \left[ \frac{1}{T-1} \sum_{t=1}^T (\hat{r}_t - \hat{r})(\hat{r}_t - \hat{r})' \right] (\hat{X}'\hat{X})^{-1},
\tag{1.1.7}
\]

\[
\hat{W}_b = \hat{X}'\hat{X}^{-1} \hat{X}'B \left[ \frac{1}{T-1} \sum_{t=1}^T (\hat{r}_t - \hat{r})(\hat{r}_t - \hat{r})' \right] B' \hat{X}'(\hat{X}'\hat{X})^{-1} \text{ and}
\tag{1.1.8}
\]

\[
\hat{W}_c = \hat{X}'\hat{X}^{-1} \hat{X}' \left[ \frac{\sum_{t=1}^T (\hat{r}_t - \hat{r})(\hat{r}_t - \hat{r})'}{T-1} \right] B' \hat{X}'(\hat{X}'\hat{X})^{-1}
\]

\[
+ \hat{X}'\hat{X}^{-1} \hat{X}' \left[ \frac{\sum_{t=1}^T (\hat{r}_t - \hat{r})(\hat{r}_t - \hat{r})'}{T-1} \right] B' \hat{X}'(\hat{X}'\hat{X})^{-1}.
\tag{1.1.9}
\]

Based on Lemmas 1.2.1.4 (details are available upon request), we obtain

\[
\hat{W} \rightarrow_p W = W_a + W_b + W_c \equiv (\Sigma_X + \Lambda)^{-1} \left[ \begin{array}{c} 0 \\ \frac{1}{T-1} (\hat{F}'\hat{F})^{-1} \end{array} \right] (\Sigma_X + \Lambda)^{-1}
\]

\[
+ (\Sigma_X + \Lambda)^{-1} \left[ \begin{array}{c} \mu'_\beta \\ \Sigma_\beta \end{array} \right] \left[ \begin{array}{c} f - E[f_t] \\ f - E[f_t] \end{array} \right] (\Sigma_X + \Lambda)^{-1}
\]

\[
+ (\Sigma_X + \Lambda)^{-1} \frac{\sigma^2}{T-1} \left[ \begin{array}{c} \mu'_\beta \\ 2\Sigma_\beta \end{array} \right] (\Sigma_X + \Lambda)^{-1}.
\tag{1.1.10}
\]

It follows that

\[
W = \left[ \begin{array}{cc} 0 & \frac{0_K'}{T-1} \\ 0_K & \frac{\mu'_\beta}{T-1} \end{array} \right].
\tag{1.1.11}
\]
Therefore, since \( \hat{W}_{k} = (k+1, 1, \ldots, K)_{k+1} \hat{W}_{k+1, 1, \ldots, K+1} \) for \( k = 0, \ldots, K \), we have \((1 + \delta)(\hat{W}_{k} - \mathbb{1}_{(k>0)}\hat{\sigma}_{k}^{2}) \rightarrow_{p} 0 \) for any value of \( \delta \). It follows that \( SE_{k}^{FM} \rightarrow_{p} \hat{\sigma}_{k}/\sqrt{T} \) and \( SE_{k}^{EM} \rightarrow_{p} 0 \). The proof of parts (i) and (ii) follows from dividing \( \gamma \) by \( \gamma_{0}, \gamma_{1}, \gamma_{1k}, \) and \( \gamma_{0k} \) by \( SE_{k}^{FM} \) and \( SE_{k}^{EM} \), for the ex ante and ex post risk premia, respectively, and then taking the limit as \( N \to \infty \). ■

**Proof of Theorem 1.1.** For part (i), starting from Eq. (1.12), we have

\[
\hat{\Gamma} = \left( \hat{\Sigma} - \hat{\Lambda} \right)^{-1} \begin{bmatrix} \hat{X}' \hat{R} \end{bmatrix} = \left( \hat{\Sigma} - \hat{\Lambda} \right)^{-1} \begin{bmatrix} \hat{X}' \Gamma + \hat{\epsilon} - (\hat{X} - X)\Gamma \end{bmatrix} = \left( \hat{\Sigma} - \hat{\Lambda} \right)^{-1} \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \Gamma + \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{X}' \hat{X} \in \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{X}' \hat{X} (\hat{X} - X) \Gamma
\end{bmatrix} = \begin{bmatrix} I_{K+1} - \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{\Lambda} \end{bmatrix}^{-1} \begin{bmatrix} \Gamma + \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{X}' \hat{X} \in \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{X}' \hat{X} (\hat{X} - X) \Gamma
\end{bmatrix}.
\]

Hence,

\[
\hat{\Gamma} - \Gamma = \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} - \hat{\Lambda} \right)^{-1} \begin{bmatrix} \hat{X}' \hat{X} \in \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{X}' \hat{X} (\hat{X} - X) \Gamma + \hat{\Lambda} \Gamma
\end{bmatrix} = \left( \hat{\Sigma} - \hat{\Lambda} \right)^{-1} \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \Gamma + \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{X}' \hat{X} \in \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{X}' \hat{X} (\hat{X} - X) \Gamma
\end{bmatrix} = \begin{bmatrix} I_{K+1} - \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{\Lambda} \end{bmatrix}^{-1} \begin{bmatrix} \Gamma + \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{X}' \hat{X} \in \left( \begin{bmatrix} \hat{X}' \hat{X} \end{bmatrix} \right)^{-1} \hat{X}' \hat{X} (\hat{X} - X) \Gamma
\end{bmatrix}.
\]

By Lemmas 1.1 and 1.2(i), \( \left( \hat{\Sigma} - \hat{\Lambda} \right) = O_{p}(1) \). In addition, Lemmas 1.3 and 1.5 imply that

\[
\begin{align*}
\hat{X}' \hat{X} & = \frac{1}{N} (\hat{X} - X)' \hat{X} + \frac{1}{N} X' \hat{X} \\
& = O_{p} \left( \frac{1}{\sqrt{N}} \right),
\end{align*}
\]

and Assumption 1.6(i) implies that

\[
P' \sum_{i=1}^{N} \epsilon_{i} = O_{p} \left( \sqrt{N} \right).
\]

Note that

\[
P' \left( \frac{\epsilon' \epsilon}{N} - \frac{1}{N} \sum_{i=1}^{N} \sigma_{i}^{2} I_{T} \right) - \hat{\sigma}^{2} \left( \hat{\Gamma} \right)^{-1} \gamma_{1}^{p}
\]

can be rewritten as

\[
P' \left( \frac{\epsilon' \epsilon}{N} - \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_{i}^{2} \right) \right) - \left( \hat{\sigma}^{2} - \hat{\sigma}^{2} \right) - \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_{i}^{2} - \sigma^{2} \right) \left( \hat{\Gamma} \right)^{-1} \gamma_{1}^{p}.
\]
Assumption 1.6(ii) implies that
\[
P^p \left( \frac{\epsilon' e' \sigma_2^2}{\sqrt{N}} - \frac{1}{N} \sum_{t=1}^N \sigma_t^2 X_t \right) \mathcal{P} \gamma_1^P = O_p \left( \frac{1}{\sqrt{N}} \right). \tag{1.B.18}
\]

Using Lemma 1.1 and Assumption 1.5(i) concludes the proof of part (i) since \( \hat{\sigma}^2 - \sigma^2 = O_p \left( \frac{1}{\sqrt{N}} \right) \) and \( \frac{1}{N} \sum_{t=1}^N \sigma_t^2 - \sigma^2 = o \left( \frac{1}{\sqrt{N}} \right) \).

For part (ii), starting from (1.B.13), we have
\[
\sqrt{N} (\hat{\gamma}^* - \gamma^p) = (\hat{\Sigma} - \hat{\lambda})^{-1} \left[ \frac{\hat{X}' e}{\sqrt{N}} - \left( \frac{\hat{X}'}{\sqrt{N}} (\hat{X} - X) \Gamma^p \right) + \sqrt{N} \hat{\Lambda} \Gamma^p \right] = (\hat{\Sigma} - \hat{\lambda})^{-1} \left[ \frac{\hat{X}' e}{\sqrt{N}} + \frac{1}{\sqrt{N}} \left[ \begin{bmatrix} 1 \\ V' \sqrt{N} \end{bmatrix} - \frac{1}{\sqrt{N}} \left[ \begin{bmatrix} 1 \\ V' \sqrt{N} \end{bmatrix} \right] \right] \Gamma^p + \sqrt{N} \hat{\Lambda} \Gamma^p \right] = (\hat{\Sigma} - \hat{\lambda})^{-1} \left[ \frac{V'}{\sqrt{N}} + \frac{1}{\sqrt{N}} \left[ \begin{bmatrix} 1 \\ V' \sqrt{N} \end{bmatrix} \right] \Gamma^p + \sqrt{N} \hat{\Lambda} \Gamma^p \right] = (\hat{\Sigma} - \hat{\lambda})^{-1} \left[ \frac{V'}{\sqrt{N}} + \frac{1}{\sqrt{N}} \left[ \begin{bmatrix} 1 \\ V' \sqrt{N} \end{bmatrix} \right] \Gamma^p + \sqrt{N} \hat{\Lambda} \Gamma^p \right]
\]
\[
= (\hat{\Sigma} - \hat{\lambda})^{-1} \frac{V'}{\sqrt{N}} + \frac{1}{\sqrt{N}} \left[ \begin{bmatrix} 1 \\ V' \sqrt{N} \end{bmatrix} \right] \Gamma^p + \sqrt{N} \hat{\Lambda} \Gamma^p
\]
\[
= (\hat{\Sigma} - \hat{\lambda})^{-1} \left[ \frac{V'}{\sqrt{N}} + \frac{1}{\sqrt{N}} \left[ \begin{bmatrix} 1 \\ V' \sqrt{N} \end{bmatrix} \right] \Gamma^p + \sqrt{N} \hat{\Lambda} \Gamma^p \right]
\]
\[
= (\hat{\Sigma} - \hat{\lambda})^{-1} \left[ \frac{V'}{\sqrt{N}} + \frac{1}{\sqrt{N}} \left[ \begin{bmatrix} 1 \\ V' \sqrt{N} \end{bmatrix} \right] \Gamma^p + \sqrt{N} \hat{\Lambda} \Gamma^p \right]
\]
\[
= (\hat{\Sigma} - \hat{\lambda})^{-1} \left[ \frac{V'}{\sqrt{N}} + \frac{1}{\sqrt{N}} \left[ \begin{bmatrix} 1 \\ V' \sqrt{N} \end{bmatrix} \right] \Gamma^p + \sqrt{N} \hat{\Lambda} \Gamma^p \right]
\]
\[
= (\hat{\Sigma} - \hat{\lambda})^{-1} (I_1 + I_2). \tag{1.B.19}
\]

Using Lemmas 1.1 and 1.2(ii), we have
\[
(\hat{\Sigma} - \hat{\lambda}) \overset{d}{\rightarrow} \left( \begin{bmatrix} 1 & \mu_\beta \\ \mu_\beta & \sigma^2 \end{bmatrix} \right) = \Sigma_X. \tag{1.B.20}
\]

Consider now the terms \( I_1 \) and \( I_2 \). Both terms have a zero mean and, under Assumption 1.5(vi), they are asymptotically uncorrelated. Assumptions 1.2, 1.5(i), 1.6(i), and 1.6(iii) imply that
\[
\text{Var}(I_1) = E \left[ \frac{Q' \frac{1}{\sqrt{N}} \sum_{i=1}^N \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^N \epsilon_j Q}{\sqrt{\frac{1}{n} \sum_{i=1}^N (Q' \otimes \beta_i) \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^N \epsilon_j Q}} \cdot \frac{Q' \frac{1}{\sqrt{N}} \sum_{i=1}^N \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^N \epsilon_j Q}{\sqrt{\frac{1}{n} \sum_{i=1}^N (Q' \otimes \beta_i) \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^N \epsilon_j Q}} \right] = \left[ \begin{array}{cc} \sigma^2 Q' Q & \sigma^2 Q' (\otimes \mu') \\ \sigma^2 (\otimes \mu') Q & \sigma^2 (\otimes \mu') Q \end{array} \right] \rightarrow \left[ \begin{array}{cc} \sigma^2 Q' Q & \sigma^2 Q' (\otimes \mu') \\ \sigma^2 (\otimes \mu') Q & \sigma^2 (\otimes \mu') Q \end{array} \right] \right] \cdot \left( \begin{array}{c} \gamma_1^P \\ \gamma_1^P \end{array} \right) \Sigma_X. \tag{1.B.21}
\]
Next, consider $I_2$. Since $\mathcal{P}_{\frac{1}{\sqrt{N}}} \sum_{i=1}^{N} \sigma_i^2 Q + \frac{1}{T-K-1} \text{tr} \left( M \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^2 I_T^T \right) \mathcal{P}^\prime \mathcal{P}^\prime \gamma_1^P = 0_K$, we have

$$I_2 = \begin{bmatrix} (Q' \otimes P') \text{vec} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (e_i e_i' - \sigma_i^2 I_T) \right) & 0 \\ 0 & \frac{1}{T-K-1} \text{tr} \left( M \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (e_i e_i' - \sigma_i^2 I_T) \right) \mathcal{P}^\prime \mathcal{P}^\prime \gamma_1^P \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ I_{22} \end{bmatrix}.$$  \hspace{1cm} (1.B.22)

Therefore, $\text{Var}(I_2)$ has the following form:

$$\text{Var}(I_2) = \begin{bmatrix} 0 & 0_K' \\ 0_K & E[I_{22}I_{22}'] \end{bmatrix}. \hspace{1cm} (1.B.23)$$

Under Assumptions 1.5(i) and 1.6(ii), we have

$$E[I_{22}I_{22}'] = E \left[ (Q' \otimes P') \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(e_i e_i' - \sigma_i^2 I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \text{vec}(e_j e_j' - \sigma_j^2 I_T)'(Q \otimes P) \right]$$

$$+ E \left[ (Q' \otimes P') \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(e_i e_i' - \sigma_i^2 I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \text{vec}(e_j e_j' - \sigma_j^2 I_T)' \frac{\text{vec}(M)}{T-K-1} \gamma_1^P \mathcal{P}^\prime \mathcal{P}^\prime \gamma_1^P \right]$$

$$+ E \left[ \mathcal{P}^\prime \mathcal{P}^\prime \gamma_1^P \frac{\text{vec}(M)}{T-K-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(e_i e_i' - \sigma_i^2 I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \text{vec}(e_j e_j' - \sigma_j^2 I_T)'(Q \otimes P) \right]$$

$$+ E \left[ \mathcal{P}^\prime \mathcal{P}^\prime \gamma_1^P \frac{\text{vec}(M)}{T-K-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(e_i e_i' - \sigma_i^2 I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \text{vec}(e_j e_j' - \sigma_j^2 I_T)' \frac{\text{vec}(M)}{T-K-1} \times \gamma_1^P \mathcal{P}^\prime \mathcal{P}^\prime \gamma_1^P \right]$$

$$\rightarrow \left[ (Q' \otimes P') + \mathcal{P}^\prime \mathcal{P}^\prime \gamma_1^P \frac{\text{vec}(M)}{T-K-1} U_e \left[ (Q \otimes P) + \frac{\text{vec}(M)}{T-K-1} \gamma_1^P \mathcal{P}^\prime \mathcal{P}^\prime \right] \right]. \hspace{1cm} (1.B.24)$$

Defining $Z = \left[ (Q \otimes P) + \frac{\text{vec}(M)}{T-K-1} \gamma_1^P \mathcal{P}^\prime \mathcal{P}^\prime \right]$ concludes the proof of part (ii). 

**Proof of Theorem 1.2.** By Theorem 1.1(i), $\hat{\gamma}_T^* \rightarrow_p \gamma_T^*$. Lemma 1.1 implies that $\hat{\Lambda}$ is a consistent estimator of $\Lambda$. Hence, using Lemma 1.2(ii), we have $\left( \hat{\Sigma}_X - \hat{\Lambda} \right) \rightarrow_p \Sigma_X$, which implies that $\hat{V} \rightarrow_p V$. A consistent estimator of $W$ requires a consistent estimate of the matrix $U_e$, which can be obtained using Lemma 1.6. This concludes the proof of Theorem 1.2. \(\blacksquare\)
Proof of Theorem 1.3. Writing

\[
\begin{align*}
(\hat{\Sigma}_X - \hat{\lambda})^{-1} X^t R_t N^{-1} &= (\hat{\Sigma}_X - \hat{\lambda})^{-1} \hat{\Sigma}_X \hat{\Gamma}_{t-1}^p + (\hat{\Sigma}_X - \hat{\lambda})^{-1} \hat{X}' \hat{\epsilon}'_{t,T} + (\hat{\Sigma}_X - \hat{\lambda})^{-1} \hat{X}'(X - \hat{X}) \hat{\Gamma}_{t-1}^p \\
&= \hat{\Gamma}_{t-1}^p + (\hat{\Sigma}_X - \hat{\lambda})^{-1} \left( \frac{\hat{X}' \hat{\epsilon}'_{t,T}}{N} + \frac{\hat{X}'(X - \hat{X})}{N} \hat{\Gamma}_{t-1}^p + \hat{\lambda} \hat{\Gamma}_{t-1}^p \right) \\
&= \hat{\Gamma}_{t-1}^p + (\hat{\Sigma}_X - \hat{\lambda})^{-1} \left( \frac{\hat{X}' \hat{\epsilon}'_{t,T}}{N} + \frac{(\hat{X} - X)' \hat{\epsilon}'_{t,T}}{N} + \frac{\hat{X}'(X - \hat{X})}{N} \hat{\Gamma}_{t-1}^p + \hat{\lambda} \hat{\Gamma}_{t-1}^p \right) \\
&= \hat{\Gamma}_{t-1}^p + (\hat{\Sigma}_X - \hat{\lambda})^{-1} \left( \left[ \begin{array}{c} Y_N' \\ B' \end{array} \right] \frac{\hat{\epsilon}'_{t,T}}{N} + \frac{1}{N} \left[ \begin{array}{c} 0' \\ P' \end{array} \right] \right) \left[ \begin{array}{c} Y_N \\ B \\
- \frac{1}{N} \frac{-Y_N' e' P \gamma_{t-1}^p}{B' e' P \gamma_{t-1}^p} + \frac{\hat{\lambda}}{\hat{\Gamma}_{t-1}^p} \right] \\
&= \hat{\Gamma}_{t-1}^p + (\hat{\Sigma}_X - \hat{\lambda})^{-1} \left( \left[ \begin{array}{c} \frac{Y_N' e' Q_{t-1}}{B' e' Q_{t-1}} \\ \frac{0'}{N} \right] \hat{\Gamma}_{t-1}^p \right) + \hat{\Gamma}_{t-1}^p + \frac{0'}{N} \left[ \begin{array}{c} 0 \\ P' \end{array} \right] \\
&= \hat{\Gamma}_{t-1}^p + (\hat{\Sigma}_X - \hat{\lambda})^{-1} \left( \left[ \begin{array}{c} \frac{Y_N' e' Q_{t-1}}{B' e' Q_{t-1}} \\ \frac{0'}{N} \right] \hat{\Gamma}_{t-1}^p \right) + \hat{\Gamma}_{t-1}^p + \frac{0'}{N} \left[ \begin{array}{c} 0 \\ P' \end{array} \right] \\
&= \hat{\Gamma}_{t-1}^p + (\hat{\Sigma}_X - \hat{\lambda})^{-1} \left( \left[ \begin{array}{c} \frac{Y_N' e' Q_{t-1}}{B' e' Q_{t-1}} \\ \frac{0'}{N} \right] \hat{\Gamma}_{t-1}^p \right) + \hat{\Gamma}_{t-1}^p + \frac{0'}{N} \left[ \begin{array}{c} 0 \\ P' \end{array} \right]
\end{align*}
\]

with

\[
E \left( \left[ \begin{array}{c} 0' \\ P' \end{array} \right] + \hat{\Gamma}_{t-1}^p \right) = E \left( \left[ \begin{array}{c} 0' \\ P' \end{array} \right] + \frac{\hat{\lambda}(M' e')}{N(T - K - 1)} \left[ \begin{array}{c} 0' \\ P' \end{array} \right] \right) = 0_{K+1}
\]

and

\[
\left[ \begin{array}{c} \frac{P' e' t_t}{N} \gamma_{1,t-1}^p \right] \rightarrow_p \left[ \begin{array}{c} 0' \sigma^2 P' \gamma_{1,t}^p \\ \sigma^2 \frac{P' t_t}{N} \gamma_{1,t}^p \right] = \left[ \begin{array}{c} 0' \\ 0' \end{array} \right]
\]

yields part (i).

Next,

\[
\hat{\Gamma}_{t-1}^* = (\hat{\Sigma}_X - \hat{\lambda})^{-1} \frac{\hat{X}' R_t}{N} - (\hat{\Sigma}_X - \hat{\lambda})^{-1} \left[ \frac{0'}{\sigma^2 P' \gamma_{t,t}^p} \right]
\]

\[
= \frac{\hat{\Gamma}_{t-1}^p}{\hat{\Gamma}_{t-1}^p} + (\hat{\Sigma}_X - \hat{\lambda})^{-1} \left( \left[ \begin{array}{c} \frac{Y_N' e' Q_{t-1}}{B' e' Q_{t-1}} \\ \frac{0'}{N} \right] \hat{\Gamma}_{t-1}^p \right) - \frac{\hat{\lambda}}{\hat{\Gamma}_{t-1}^p} \left[ \begin{array}{c} 0' \\ P' \end{array} \right] \\
&= (\hat{\Sigma}_X - \hat{\lambda})^{-1} \hat{X}' R_t N^{-1} - (\hat{\Sigma}_X - \hat{\lambda})^{-1} \frac{0'}{\sigma^2 P' \gamma_{t,t}^p}.
\]

The part of \( \sqrt{N} (\hat{\Gamma}_{t-1}^* - \hat{\Gamma}_{t-1}^p) \) that depends on \( \epsilon' \) can be written as

\[
(\hat{\Sigma}_X - \hat{\lambda})^{-1} \left( (Q_{t-1}' - \frac{P'}{\sigma^2} \hat{Q}_{t-1} \text{vec}(M)^{\prime} \text{vec} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon' \epsilon'_i - \sigma_i^2 I_T) \right) \\ = (\hat{\Sigma}_X - \hat{\lambda})^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon' \epsilon'_i - \sigma_i^2 I_T),
\]

and the result follows along the proof of Theorem 1.1(ii).
1.B. PROOFS OF PROPOSITIONS AND THEOREMS

Proof of Theorem 1.4. We first establish a simpler, asymptotically equivalent, expression for \( \sqrt{N} \left( \frac{\hat{e}^P}{N} - \hat{\sigma}^2 \hat{Q}' \hat{Q} \right) \).

Then, we derive the asymptotic distribution of this approximation. Consider the sample ex post pricing errors,

\[
\hat{e}^P = \hat{R} - \hat{X} \hat{\Gamma}^*.
\]  

(1.B.30)

Starting from \( \hat{R} = \hat{X} \hat{\Gamma}^P + \eta^P \) with \( \eta^P = \bar{e} - (\hat{X} - X)\hat{\Gamma}^P \), we have

\[
\hat{e}^P = \hat{X} \Gamma^P + \bar{e} - (\hat{X} - X)\Gamma^P - \hat{X} \hat{\Gamma}^* \\
= \bar{e} - \hat{X} (\hat{\Gamma}^* - \Gamma^P) - (\hat{X} - X) \Gamma^P.
\]  

(1.B.31)

Then,

\[
\hat{e}^P \hat{e}^P = \bar{e} \bar{e} + \Gamma^P (\hat{X} - X)'(\hat{X} - X) \Gamma^P - 2(\hat{\Gamma}^* - \Gamma^P)' \bar{e} \bar{e} - 2\Gamma^P (\hat{X} - X)' \bar{e} \\
+ 2\Gamma^P (\hat{X} - X)' \hat{X} (\hat{\Gamma}^* - \Gamma^P) + (\hat{\Gamma}^* - \Gamma^P)' \hat{X} \hat{X} (\hat{\Gamma}^* - \Gamma^P).
\]

Note that

\[
\frac{\bar{e} \bar{e}}{N} = \frac{1}{T^2} \mathbf{1}' \frac{\epsilon \epsilon'}{N} \mathbf{1} \to_p \frac{\sigma^2}{T}.
\]  

(1.B.32)

and, by Lemma 1.2(iii),

\[
\Gamma^P (\hat{X} - X)'(\hat{X} - X) \Gamma^P - \gamma^P (\hat{\Gamma}^* - \Gamma^P)' \gamma^P - \gamma^P (\hat{\Gamma}^* - \Gamma^P) = o_p \left( \frac{\sigma^2}{T} \right).
\]

(1.B.33)

Using Lemmas 1.3 and 1.5 and Theorem 1.1, we have

\[
\frac{(\hat{\Gamma}^* - \Gamma^P)' \bar{e} \bar{e}}{N} = \frac{(\hat{\Gamma}^* - \Gamma^P)' \bar{e} \bar{e}}{N} + \frac{(\hat{\Gamma}^* - \Gamma^P)' X' \bar{e}}{N} = o_p \left( \frac{1}{N} \right)
\]  

(1.B.34)

and

\[
\frac{\Gamma^P (\hat{X} - X)' \bar{e}}{N} = o_p \left( \frac{1}{\sqrt{N}} \right).
\]  

(1.B.35)

In addition, using Lemmas 1.2(i), 1.2(iii), 1.4 and Theorem 1.1, we have

\[
\frac{\Gamma^P (\hat{X} - X)' (\hat{\Gamma}^* - \Gamma^P)}{N} = \frac{\Gamma^P (\hat{X} - X)' (\hat{\Gamma}^* - \Gamma^P)}{N} + \frac{\Gamma^P (\hat{X} - X)' X (\hat{\Gamma}^* - \Gamma^P)}{N} = o_p \left( \frac{1}{\sqrt{N}} \right) + o_p \left( \frac{1}{N} \right)
\]  

(1.B.36)

and

\[
\frac{(\hat{\Gamma}^* - \Gamma^P)' \hat{X}' \hat{X} (\hat{\Gamma}^* - \Gamma^P)}{N} = o_p \left( \frac{1}{N} \right).
\]  

(1.B.37)

It follows that

\[
\frac{\hat{e}^P \hat{e}^P}{N} \to_p \frac{\sigma^2}{T} + \sigma^2 \Gamma^P (\hat{\Gamma}' \hat{\Gamma})^{-1} \Gamma^P = \sigma^2 Q'Q.
\]  

(1.B.38)
Collecting terms and rewriting explicitly only the ones that are $O_p\left(\frac{1}{\sqrt{N}}\right)$, we have

\[
\frac{\hat{\varepsilon}^P\hat{\varepsilon}^P}{N} = \frac{\bar{\varepsilon}\bar{\varepsilon}}{N} + \frac{\Gamma^P(\hat{X} - X)'(\hat{X} - X)\Gamma^P}{N} + \frac{2\Gamma^P(\hat{X} - X)'\bar{\varepsilon}}{N} + O_p\left(\frac{1}{N}\right),
\]

Consider the sum of the three terms in Eqs. (1.39)-(1.41). Under Assumption 1.5(i), we have

\[
\frac{\bar{\varepsilon}\bar{\varepsilon}}{N} + \frac{\Gamma^P(\hat{X} - X)'(\hat{X} - X)\Gamma^P}{N} + 2\frac{\Gamma^P(\hat{X} - X)'\bar{\varepsilon}}{N} = \frac{1_T\hat{\varepsilon}\hat{\varepsilon}}{T N} - \frac{\gamma_1 P'\hat{\varepsilon}\hat{\varepsilon}'}{N} + \frac{1_T\hat{\varepsilon}\hat{\varepsilon}'}{T N} - \frac{1_T P'\hat{\varepsilon}\hat{\varepsilon}'}{N} + \gamma_1 P' P' P\gamma_1 - 2 \frac{1_T P'\hat{\varepsilon}\hat{\varepsilon}'}{T N} P\gamma_1
\]

\[
= \frac{1_T\hat{\varepsilon}\hat{\varepsilon}}{T N} - \frac{1_T P'\hat{\varepsilon}\hat{\varepsilon}'}{T N} Q - \frac{1_T P'\hat{\varepsilon}\hat{\varepsilon}'}{T N} P\gamma_1
\]

\[
= Q'\hat{\varepsilon}\hat{\varepsilon} - Q'\hat{\varepsilon}\hat{\varepsilon}'P\gamma_1
\]

\[
= Q'\hat{\varepsilon}\hat{\varepsilon} - Q'\hat{\varepsilon}\hat{\varepsilon} - \frac{1}{N}\bar{\sigma}^2 I_T
\]

\[
= Q'\left(\hat{\varepsilon}\hat{\varepsilon} - \bar{\sigma}^2 I_T\right)Q + \sigma^2 Q'Q + o\left(\frac{1}{\sqrt{N}}\right),
\]

where the $o\left(\frac{1}{\sqrt{N}}\right)$ term comes from $(\bar{\sigma}^2 - \sigma^2)Q'Q$. As for the term in Eq. (1.42), define

\[
\left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} = \begin{bmatrix}
\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\
\hat{\Sigma}_{21} & \hat{\Sigma}_{22}
\end{bmatrix},
\]

where every block of $\left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1}$ is $O_p(1)$ by the nonsingularity of $\Sigma_X$ and Slutsky's theorem. Using the
same arguments as for Theorem 1.2, we have
\[
\begin{align*}
2 \frac{\Gamma^{P^*}(\tilde{X} - X)'(\tilde{X} - X)(\hat{\Sigma}^* - \Gamma^P)}{N} &= 2 \left[ \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{21} + \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} \right] \left[ \frac{\epsilon'_N}{N} + Z' \text{vec} \left( \epsilon'_N - \sigma^2 I_T \right) \right] \\
&= 2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{21} + 2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} \frac{B'}{N} \\
&+ 2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} Z' \text{vec} \left( \epsilon'_N - \sigma^2 I_T \right) \\
&+ 2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{21} \frac{B'}{N} \\
&+ 2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} Z' \text{vec} \left( \epsilon'_N - \sigma^2 I_T \right) + o_p \left( \frac{1}{N} \right) \\
&= 2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{21} + 2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} \frac{B'}{N} \\
&+ 2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} Z' \text{vec} \left( \epsilon'_N - \sigma^2 I_T \right) + o_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{N} \right). \tag{1.B.46}
\end{align*}
\]

where the two approximations on the right-hand side of the previous expression refer to
\[
2(\tilde{P} - \sigma^2) \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{21} \frac{B'}{N} + 2(\tilde{P} - \sigma^2) \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} \frac{B'}{N} + 2(\tilde{P} - \sigma^2) \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} Z' \text{vec} \left( \epsilon'_N - \sigma^2 I_T \right) = o_p \left( \frac{1}{N} \right) \tag{1.B.47}
\]

and
\[
2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{21} + 2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} \frac{B'}{N} + 2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} Z' \text{vec} \left( \epsilon'_N - \sigma^2 I_T \right) = o_p \left( \frac{1}{N} \right), \tag{1.B.48}
\]

respectively. Therefore, we have
\[
\frac{\hat{\epsilon}^P \hat{\epsilon}^{P^*}}{N} = Q' \left( \frac{\epsilon'_N}{N} - \sigma^2 I_T \right) Q + \sigma^2 Q'Q \\
+ 2 \sigma^2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{21} \frac{B'}{N} + 2 \sigma^2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} \frac{B'}{N} \\
+ 2 \sigma^2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} Z' \text{vec} \left( \epsilon'_N - \sigma^2 I_T \right) + o_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{N} \right) + o \left( \frac{1}{\sqrt{N}} \right). \tag{1.B.49}
\]

It follows that
\[
\frac{\hat{\epsilon}^P \hat{\epsilon}^{P^*}}{N} - \sigma^2 Q'Q = Q' \left( \frac{\epsilon'_N}{N} - \sigma^2 I_T \right) Q - \left( \hat{\epsilon}^P Q'Q - \sigma^2 Q'Q \right) \\
+ 2 \sigma^2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{21} \frac{B'}{N} + 2 \sigma^2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} \frac{B'}{N} \\
+ 2 \sigma^2 \gamma_1^{\tilde{P}^*} p \left( \epsilon'_N - \sigma^2 I_T \right) \hat{\Sigma}^{22} Z' \text{vec} \left( \epsilon'_N - \sigma^2 I_T \right) + o_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{N} \right) + o \left( \frac{1}{\sqrt{N}} \right). \tag{1.B.50}
\]
Note that
\[ \tilde{\sigma}^2 \tilde{Q}' \tilde{Q} - \sigma^2 Q'Q = \frac{1}{T} (\tilde{\sigma}^2 - \sigma^2) + \tilde{\sigma}^2 \tilde{\gamma}_1'(\tilde{F}' \tilde{F})^{-1} \tilde{\gamma}_1 - \sigma^2 \gamma_1' (\tilde{F}' \tilde{F})^{-1} \gamma_1' \]
\[ = \frac{1}{T} (\tilde{\sigma}^2 - \sigma^2) + (\tilde{\sigma}^2 - \sigma^2) \gamma_1' (\tilde{F}' \tilde{F})^{-1} \gamma_1' + 2 \sigma^2 (\gamma_1' - \gamma_1')' (\tilde{F}' \tilde{F})^{-1} \gamma_1' + O_p \left( \frac{1}{N} \right) 
\]
\[ = (\tilde{\sigma}^2 - \sigma^2) \left( \frac{1}{T} + \gamma_1' (\tilde{F}' \tilde{F})^{-1} \gamma_1' \right) + 2 \sigma^2 (\gamma_1' - \gamma_1')' (\tilde{F}' \tilde{F})^{-1} \gamma_1' + O_p \left( \frac{1}{N} \right) 
\]
\[ = (\tilde{\sigma}^2 - \sigma^2) \left( \frac{1}{T} + \gamma_1' (\tilde{F}' \tilde{F})^{-1} \gamma_1' \right) + 2 \sigma^2 \gamma_1' P' \tilde{P} \Sigma_{21} \frac{1}{N} + 2 \sigma^2 \gamma_1' P' \tilde{P} \Sigma_{22} \frac{B' \epsilon' Q}{N \sqrt{N}} \right) + O_p \left( \frac{1}{N} \right) + O_p \left( \frac{1}{N \sqrt{N}} \right), \quad (1.B.51) \]

where \( \sigma^2 (\gamma_1' - \gamma_1')' (\tilde{F}' \tilde{F})^{-1} (\gamma_1' - \gamma_1') + 2 (\tilde{\sigma}^2 - \sigma^2) (\gamma_1' - \gamma_1')' (\tilde{F}' \tilde{F})^{-1} \gamma_1' = O_p \left( \frac{1}{N} \right) \) and \( (\tilde{\sigma}^2 - \sigma^2) (\gamma_1' - \gamma_1')' (\tilde{F}' \tilde{F})^{-1} (\gamma_1' - \gamma_1') = O_p \left( \frac{1}{N \sqrt{N}} \right) \). It follows that
\[
\frac{\tilde{\sigma}^2 \tilde{Q}'}{N} - \tilde{\sigma}^2 \tilde{Q}' \tilde{Q} = \tilde{Q}' \left( \frac{\tilde{\epsilon}' \epsilon}{N} - \tilde{\sigma}^2 I_T \right) \tilde{Q} - (\tilde{\sigma}^2 - \sigma^2) \left( \frac{1}{T} + \gamma_1' (\tilde{F}' \tilde{F})^{-1} \gamma_1' \right) + O_p \left( \frac{1}{N \sqrt{N}} \right) + O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{\sqrt{N}} \right) 
\]
\[ = \left[ (Q' \otimes \tilde{Q}') - \frac{Q'Q}{T - K - 1} \text{vec}(M)' \right] \text{vec} \left( \frac{\tilde{\epsilon}' \epsilon}{N} - \tilde{\sigma}^2 I_T \right) + O_p \left( \frac{1}{\sqrt{N}} \right) 
\]
\[ = Z_Q \text{vec} \left( \frac{\tilde{\epsilon}' \epsilon}{N} - \tilde{\sigma}^2 I_T \right) + o_p \left( \frac{1}{\sqrt{N}} \right), \quad (1.B.52) \]

where, for simplicity, we have condensed \( O_p \left( \frac{1}{N \sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{N}} \right) + o_p \left( \frac{1}{\sqrt{N}} \right) \) into the single term \( o_p \left( \frac{1}{\sqrt{N}} \right) \). Hence,
\[
\sqrt{N} \left( \frac{\tilde{\sigma}^2 \tilde{Q}'}{N} - \tilde{\sigma}^2 \tilde{Q}' \tilde{Q} \right) = \sqrt{N} Z_Q \text{vec} \left( \frac{\tilde{\epsilon}' \epsilon}{N} - \tilde{\sigma}^2 I_T \right) + o_p (1), \quad (1.B.53) \]

implying that the asymptotic distribution of \( \sqrt{N} \left( \frac{\tilde{\sigma}^2 \tilde{Q}'}{N} - \tilde{\sigma}^2 \tilde{Q}' \tilde{Q} \right) \) is equivalent to the asymptotic distribution of \( \sqrt{N} Z_Q \text{vec} \left( \frac{\tilde{\epsilon}' \epsilon}{N} - \tilde{\sigma}^2 I_T \right) \). Finally, by Assumption 1.6(ii), we have
\[
\sqrt{N} Z_Q \text{vec} \left( \frac{\tilde{\epsilon}' \epsilon}{N} - \tilde{\sigma}^2 I_T \right) \rightarrow_d N \left( 0, Z'_Q U Z_Q \right). \quad (1.B.54) \]

**Proof of Theorem 1.5.** For part (i), in view of Eq. (1.65), we obtain \( \tilde{R} = X \tilde{\Gamma}' + e + \tilde{\epsilon} \), where \( \tilde{\Gamma}' = \Gamma + \tilde{f} - E[f_t] \). Using the same arguments as for Theorem 1.1,
\[
\tilde{\epsilon}' = \tilde{\Gamma}' - \tilde{\Gamma}' = \left( \tilde{X}' \tilde{X}' - \tilde{\Lambda}' \right)^{-1} \left[ \tilde{X}' \tilde{X}' - \left( \tilde{X}' \tilde{X}' - \tilde{\Lambda}' \right)^{-1} \tilde{X}' \tilde{X}' \right] \tilde{\Gamma}' + \tilde{X}' e 
\]
\[ = \tilde{X}' e \frac{X'e}{N} + \left( \tilde{X}' - X \right)' e \frac{1}{N} = o_{K+1} \left( \frac{1}{N} \right) \] (1.B.55)
\[ \sqrt{N} \left( \hat{\Gamma}^* - \hat{\Gamma}^P \right) = \left( \Sigma_X - \hat{\Lambda} \right)^{-1} \left( \frac{1}{\sqrt{N}} \hat{\Sigma}_X + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \ell_i \epsilon_i e_j \right) \]

\[ = \left( \Sigma_X - \hat{\Lambda} \right)^{-1} (I_1 + I_2 + I_3). \quad (1.57) \]

As for terms \( I_1 \) and \( I_2 \), Theorem 1.1 applies, that is, \( \hat{\Sigma}_X - \hat{\Lambda} \rightarrow_p \Sigma_X \), \( \text{Var}(I_1) = \frac{\tau_\Phi}{T} \left[ 1 + \gamma_1 P \left( \hat{\gamma}_e P \right)^{-1} \gamma_1 P \right] \Sigma_X \) and \( \text{Var}(I_2) = 0 \), with \( E[I_{22} I_{22}^T] = \left( \left( Q' \otimes P' \right) + P' P \hat{\gamma}_e P \frac{\gamma_1 P}{T - K - 1} \right) U_c \left( \left( Q' \otimes P \right) + \frac{\gamma_1 P}{T - K - 1} \gamma_1 P \right), \)

where \( \text{Cov}(I_1, I_2) = 0 \) by the assumption of zero third moment of the error term. Using Lemmas 1.8 and 1.9, the proof of part (ii) becomes very similar to the proof of Theorem 1.2 and is omitted.

**Proof of Theorem 1.6.** For part (i), rewrite

\[ \left[ \hat{\delta}^* \right] = \left[ \hat{\delta}^P \right] + \left[ \hat{\hat{X}}^C \hat{X} - \hat{\Lambda} \hat{X}^C \right]^{-1} \left[ \left[ \hat{\Lambda} \right] P \right] + \left[ \hat{\hat{X}}^C \right] \left( \tau + (X - \hat{X}) \hat{\Gamma}^P \right). \]

As for the bias associated with \( \hat{\Gamma}^* \) (see the proof of Theorem 1.1), we have

\[ \hat{\Lambda}^P + \frac{1}{N} \hat{\hat{X}}^C (\tau + (X - \hat{X}) \hat{\Gamma}^P) = O_p(N^{-1/2}). \quad (1.60) \]

As for the bias associated with \( \hat{\delta}^* \), we have

\[ \frac{1}{N} C' (\tau + (X - \hat{X}) \hat{\Gamma}^P) = \frac{1}{N} C' \hat{\gamma}_e \left( \frac{1}{T} - P \gamma_1 P \right) = \frac{1}{N} C' \hat{\gamma}_e Q = O_p(N^{-1/2}) \quad (1.61) \]

since \( N^{-1} C' \hat{\gamma}_e \rightarrow_p 0_{K \times T} \) and

\[ \text{Var} \left( \frac{1}{N} C' \hat{\gamma}_e Q \right) = (Q' \otimes I_K) \frac{1}{N} \sum_{i,j=1}^{N} \sigma_{22, ij} (Q \otimes I_K) = \frac{1}{N^2} (Q' \otimes I_K) \sum_{i,j=1}^{N} \sigma_{ij} (I_T \otimes \epsilon_i \epsilon_j) (Q \otimes I_K) \]

\[ = \frac{1}{N^2} \sum_{i,j=1}^{N} \sigma_{ij} (Q' \Sigma \epsilon_i \epsilon_j) = \frac{1}{N^2} \sum_{i,j=1}^{N} \sigma_{ij} (Q' \Sigma \epsilon_i \epsilon_j) = \frac{1}{N^2} \sum_{i,j=1}^{N} \sigma_{ij} (Q' \Sigma \epsilon_i \epsilon_j) = o \left( \frac{1}{N} \right) \quad (1.62) \]

by Assumption 1.8.
For part (ii), by straightforward generalizations of Lemmas 1.1 and 1.2(ii), we have

\[
\frac{1}{N} \left[ \hat{X}' \hat{X} - N \hat{\Lambda} \right] = \Sigma_{CC}^{-1} \left[ \mu'_{c} \Sigma_{CB} \right] = L. \tag{1.63}
\]

We now prove that \( L \) is positive-definite. Using the blockwise formula for the inverse of a matrix, the invertibility of \( L \) follows from \( \Sigma_{CC} \) being positive-definite (see Assumption 1.8(i)) and the invertibility of

\[
\begin{bmatrix}
\frac{1}{\mu'_{c}} - \frac{1}{\Sigma_{CB}} \\
\mu_{c} - \Sigma_{CB}
\end{bmatrix}
\]

in turn, this holds if

\[
D = \Sigma_{\beta} - \Sigma_{CB}^{-1} \Sigma_{CB}^{-1} \Sigma_{CB}
\]

is positive-definite and

\[
1 - \mu'_{c} \Sigma_{CB}^{-1} \mu_{c} - (\mu'_{\beta} - \mu'_{c} \Sigma_{CB}^{-1} \Sigma_{CB}) D^{-1} (\mu_{\beta} - \Sigma_{CB}^{-1} \mu_{c})
\]

is nonzero. The last equation can be rewritten as

\[
1 - |\mu'_{c} \mu'_{\beta}| \begin{bmatrix} \Sigma_{CC}^{-1} & \Sigma_{CB}^{-1} \\ \Sigma_{CB}^{-1} & \Sigma_{\beta}^{-1} \end{bmatrix} \begin{bmatrix} \mu_{c} \\ \mu_{\beta} \end{bmatrix}.
\]

The positiveness of Eq. (1.66) and the positive-definiteness of \( D \) follow from Assumption 1.8(i). Next, following the proof of Theorem 1.1,

\[
\sqrt{N} \begin{bmatrix}
\hat{\Gamma}^{*} - \Gamma^{P}
\hat{\delta}^{*} - \delta
\end{bmatrix} = \begin{bmatrix}
\hat{X}' \hat{X} - \hat{\Lambda} \\
\hat{\delta}' \hat{\delta}
\end{bmatrix}^{-1}
\times \begin{bmatrix}
\hat{X}' \hat{Q}' \\
\hat{\delta}' \hat{Q}
\end{bmatrix} + \begin{bmatrix}
0 \\
\frac{\nu(M_{c})}{\sqrt{N(N-K-1)}}
\end{bmatrix} + \begin{bmatrix}
0 \\
\frac{\nu}{\sqrt{N} \hat{Q}}
\end{bmatrix} (I_{1} + I_{2} + I_{3}). \tag{1.67}
\]

We now derive \( \text{Var}(I_{3}) \) and \( \text{Cov}(I_{1}, I_{3}') \) because the other terms can be directly obtained from Theorem 1.1 and \( \text{Cov}(I_{2}, I_{3}') = 0_{K+K_{c}-1 \times (K+K_{c}-1)} \). We have

\[
\text{Var}(I_{3}) = \begin{bmatrix}
0_{(K+1) \times (K+1)} \\
0_{K_{c} \times (K+1)}
\end{bmatrix} \begin{bmatrix}
0_{(K+1) \times (K_{c} \times K_{c})} \\
0_{K_{c} \times (K+1)}
\end{bmatrix} \begin{bmatrix}
0_{(K+1) \times (K+1)} \\
0_{K_{c} \times (K+1)}
\end{bmatrix}
\]

and, by Theorem 1.1,

\[
\text{Cov}(I_{1}, I_{3}') = \begin{bmatrix}
0_{(K+1) \times (K+1)} \\
0_{K_{c} \times (K+1)}
\end{bmatrix} \begin{bmatrix}
\hat{\sigma}^{2} \hat{Q}' \hat{Q} \Sigma_{CC}^{-1} \\
\hat{\sigma}^{2} \hat{Q}' \hat{Q} \Sigma_{CB}^{-1}
\end{bmatrix} \begin{bmatrix}
0_{(K+1) \times (K+1)} \\
0_{K_{c} \times (K+1)}
\end{bmatrix}.
\]
1.C. Explicit Form of $U_e$

Denote by $U_e$ the $T^2 \times T^2$ matrix

$$
U_e = 
\begin{pmatrix}
U_{11} & \cdots & U_{1T} & \cdots & U_{1T} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
U_{t1} & \cdots & U_{tt} & \cdots & U_{tT} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
U_{T1} & \cdots & U_{TT} & \cdots & U_{TT}
\end{pmatrix}.
$$

(1.C.1)

Each block of $U_e$ is a $T \times T$ matrix. The blocks along the main diagonal, denoted by $U_{tt}$, $t = 1, 2, \ldots, T$, are themselves diagonal matrices, with $(\kappa_4 + 2\sigma_4)$ in the $(t, t)$-th position and $\sigma_4$ in the $(s, s)$ position for every $s \neq t$; that is,

$$
U_{tt} = 
\begin{pmatrix}
\sigma_4 & \cdots & 0 & \cdots & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \sigma_4 & 0 & \cdots & 0 \\
0 & \cdots & 0 & (\kappa_4 + 2\sigma_4) & \cdots & 0 \\
0 & \cdots & 0 & \cdots & \sigma_4 & \cdots \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \sigma_4
\end{pmatrix}.
$$

(1.C.2)

The blocks outside the main diagonal, denoted by $U_{ts}$, $s, t = 1, 2, \ldots, T$ with $s \neq t$, are all made of zeros except for the $(s, t)$-th position that contains $\sigma_4$, that is,

$$
U_{ts} = 
\begin{pmatrix}
0 & \cdots & 0 & \cdots & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & \sigma_4 & 0 & \cdots \\
0 & \cdots & 0 & 0 & 0 & \cdots \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & 0
\end{pmatrix}.
$$

(1.C.3)

Under Assumption 1.5, by Lemma 1.6, it is easy to show that $\hat{U}_e$ in Theorem 1.2 is a consistent plug-in estimator of $U_e$ that only depends on $\hat{\sigma}_4$.

1.D. Random Betas

In this section, we discuss the modifications of the analysis that are necessary to accommodate random betas\footnote{Dealing with random betas requires a different specification of the sampling scheme (see, for example Gagliardini et al. (2016)). For simplicity, we do not provide the full details of the analysis here.}. First, consider the case where the random betas are mutually independent of the innovations in
individual asset returns. In this scenario, the asymptotic properties of the bias-adjusted Shanken (1992) estimator, \( \hat{\Gamma}^* \), are not affected. The only changes involve Assumptions 1.1 and 1.2. In particular, Eq. (1.3) in Assumption 1.1 must be replaced with \( E[R_t | X] = X\beta \). Moreover, Eqs. (1.17) and (1.18) in Assumption 1.2 must be stated in terms of convergence in probability, instead of conventional convergence, which is applicable to non-random sequences only. All the other assumptions remain unchanged, except that now Eq. (1.38) involves random betas. As \( N \to \infty \), we have

\[
\lim \text{Var} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( C_T^* \otimes [1, \beta' i] \right) \epsilon_i \right) = \lim \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} E \left[ (C_T^* \otimes [1, \beta' i]) \epsilon_i \epsilon_j' (C_T \otimes [1, \beta' j]) \right] = (C_T^*C_T) \lim \frac{1}{N} \sum_{i,j=1}^{N} \sigma_{ij} E \left[ 1, \beta' i \right] \right] = (C_T^*C_T) \sigma^2 \Sigma_X.
\]

(1.D.4)

The second term in the second line of Eq. (1.D.4) converges to zero under our assumptions since \( E[\|\beta_i \beta' j\| \leq E(\|\beta_i \beta' j\| \leq C < \infty \) and \( \sum_{i,j=1}^{N} |\sigma_{ij}| = O(N) \). Eq. (1.D.4) coincides with the asymptotic covariance matrix in Eq. (1.38), which holds for non-random \( \beta_i \).

Consider now the case in which the \( \beta_i \) are potentially cross-sectionally correlated with the \( \epsilon_i \). When \( T \) is fixed, such covariance structure cannot be identified based on the OLS estimators \( \hat{\beta}_i \) and \( \hat{\epsilon}_i \) (either for a finite or an arbitrarily large \( N \)). Therefore, the possibility of cross-correlation between the \( \beta_i \) and the \( \epsilon_i \) needs to be ruled out. By inspection of the proof of Theorem 1.1, the asymptotic covariance of \( \sqrt{N}(\hat{\Gamma}^* - \Gamma^*) \) depends on, among other things, \( N^{-1/2} \sum_{i=1}^{N} \beta_i \epsilon_i' Q \), where \( Q = \frac{1}{T} - \frac{1}{T} I_T \). Letting the \( K \)-vector \( \omega_i = E[\beta_i \epsilon_i] = \text{Cov}(\beta_i, \epsilon_i) \), we have

\[
E[\beta_i \epsilon_i'] = \Omega_i = \omega_i 1_T,' \]

(1.D.5)

where the second equality follows from the i.i.d. assumption over time for the \( \epsilon_i \) (see Assumption 1.3). Then,

\[
1 \sqrt{N} \sum_{i=1}^{N} \beta_i \epsilon_i' Q = 1 \sqrt{N} \sum_{i=1}^{N} (\beta_i \epsilon_i' - \Omega_i) Q + 1 \sqrt{N} \sum_{i=1}^{N} \Omega_i Q.
\]

(1.D.6)

A straightforward generalization of Assumption 1.6 (iii) implies that the first term of Eq. (1.6) converges to a normal distribution as \( N \to \infty \). Given Eq. (1.6), the second term of Eq. (1.6) can be re-written as \( \sqrt{N} \sum_{i=1}^{N} \omega_i 1_T \sim \sqrt{N} \sum_{i=1}^{N} \omega_i 1_T Q = \sqrt{N} \sum_{i=1}^{N} \omega_i Q \) because \( 1_T Q = 1 \). As we show below, this latter term cannot be consistently estimated by OLS when \( T \) is fixed. Therefore, in order to avoid lack of identification in the asymptotic covariance of \( \hat{\Gamma}^* \), the \( \omega_i \) must satisfy the restriction \( \sqrt{N} \sum_{i=1}^{N} \omega_i = O(1) \), which contains \( \omega_i = 0_K \) as a special case.

We now illustrate how this restriction is needed when considering the OLS estimator of the second term of Eq. (1.6). Starting with a fixed \( N \), the OLS estimator of \( \sqrt{N}^{-1} \sum_{i=1}^{N} \Omega_i Q = \sqrt{N}^{-1} \sum_{i=1}^{N} E[\beta_i \epsilon_i'] Q \) is \( \sqrt{N}^{-1} \sum_{i=1}^{N} \hat{\Omega}_i \hat{Q} \) with \( \hat{Q} = \hat{\epsilon}_i \hat{Q} \). Since \( \hat{\epsilon}_i \) and \( \hat{Q} \) are orthogonal for any finite \( T \) and \( N \), the estimated term \( \sqrt{N}^{-1} \sum_{i=1}^{N} \hat{\Omega}_i \hat{Q} \) is a zero vector and \( \sqrt{N}^{-1} \sum_{i=1}^{N} \Omega_i Q \) cannot be identified. Next, when \( N \) diverges, even without post-multiplying by \( \hat{Q} \), it can be shown that \( N^{-1} \sum_{i=1}^{N} \hat{\Omega}_i \rightarrow_p 0_{K \times T} \), and once again \( \omega \) is \( \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \omega_i \) cannot be identified.

Therefore, under our fixed-\( T \) sampling scheme, the assumption \( \omega_i = \text{Cov}(\beta_i, \epsilon_i) = 0_K \) or, alternatively, the slightly more general assumption \( \sqrt{N}^{-1} \sum_{i=1}^{N} \omega_i = o(1) \) is needed for identification purposes.

\[\text{In particular, recalling that } M = I_T - D(D' D)^{-1} D' \text{ with } D = [I_T, \ F], \text{ under our assumptions } N^{-1} \sum_{i=1}^{N} \hat{\Omega}_i \rightarrow_p (\omega_i' 1_T + \sigma^2 P') M = 0_{K \times T} \text{ because } M \text{ is orthogonal to both } 1_T \text{ and } P.\]
1.E. Nonparametric Estimation of Risk Premia on Traded Factors

Under Eq. (1.48) and assuming that the factors are traded, it is well-known that the time-varying risk premia are given by

$$\gamma_{1,t} = E_t[f_{t+1}] - 1_{K} \gamma_{0,t}. \tag{1.6.E}$$

Moreover, when a risk-free asset with one-period return $r_{f,t}$ is available for investment and we assume that the zero-beta rate is equal to the risk-free rate, the latter expression simplifies to $\gamma_{1,t} = E_t[f_{t+1}] - 1_{K} r_{f,t}$, that is, the risk premia coincide with the conditional expected excess factor returns. This suggests that any estimator of the conditional mean can be used for risk premia estimation. A popular estimator of $E_t[f_{t+1}]$ is the sample mean of $T$ consecutive observations, that is,

$$\tilde{f}_t = \frac{1}{T} \sum_{h=-T/2}^{T/2-1} f_{t+h} \quad \text{or} \quad \tilde{f}_t = \frac{1}{T} \sum_{h=1}^{T} f_{t+h}. \tag{1.6.E}$$

Typically, the estimators in Eq. (1.6.E) are evaluated over consecutive rolling samples. In the absence of time variation, the risk premia are given by $\gamma_t = E_t[f_{t+1}] - 1_{K} r_{f}$, and the unconditional mean $E_t[f_{t+1}]$ is consistently estimated (as $T$ diverges) by the sample mean of $f_{t+1}$ over the full sample.

We now summarize the statistical properties of the risk premia estimators in Eq. (1.6.E) as $T$ diverges. It is convenient to simplify the exposition by setting $K = 1$ and assuming that the risk-free rate is constant over time. In addition, assume that the realized factor return, $f_{t+1}$, can be written as

$$f_{t+1} - r_f = \gamma_{1,t} + u_{t+1} \quad \text{for some i.i.d. error } u_{t+1} \sim (0, \sigma^2). \tag{1.6.E.9}$$

Finally, assume that the $T$ observations used to compute the estimators above represent a subset of a possibly larger number of observations, $T_0 \geq T$. Then, we consider two alternative sampling schemes. First, we evaluate the estimators’ behavior under the conventional scheme $T = T_0$, that is, using all the available data. Next, we consider a scheme where, even though $T$ diverges with $T_0$, $T/T_0 \to 0$. The latter is the sampling scheme adopted in nonparametric kernel estimation, and it implies that the estimators $\tilde{f}_t$ and $\tilde{f}$ are evaluated over a shrinking time interval (around period $t$) of relative length $T/T_0$ as $T_0$ diverges. Typically, samples of size $T$ are rolled over the entire length $T_0$.

Theorem 1.11 Under Eq. (1.6.E),

(i) When $T = T_0$, $\tilde{f}_t - r_f$ is an unbiased estimator of $\gamma_{1,t} = \gamma_{1,t} = T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} \gamma_{1,t+h}$ for any $T_0$. Moreover, when $T_0 \to \infty$,

$$T_0^{1/2} (\tilde{f}_t - r_f - \gamma_{1,t}) \rightarrow_d N(0, \sigma^2). \tag{1.6.E.10}$$

(ii) Finally, $\sigma^2$ can be consistently estimated by means of $T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} (f_{t+h} - \tilde{f}_{t+h})^2$ under the smoothness assumption $T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} (\gamma_{1,t+h} - \gamma_{1,t})^2 = o_p(1)$.

The same properties apply to $\tilde{f}_t$ with respect to $\gamma_{1,t} = T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} \gamma_{1,t+h}$.

Theorem 1.11. Under Eq. (1.6.E),

(i) When $T = T_0$, $\tilde{f}_t - r_f$ is an unbiased estimator of $\gamma_{1,t} = \gamma_{1,t} = T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} \gamma_{1,t+h}$ for any $T_0$. Moreover, when $T_0 \to \infty$,

$$T_0^{1/2} (\tilde{f}_t - r_f - \gamma_{1,t}) \rightarrow_d N(0, \sigma^2). \tag{1.6.E.10}$$

(ii) Finally, $\sigma^2$ can be consistently estimated by means of $T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} (f_{t+h} - \tilde{f}_{t+h})^2$ under the smoothness assumption $T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} (\gamma_{1,t+h} - \gamma_{1,t})^2 = o_p(1)$.

The same properties apply to $\tilde{f}_t$ with respect to $\gamma_{1,t} = T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} \gamma_{1,t+h}$.

Theorem 1.11. Under Eq. (1.6.E),

(i) When $T = T_0$, $\tilde{f}_t - r_f$ is an unbiased estimator of $\gamma_{1,t} = \gamma_{1,t} = T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} \gamma_{1,t+h}$ for any $T_0$. Moreover, when $T_0 \to \infty$,

$$T_0^{1/2} (\tilde{f}_t - r_f - \gamma_{1,t}) \rightarrow_d N(0, \sigma^2). \tag{1.6.E.10}$$

(ii) Finally, $\sigma^2$ can be consistently estimated by means of $T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} (f_{t+h} - \tilde{f}_{t+h})^2$ under the smoothness assumption $T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} (\gamma_{1,t+h} - \gamma_{1,t})^2 = o_p(1)$.

The same properties apply to $\tilde{f}_t$ with respect to $\gamma_{1,t} = T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} \gamma_{1,t+h}$.

Theorem 1.11. Under Eq. (1.6.E),

(i) When $T = T_0$, $\tilde{f}_t - r_f$ is an unbiased estimator of $\gamma_{1,t} = \gamma_{1,t} = T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} \gamma_{1,t+h}$ for any $T_0$. Moreover, when $T_0 \to \infty$,

$$T_0^{1/2} (\tilde{f}_t - r_f - \gamma_{1,t}) \rightarrow_d N(0, \sigma^2). \tag{1.6.E.10}$$

(ii) Finally, $\sigma^2$ can be consistently estimated by means of $T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} (f_{t+h} - \tilde{f}_{t+h})^2$ under the smoothness assumption $T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} (\gamma_{1,t+h} - \gamma_{1,t})^2 = o_p(1)$.

The same properties apply to $\tilde{f}_t$ with respect to $\gamma_{1,t} = T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} \gamma_{1,t+h}$.
When (i) \( T^{-1} + T_0^{1/2} (T/T_0)^{1.5} \to 0 \), (ii) the kernel \( \kappa(\cdot) \) integrates to unity and is differentiable, and both \( \kappa(\cdot) \) and \( \kappa'(\cdot) \) go to zero faster than \( O((1+u^2)^{-1}) \) for \( u \) large, and (iii) the true risk premium satisfies \( \gamma_{1,t} = \gamma_1(\frac{T}{T_0}) \) for a differentiable function \( \gamma_1(\cdot) \), then

\[
\hat{\gamma}_{1,t}^\kappa - \gamma_{1,t} = O_p(T^{-\frac{1}{2}}), \quad (1.13)
\]

\[
T^\frac{1}{2} (\hat{\gamma}_{1,t}^\kappa - \gamma_{1,t}) \to_d \mathcal{N}(0, \sigma^2 \int_{-\infty}^{\infty} \kappa(u)du), \quad (1.14)
\]

Finally, \( \sigma^2 \) can be consistently estimated by means of \( T_0^{-1} \sum_{h=1}^{T_0} (f_{h+1} - \hat{f}_t - \hat{\gamma}_{1,h}^\kappa)^2 \) under the assumptions above.

The proof is available upon request. Part (i) follows from noting that

\[
\hat{f}_t - r_f = \hat{\gamma}_{1,t} + T_0^{-1} \sum_{h=-T_0/2}^{T_0/2-1} u_{t+1+h}, \quad (1.15)
\]

and exploiting the properties of the sample mean of i.i.d. random variables. The proof of part (ii) follows from Robinson (1997), where the more general framework with non-i.i.d. innovations \( u_{t+1} \) is considered.

Part (i) of Theorem 1.11 indicates that the rolling estimators accurately estimate the average risk premia over a given time interval, but they will not converge to the true risk premia at some specific time \( t \). In particular, note that the rolling estimators converge to the \textit{integrated} risk premium \( \gamma_1 = \lim_{T_0 \to \infty} \frac{1}{T_0} \int_0^{T_0} \gamma_{1,s} ds \) (assuming that \( \gamma_1 \) is bounded). For inference, a smoothness condition that limits the degree of time variation in the true risk premia is required.

Part (ii) shows how the traditional rolling sample mean estimators in Eq. (1.1.8) can be obtained as special cases of the nonparametric Nadaraya-Watson estimator by suitably choosing the kernel function.\textsuperscript{55} Although inference can be conducted for the \( t \)-th risk premium, the rate of convergence is slower than the usual “square-root” speed. For example, condition \( T^{-1} + T_0^{1/2} (T/T_0)^{1.5} \to 0 \) is satisfied when \( T \) is not larger than \( T_0^{2/3} \). It follows that the rolling sample mean estimators will converge at rate \( O(T^{-1/3}) \). For instance, when \( T_0 = 100 \), only about \( T = 20 \) observations are available for inference on the time-varying risk premia. This small \( T \) implies that the confidence interval for the time-varying risk premium will be quite large. Therefore, not only \( T \) must diverge for the asymptotic theory to be valid, but \( T \) needs to be larger than what is required by the usual parametric rate.\textsuperscript{56} Notice how the results in parts (i) and (ii) require additional smoothness assumptions on the form and degree of time variation in the true risk premia. Using high-frequency data, Ang and Kristensen (2012) rely on similar nonparametric techniques to develop tests of conditional beta-pricing models.

\textsuperscript{55} For the estimator \( \hat{f}_t \) we provide the appropriate kernel function but a simpler, yet asymptotically equivalent, expression can be obtained by noting that \( \frac{1}{\sqrt{T}} \sum_{|u| \leq \frac{\sqrt{T}}{4}} \approx \frac{1}{2} 1_{|u| \leq 0.5} \) for \( T \) large.

\textsuperscript{56} As emphasized by Robinson (1997), the above results can be extended to non-i.i.d. errors by replacing \( \sigma^2 \) with \( \frac{2\pi f_0(0)}{2\pi f_0(0)} \), the spectral density of \( u_{t+1} \) at frequency zero (multiplied by \( 2\pi \)), and assuming boundedness of \( f_0(\cdot) \).
Table 1.11
Percentage difference between estimated risk premia

<table>
<thead>
<tr>
<th>Factor</th>
<th>$T = 36$</th>
<th>$T = 120$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: CAPM (with liquidity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mkt</td>
<td>64.3%</td>
<td>27.2%</td>
</tr>
<tr>
<td>liq</td>
<td>41.3%</td>
<td>54.2%</td>
</tr>
<tr>
<td>Panel B: FF3 (with liquidity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mkt</td>
<td>13.9%</td>
<td>7.3%</td>
</tr>
<tr>
<td>smb</td>
<td>14.7%</td>
<td>12.3%</td>
</tr>
<tr>
<td>hml</td>
<td>51.6%</td>
<td>31.2%</td>
</tr>
<tr>
<td>liq</td>
<td>22.9%</td>
<td>46.1%</td>
</tr>
<tr>
<td>Panel C: FF5 (with liquidity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mkt</td>
<td>15.3%</td>
<td>11.1%</td>
</tr>
<tr>
<td>smb</td>
<td>13.2%</td>
<td>9.7%</td>
</tr>
<tr>
<td>hml</td>
<td>14.1%</td>
<td>15.2%</td>
</tr>
<tr>
<td>rmw</td>
<td>13.3%</td>
<td>15.2%</td>
</tr>
<tr>
<td>cma</td>
<td>43.3%</td>
<td>33.0%</td>
</tr>
<tr>
<td>liq</td>
<td>13.9%</td>
<td>38.7%</td>
</tr>
</tbody>
</table>

The table reports the percentage difference between the Shanken (1992) estimator, $\hat{\gamma}_T$, and the OLS CSR estimator, $\hat{\gamma}_t$, averaged over rolling windows of size $T = 36$ and $T = 120$, respectively. The three panels refer to the CAPM, Fama and French (1993) three-factor model (FF3), and Fama and French (2015) five-factor model (FF5). Each of these models has been augmented with the non-traded liquidity factor of Pástor and Stambaugh (2003). We use monthly return data on individual stocks from CRSP and factor data from Kenneth French’s and Lubos Pástor’s websites from January 1966 to December 2013.
## Table 1.12
Betaz versus Characteristics

<table>
<thead>
<tr>
<th></th>
<th>CAPM</th>
<th>FF3</th>
<th>FF5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: F-tests and rejection frequencies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0 : \gamma^p_i = 0_K$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$-tests</td>
<td>14.54</td>
<td>17.33</td>
<td>21.14</td>
</tr>
<tr>
<td>Rejection frequencies</td>
<td>25.84%</td>
<td>28.72%</td>
<td>29.91%</td>
</tr>
<tr>
<td>$H_0 : \delta = 0_{K+}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$-tests</td>
<td>888.27</td>
<td>960.01</td>
<td>927.04</td>
</tr>
<tr>
<td>Rejection frequencies</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

|                |      |      |      |
| **Panel B: Variance ratios** |      |      |      |
| $100 \times \frac{s^2_{\alpha_i/C}}{s^2_{\alpha}}$ | 73.84% | 76.36% | 76.70% |
| $100 \times \frac{s^2_{\alpha,C}}{s^2_{\alpha}}$ | 2.21% | 3.11% | 3.19% |

The top panel of the table reports the $F$-tests (average over rolling windows of size $T = 36$) for the null hypotheses $H_0 : \gamma^p_i = 0_K$ and $H_0 : \delta = 0_{K+}$, respectively, and the rejection frequencies at the 95% confidence level (average over rolling windows of size $T = 36$). Each column refers to a different beta-pricing model, that is, the CAPM (first column), the Fama and French (1993) three-factor model (FF3, second column), and the Fama and French (2015) five-factor model (FF5, third column). The bottom panel reports the variance ratios $100 \times \frac{s^2_{\alpha_i/C}}{s^2_{\alpha}}$ and $100 \times \frac{s^2_{\alpha,C}}{s^2_{\alpha}}$ defined in Section 1.7.4 (average over rolling windows of size $T = 36$). The data is from DeMiguel et al. (2018) and Kenneth French’s website (from January 1980 to December 2015).
Chapter 2

Portfolio Choice with Model Misspecification

Valentina Raponi, Imperial College London
Raman Uppal, EDHEC Business School
Paolo Zaffaroni, Imperial College London
Abstract

We investigate the effect of model misspecification on mean-variance portfolios and show how asset-pricing theory and asymptotic analysis (for a large number of assets) can mitigate misspecification. Our analysis is founded on the Arbitrage Pricing Theory (APT), because it allows for pricing errors. We extend the APT to show it can capture not just small pricing errors unrelated to factors but also large pricing errors from mismeasured and missing factors. Our key insight is that, instead of treating misspecification directly in the mean-variance portfolio, it is better to first decompose the portfolio into a “beta” portfolio that depends only on factor risk premia and an “alpha” portfolio that depends only on pricing errors. Then, we show that as the number of assets increases, the weights of the alpha portfolio dominate those of the beta portfolio, leading to mean-variance portfolio weights that are immune to beta misspecification. For the alpha portfolio, we treat misspecification by imposing the APT restriction, which serves as an identification condition and a shrinkage constraint. Using simulations, we illustrate how our theoretical insights lead to a significant improvement in the out-of-sample performance of mean-variance portfolios.

Keywords: Active and passive portfolios, pricing errors, factor models, factor investing, mean-variance portfolio, estimation error, robust estimation, ambiguity.

JEL classification: G11, G12, C58, C53.
2.1. Introduction

In this paper, our objective is to study the effect of model misspecification on mean-variance portfolios and to show how asset-pricing theory and asymptotic analysis (as the number of assets increases) can be used to provide powerful solutions to mitigate it. In particular, we show how to improve the performance of mean-variance portfolios out of sample in the presence of model misspecification. In our context, one form of model misspecification is represented by the pricing error, often referred to as alpha.\(^1\) Another form of misspecification is associated with the beta component of returns; this arises even when all relevant factors are observable but one specifies the incorrect risk premia or covariances for the factors. Our key insight is that, instead of treating misspecification in the mean-variance portfolio, it is better to first decompose the mean-variance portfolio into two parts that correspond to the alpha and beta components of returns, and then to treat misspecification in the two parts using different methods.

We study portfolio choice assuming that asset returns satisfy the Arbitrage Pricing Theory (APT) of Ross (1976). The APT is particularly well-suited for our purpose because it allows for the possibility of model misspecification, and hence, mispricing (alpha), while still imposing no arbitrage. Moreover, the APT is a very general asset-pricing model that can accommodate a variety of observed factors. The factors could be statistical, for example, based on a principal-component decomposition of returns; macroeconomic, for example, shocks to inflation, interest rates, and exchange rates; or, characteristic-based, for instance, industry, country, size, value, return momentum, and liquidity.\(^2\)

Our first contribution is to extend the notion of pricing errors in the APT to show that it applies much more broadly than typically assumed. Traditionally, the APT is interpreted as applying to pricing errors that are small and unrelated to factors (see, for example, Cochrane (2005, Ch. 9.4)), implying that the covariance matrix of residuals has bounded eigenvalues. We show that the APT can capture not just small but also large pricing errors, such as those arising from latent pervasive factors or mismeasured. This implies that not all eigenvalues of the residual-covariance matrix are bounded, and hence, even well-spread portfolios may not be well diversified. In fact, our work shows that the APT is much more than just a statistical model of returns, and our results allow us to illustrate the deep economic content of the APT.

Our second contribution is to demonstrate how this extended interpretation of the APT can be used to capture, and then mitigate, model misspecification in the class of mean-variance portfolios. We treat model misspecification in three steps. In the first step, we show that under the APT, the optimal mean-variance portfolio can be decomposed into an "alpha" portfolio, which depends only on pricing errors with zero exposure to common risk, and a "beta" portfolio, which depends on factor risk premia and their loadings.\(^3\) We also show that the alpha and beta portfolios, despite being inefficient, span the entire mean-variance efficient frontier, thus leading to two-fund separation.\(^4\)

In the second step, we treat misspecification in the beta portfolio. In the environment with an asymptotically large number of assets, we show that the weights of the alpha portfolio typically dominate the weights of the beta portfolio.\(^5\) Given the secondary role played by the beta portfolio, we show that, under a set of

\(^1\)As in Hansen and Jagannathan (1997), this alpha could arise if the asset-pricing model is correct (that is, we have the correct set of factors) but the factors are measured with error (Roll, 1977; Green, 1986a) or if the factors are measured without error but the model is incorrect in the sense that some factors are missing (MacKinlay and Pastor, 2000) or are latent (Connor and Korajczyk, 1986; Lehmann and Modest, 1988). These alphas could also arise if the views of the investor disagree with the predictions of the asset-pricing model (Black and Litterman, 1990, 1992).

\(^2\)For further details of the variety of applications of the APT, see Connor et al. (2010, Ch. 4–6). For the importance of factor investing, see the excellent discussion in Ang (2014).

\(^3\)This decomposition is important because the world's largest hedge funds, such as Bridgewater Associates, offer alpha and beta portfolios. Similarly, sovereign-wealth funds, such as Norges Bank, separate the management of their alpha and beta funds. In fact, today most asset managers offer "portable alpha" products, and a large proportion of institutional investors have invested in these products. The returns on these alpha portfolios are often referred to as "absolute returns" because they are supposed to remain positive under all market conditions.

\(^4\)Even though the alpha and beta portfolios themselves are not on the Markowitz efficient frontier, they satisfy an optimality condition: each is the minimum-variance portfolio that is orthogonal to the other, extending the result in Roll (1980) about orthogonal portfolios to the case in which a risk-free asset is available.

\(^5\)The asymptotic analysis for a large number of assets is not just an abstract mathematical device but also corresponds to practice: hedge funds and sovereign-wealth funds hold a large number of assets in their portfolios; for instance, the portfolio of
mild conditions, it can be replaced, without any loss of efficiency, by a class of benchmark portfolios that by construction are independent of the mean vector and covariance matrix of the observed factors, and hence, immune to beta misspecification. The dominant role played by the pricing errors (alphas) asymptotically also implies that it is critical to estimate the pricing errors precisely.

In the third step, we treat misspecification in the alpha portfolio. We do this by imposing the APT restriction on the pricing errors when estimating the model, which reduces the error in the estimated parameters of the factor model generating returns. In our analysis, we consider both the case of pricing errors that are unrelated to factors and pricing errors that arise from latent factors. In the case in which we have both errors that are unrelated and those that are related to factors, the APT no-arbitrage restriction plays a second, even more fundamental, role: in the absence of this restriction, the model is not (econometrically) identified, and hence, cannot be estimated. This part of our work extends the rich insights of MacKinlay and Pastor (2000), who study estimation of models with missing factors.

Finally, we illustrate how these results regarding the decomposition of portfolio weights, the asymptotic analysis of these weights, together with the restriction arising from the extended APT, can and should be used to improve the estimation of the return-generating model and the portfolio weights in the presence of model misspecification. Using simulations, we show that it is possible to take advantage of our theoretical insights to achieve economically and statistically significant improvement in the out-of-sample performance of mean-variance portfolios.

In summary, traditional models of portfolio choice either have no alpha or do not distinguish between the alpha and beta components of returns, and therefore, when correcting for misspecification, they shrink only the beta component of returns. In contrast, our model, founded on the APT, distinguishes between the alpha and beta components of returns. Our portfolio strategy exploits alpha to improve portfolio performance, while it completely shrinks the beta components of returns, so that the beta portfolio does not depend at all on the distribution of the factor returns.

The rest of the paper is organized as follows. In Section 2.2, we discuss the literature related to our work. In Section 2.3, we specify the linear factor model for asset returns, summarize the results in the existing literature for the APT, and extend the APT to the case of unbounded eigenvalues for the residual covariance matrix. In Section 2.4, we describe the three steps that allow us to mitigate model misspecification in mean-variance portfolios. In Section 2.5, we demonstrate how these results can be applied to improve the estimation of portfolio weights that achieve superior out-of-sample performance. We conclude in Section 2.6. Proofs and technical details for all our results are collected in Appendix 2.A. Additional results, including the mitigation of model misspecification for the global-minimum-variance portfolio and the Markowitz efficient-frontier portfolios, are available in an online appendix (which is appended to this manuscript).

2.2. Related Literature

Each of our contributions is related to a distinct stream of the literature, which we discuss below.

Ross (1976, 1977) develops the APT by showing that if asset returns have a strict factor structure, then mean returns are approximately linear functions of factor loadings. Huberman (1982) formalizes the argument proposed for linear factor pricing in Ross. He also shows that the arbitrage portfolios used to prove the APT need not be well diversified (that is, these portfolios could have idiosyncratic risk); instead, they need to satisfy only two conditions: they are zero-cost portfolios and their weights are orthogonal to factor loadings (which implies that they have zero factor risk). Chamberlain (1983b), Chamberlain and Rothschild (1983), and Ingersoll (1984) extend the results of Ross (1976, 1977) and Huberman (1982) to a setting where returns need to satisfy only an approximate factor structure; that is, the idiosyncratic components of returns are allowed to be mildly correlated across assets. Chamberlain and Rothschild (1983) also show that if the covariance matrix of the asset returns has only K unbounded eigenvalues, then there is an approximate factor structure, and it is unique.6 Just as in Chamberlain (1983b), Chamberlain and Rothschild (1983), and Ingersoll (1984), we do not restrict the covariance matrix of the residuals to be diagonal; that is, we allow

---

Norges Bank has over 9,000 assets. Our asymptotic results hold even when the number of assets is as small as 100.

6See Reissman (1988, 1992a,b) for further extensions of the APT.
for correlated error terms. However, in contrast to Chamberlain (1983b) and Chamberlain and Rothschild (1983), we study also the case in which the maximum eigenvalue of the residual covariance matrix is not restricted to be bounded as the number of assets increases, which is the case when the pricing errors are large, for instance, when they are related to some latent pervasive factors.

The reason that Chamberlain (1983b) and Chamberlain and Rothschild (1983) do not consider the case in which the maximum eigenvalue of the covariance matrix is unbounded when the number of assets is large is that they view all factors as latent (in fact, they advocate the use of principal components to estimate the overall factor structure); consequently, the covariance matrix for the residuals left over after extracting the principal components has necessarily bounded maximum eigenvalue. However, in contrast to Chamberlain (1983b) and Chamberlain and Rothschild (1983), the usual practice is to consider that returns are driven by a set of observed factors, which are assumed to be the only source of commonality. In this case, if there are missing factors, then the mean of the missing factors will show up in the pricing errors and the covariance of the missing factors will contribute to the residual covariance matrix. Therefore, the maximum eigenvalue of the residual covariance matrix could be unbounded as the number of assets increases.

The second part of our analysis studies the implications for portfolio selection of mispricing, as modeled by the APT.\footnote{For work showing how beliefs about an asset-pricing model can influence portfolio choice, see Pástor (2000) and Pástor and Stambaugh (2000).} There is a large body of literature that studies how one should form portfolios in the presence of mispricing, which could arise because the index portfolio is inefficient (Dybvig and Ross (1985a), Green (1986a)) or because of superior information, analyst recommendations, or managerial skill (see, for example, Dybvig and Ross (1985b), Grinblatt and Titman (1989), and Kosowski et al. (2006)). The seminal paper in this body of literature is Treynor and Black (1973), that takes as its starting point the single-factor model with a diagonal residual covariance matrix as in Sharpe (1963), and asks whether it is desirable to form an “active” portfolio by going long underpriced assets and shorting overpriced assets so that market risk is fully hedged, or, should one invest in only a diversified “passive” portfolio so that it is exposed only to market risk. Most importantly, they highlight the vital role in active asset management of pricing errors, which they call “independent returns.”

Treynor and Black (1973), and the papers that build on their work,\footnote{See, for example, Kane et al. (2003), who use shrinkage estimators to improve the econometric methods used in constructing portfolios, He (2007), who extends the Treynor and Black model to a Bayesian setting, and Brown and Tu (2010), who use the Treynor and Black model to assess ex post whether the proper active-passive allocation strategy was adopted by the portfolio manager. For surveys of this literature, see Aragon and Penson (2006) and Ferson (2014).} study the case of mean-variance portfolios with a target mean for the case of a finite number of assets that may be overpriced or underpriced, with these pricing errors uncorrelated to the single factor. However, these papers do not restrict the pricing errors in any way, so it is not clear how exactly arbitrage opportunities are ruled out. Moreover, the pricing errors they consider are unrelated to latent factors. Our work extends their analysis along these dimensions. We provide a firm theoretical foundation for the important issue of active versus passive portfolio management highlighted in Treynor and Black (1973) by placing this analysis in the context of a no-arbitrage asset pricing model such as the APT. We allow for multiple factors, do not restrict the residual covariance matrix to be diagonal, and study the case of both pricing errors that are independent of factors and those that arise from latent factors. Finally, we show that the no-arbitrage restriction imposed by the APT plays a central role in the estimation of optimal active and passive portfolios.

There is also a body of literature that considers whether weights in mean-variance portfolios can be extreme. In a setting without pricing errors, Green and Hollifield (1992) show that the presence of a single pervasive factor in asset returns would lead to extreme positive and negative weights in mean-variance efficient portfolios.\footnote{Green (1986b) provides necessary and sufficient conditions for portfolios on the minimum-variance frontier to have positive weights in all assets.} We complement this result by showing that, in the presence of pricing errors, the long-short alpha portfolio weights dominate the portfolio weights as the number of assets increases. Pesaran and Zaffaroni (2009) show that under some technical conditions, the limiting properties of the mean-variance portfolio and a factor-neutral portfolio can be similar.

The last part of our analysis examines the implications of the no-arbitrage constraint imposed by the
APT and the insights from the asymptotic analysis of portfolio weights for the empirical estimation of mean-variance portfolio weights in order to improve their out-of-sample performance. It is well known that mean-variance efficient portfolios that are based on sample estimates of first and second moments perform poorly out of sample; see, for example, Jobson and Korkie (1980), Frost and Savarino (1986), Michaud (1989), Black and Litterman (1990), and DeMiguel et al. (2009b).

Of the many approaches considered to improve the out-of-sample performance of mean-variance portfolios, one is to impose portfolio constraints. For example, Frost and Savarino (1988) find that imposing shortsale constraints can lead to significant improvement in performance. In an insightful paper, Jagannathan and Ma (2003) explain, in the context of the global-minimum-variance portfolio, that the reason for the improved performance is that imposing shortsale constraints is equivalent to shrinking the covariance matrix, and that this constraint can help even when returns are driven by a dominant factor. DeMiguel et al. (2009a) show that further gains are possible by imposing a more general form of the shortsale constraint, a norm constraint on the portfolio weights; they also show that these constraints can be interpreted as leading to portfolio weights with Bayesian shrinkage, just as Tibshirani (1996) does for the lasso and ridge-regression techniques. Olivares-Nadal and DeMiguel (2015) show that the portfolio-optimization problem with a constraint that is motivated by transaction costs can be interpreted in three ways: as a robust optimization problem, a robust regression problem, and a Bayesian problem.

MacKinlay and Pástor (2000) recognize that a missing factor implies the presence of the pricing error in the residual covariance matrix and that taking this into account in the estimation improves portfolio selection. They also find that, even if the true covariance matrix of returns is not the identity matrix, using the identity matrix as the covariance matrix, which is implicitly another kind of restriction, improves portfolio performance. Pettenuzzo et al. (2014) demonstrate that economic constraints, such as restricting the equity risk premium to be positive and bounding the Sharpe ratio, improve the estimation of time-series forecasts of the equity risk premium. In our work, the constraint to be imposed when estimating asset returns follows naturally from the APT. Our novel approach of imposing the no-arbitrage constraint for estimation of the equity return-generating model is similar in spirit to the usual practice of imposing a no-arbitrage constraint when estimating models of fixed-income returns; for further analogies between returns on equities and returns on fixed-income instruments, see Binsbergen and Koijen (2017).

As highlighted by the literature described above, an investor choosing optimal portfolio weights faces both model and parameter uncertainty. Two features of the APT allow us to deal with both sources of uncertainty. The first feature is allowing for the presence of the pricing error, denoted by the vector \( \alpha \). The pricing error in the model tells us how the incorrectly specified mean and variance of returns are to be adjusted in the presence of misspecification.

The second feature of the APT is the restriction on the magnitude of the pricing error: \( \alpha' \Sigma^{-1} \alpha \leq \delta < \infty \), where \( \Sigma^{-1} \) is the inverse of the covariance matrix of residuals and \( \delta \) is an arbitrary positive constant. By imposing the APT restriction when estimating the parameters, one ensures that the "approximating model" lies within the set of no-arbitrage models. This restriction limits the magnitude of estimates for \( \alpha \), and hence, also allows one to deal with parameter uncertainty associated with \( \alpha \). Interestingly, imposing the APT restriction is analogous to the approach adopted in Garlappi et al. (2007), in which investors are assumed to be averse to their uncertainty (ambiguity) about the model, and one accounts for parameter uncertainty using the minimax approach originally proposed in Gilboa and Schmeidler (1989). In the

---

10 Model uncertainty arises when the investor is not confident about the true data-generating process. Parameter uncertainty, on the other hand, refers to not knowing the true parameter values of the data-generating process. Thus, parameter uncertainty arises even in the absence of model uncertainty; that is, even if one knows the true data-generating process, one may not know the parameter values for this process. Conditional on knowing the true model, one can always resolve parameter uncertainty with an infinite number of observations.

11 Adopting the terminology of Hansen and Sargent (2008), this means that any "approximating model" that is estimated with the data must lie within the set bounded by the APT restriction. Hansen and Sargent (2008) also provide a comprehensive discussion of how one can analyze decision makers who regard their model as an approximation and who desire decision rules that work over a set of models in the neighborhood of the approximating model.

12 The uncertainty arising from having to estimate the parameters associated with the beta component of returns, referred to as beta misspecification, is dealt with using asymptotic analysis as described later in the text.

13 There is extensive literature modeling decision making in the presence of ambiguity: see, for example, Chen and Epstein.
minmax approach, portfolio weights are obtained by first minimizing the mean-variance objective function, equivalent to minimizing the Sharpe ratio, over the set of expected returns subject to the constraint that these expected returns are not too distant from the estimated mean returns, and then maximizing over the portfolio weights; see, for example, Garlappi et al. (2007, equation (14)). Analogously, our approach can be interpreted as first estimating the parameters subject to the constraint that the maximum-likelihood (ML) estimates satisfy the APT restriction, which is equivalent to restricting the Sharpe ratio of the alpha portfolio. We then plug in the estimated parameters into the mean-variance objective function and choose portfolio weights to maximize its Sharpe ratio. The key difference between our approach and that of the minmax approach is that the minmax approach constrains total expected returns, while we constrain only the alpha component of expected returns and account for misspecification in the beta component of returns using asymptotic analysis.\footnote{Garlappi et al. (2007) and DeMiguel et al. (2009a) also provide a Bayesian interpretation of the portfolio weights that account for parameter uncertainty. Specifically, all these approaches—Bayesian, minmax, and imposing the APT restriction—that address the problem of parameter uncertainty lead to shrinkage-type estimators; see, for example, Bawa et al. (1979), Jorion (1986), Pástor (2000), Pástor and Stambaugh (2000), and Garlappi et al. (2007).}

2.3. Generalizing the APT

In this section, we start by describing our notation and assumptions. Our analysis is founded on precisely the same assumptions as the ones underlying the APT. After stating the APT result in the existing literature, we show how it can be extended to incorporate large pricing errors, that is, where the covariance matrix of returns has unbounded eigenvalues implying that even a well-spread portfolio will not be diversified. We conclude this section by describing how the APT model can capture misspecification arising from the beta and alpha components of returns.

2.3.1 Notation

The number of risky assets is denoted by $N$. Just like in Chamberlain and Rothschild (1983) and Ingersoll (1984), we study a market with an infinite number of assets. To make clear the dependence on the number of assets, we index quantities that are $N$-dimensional by the subscript $N$, except for random variables, such as the returns on risky assets, which have the subscript $t$. Instead of considering a sequence of distinct economies, we consider a fixed infinite economy in which we study a sequence of nested subsets of assets. Therefore, in the $N$th step, as a new asset is added to the first $N-1$ assets, the parameters of the first $N-1$ stay unchanged. These unchanging parameters can be interpreted as the parameters one would get in the limit as the number of assets becomes asymptotically large.

Let $\tau_t$ denote the return on the risk-free asset and let the $N$-dimensional vector $r_t = (r_{1t}, r_{2t}, \ldots, r_{Nt})'$ denote the vector of returns on risky assets.

Given an arbitrary portfolio strategy $a$ with weights $w_N^a = (w_1^a, w_2^a, \ldots, w_N^a)'$ of $N$ risky assets, and using $1_N$ to denote an $N$-dimensional vector of ones, we define the associated portfolio return as

$$r_t^a = r_t'w_N^a + \tau_t(1 - 1_N'w_N^a), \tag{2.1}$$

with finite conditional mean, standard deviation, and Sharpe ratio defined as

$$\mu^a = E(r_t^a) = E(r_t)'w_N^a + \tau_t(1 - 1_N'w_N^a), \tag{2.2}$$

$$\sigma^a = \sqrt{\text{Var}(r_t')} \quad \text{and} \quad \tag{2.3}$$

$$\text{SR}^a = \frac{\mu^a - \tau_t}{\sigma^a}. \tag{2.4}$$

\footnote{2002}, Ghirardato et al. (2004), Kilianoff et al. (2005), Ju and Miao (2012), and Maccheroni et al. (2013). For surveys of this literature, see Hansen and Sargent (2008) and Epstein and Schneider (2010).
2.3.2 Linear factor model for asset returns with misspecification

We start our analysis with the assumption of a linear latent-factor structure for returns. Just as in Chamberlain (1983b), Chamberlain and Rothschild (1983), and Ingersoll (1984), we do not restrict the covariance matrix of the residuals, $\Sigma_N$, to be diagonal; that is, we allow for correlated error terms. Furthermore, in contrast to these papers, we also study in Section 2.3.4 the case in which not all eigenvalues of $\Sigma_N$ are restricted to be bounded when $N$ is large.

**Assumption 2.1 (Linear factor model)** We assume the $N$-dimensional vector $r_t$ of asset returns can be characterized by

$$ r_t = \mu_N + B_N z_t + \epsilon_t, $$

where $z_t = (z_{1t}, z_{2t}, ..., z_{Kt})'$ is the $K \times 1$ vector of common unobserved factors, $B_N = (\beta_1, \beta_2, ..., \beta_N)'$ is an $N \times K$ full-rank matrix of factor loadings with $i$th row $\beta_i'$, $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{Nt})'$ is an $N \times 1$ vector of residuals, and the $N \times 1$ vector $\mu_N$ represents the mean of the vector of returns, $r_t$. At any time $t$, $z_t$ is distributed with zero mean and $K \times K$ covariance matrix $\Omega$, and $\epsilon_t$ is distributed with zero mean and the $N \times N$ covariance matrix $\Sigma_N$, with $\Omega$ and $\Sigma_N$ being positive definite. Moreover, $\epsilon_t$ and $z_t$ are uncorrelated; that is, $E(\epsilon_t z_t') = 0$.

It is important to note that the APT is a model of the random component of returns, $r_t - \mu_N = B_N z_t + \epsilon_t$, and is silent about expected returns. Black (1995, p. 168) recognizes this and states that the “Arbitrage Pricing Theory (APT) is a model of returns. It says that the number of independent factors influencing return is limited, but it is silent on the pricing of these factors, so it is silent on expected return.” For instance, Assumption 2.1 implies that the conditional variance-covariance matrix for asset returns is

$$ E[(r_t - \mu_N)(r_t - \mu_N)'] = V_N = B_N \Omega B_N' + \Sigma_N, $$

regardless of the vector of expected returns, $\mu_N$.

2.3.3 Arbitrage Pricing Theory (APT)

We now describe the APT result in the literature. In the definition below, as well as throughout the paper, we use $\delta$ to denote an arbitrary positive scalar.

**Definition 2.12 (Asymptotic arbitrage)** A sequence of portfolios is said to generate an asymptotic arbitrage opportunity if along some subsequence $N'$:

$$ \text{Var}(r_t' w_N') \to 0 \quad \text{as} \quad N' \to \infty \quad \text{and} \quad (\mu_{N'} - \gamma_1 1_N)' w_N \geq \delta > 0 \quad \text{for all} \quad N'. $$

**Assumption 2.2 (No asymptotic arbitrage)** There are no asymptotic arbitrage opportunities.

We now state the APT result. This result is derived in Huberman (1982) and Ingersoll (1984), so we do not include the proof.

**Theorem 2.1 (The APT restriction (Huberman, 1982; Ingersoll, 1984))** Assumptions 2.1 and 2.2 imply that for all $N$ there exists some positive number $\delta$ such that the weighted sum of the squared pricing errors is uniformly bounded:

$$ \alpha_N' \Sigma_N^{-1} \alpha_N \leq \delta < \infty, $$

where the vector of pricing errors is

$$ \alpha_N = (\mu_N - \gamma_1 1_N) - B_N \lambda. $$

---

15 Throughout the paper, it will be assumed that $K < N$. 
and the vector of risk premia is

$$\hat{\lambda} = (B'_N\Sigma^{-1}_N B_N)^{-1} B'_N \Sigma^{-1}_N (\mu_N - r_f 1_N),$$

(2.10)

with the $'$ symbol used to denote that the quantity is obtained from a projection, and hence, every component of it will change with $N$.\footnote{Note that $\hat{\alpha}_N$ has the subscript $N$ because it is $N$ dimensional, while $\hat{\lambda}$ is not subscripted by $N$ because it is not $N$ dimensional.}

Remark 2.1 Observe that $\hat{\lambda}$ is the projection coefficient when projecting $(\mu_N - r_f 1_N)$ on $B_N$, and $\hat{\alpha}_N$ is the projection residual (pricing error) that satisfies

$$B'_N \Sigma^{-1}_N \hat{\alpha}_N = 0.$$

(2.11)

2.3.4 Extending the APT

Above, we described the standard APT. In this section, we explain our first contribution, which is to extend the standard APT to the case in which the residual covariance matrix does not necessarily have bounded eigenvalues; that is, even a well-spread portfolio may not be well diversified. This occurs precisely because of model misspecification.

We adopt the following notation. Consider a symmetric $M \times M$ matrix $A$. Let $g_{M}(A)$ denote the $i$th eigenvalue of $A$ in decreasing order for $1 \leq i \leq M$. Thus, the maximum eigenvalue is $g_{IM}(A)$ and the minimum eigenvalue is $g_{MM}(A)$.

We start with the definition of a regular factor economy.

Definition 2.2 (Regular factor economy (Ingersoll, 1984, p. 1028 and fn. 10)) A factor economy is regular if the maximum eigenvalue $g_{KK}(B'_N \Sigma^{-1}_N B_N)$ goes to zero as $N \to \infty$. If the limit is positive instead of zero, then the factor representation is irregular. Equivalently, in a regular economy, the minimum eigenvalue $g_{KK}(B'_N \Sigma^{-1}_N B_N)$ diverges as $N \to \infty$.

The above definition implies that in a regular economy the risk arising from factors cannot be diversified away; see Connor et al. (2010) for additional details.

Next, we introduce the limit of $\hat{\lambda}$ as $N \to \infty$, which we label $\lambda$, and the associated vector of pricing errors, $\alpha_N = (\mu_N - r_f 1_N - B_N \lambda)$.\footnote{Note that $\alpha_N$ is an $N$-dimensional vector that is a component of the infinite-dimensional vector $\alpha$; as $N$ increases, the number of elements in $\alpha_N$ increases, but the elements themselves do not change.} Under Assumptions 2.1 and 2.2 and the assumption of a regular factor economy, Ingersoll (1984, Theorem 3 and fn. 10) shows that $\lambda$ is unique and prices assets with bounded squared error:

$$\alpha'_N \Sigma^{-1}_N \alpha_N \leq \delta < \infty,$$

(2.12)

where $\delta$ represents some positive arbitrary scalar. Observe that the APT restriction in (2.8) is expressed in terms of $\hat{\alpha}_N$, while the restriction in (2.12) is expressed in terms of $\alpha_N$; in Lemma 2.4, we show the equivalence between these two conditions.

Observe that, as discussed in Ingersoll (1984, fn. 10), in (2.12) the weighting of the squared pricing error uses $\Sigma^{-1}_N$. The intuition for why the squared-error bound in (2.12) is weighted by (the inverse of) the matrix of residual variances, $\Sigma_N$, is that an asset's usefulness in arbitrage trades is limited by its residual variance. As Ingersoll (1984) explains: "Roughly speaking, the smaller an asset's residual variance, the more extreme its weighting can be in a portfolio, while maintaining diversification, and the more extreme its weighting, the bigger is its effect on the portfolio's expected return. Thus, an asset with a small residual variance can have a major impact on a diversified portfolio's expected return, and to prevent arbitrage its own expected return must be more nearly in line with the prediction of the APT. Assets with larger residual variances can only have small effects on a diversified portfolio's expected return, so their own expected returns need not be as closely in line with the prediction."
The restriction in (2.12), which is a consequence of asymptotic no arbitrage, links the pricing error $\alpha_N$ to the residual covariance matrix, $\Sigma_N$. There are two possible, complementary, cases for $\Sigma_N$ as $N \to \infty$; we examine both. For large $N$, in the first case all the eigenvalues of $\Sigma_N$ are bounded, and in the second case at least one of the eigenvalues is unbounded. The existing APT literature has focused on studying the case in which all the eigenvalues of $\Sigma_N$ are bounded, which—for completeness—is restated in the theorem below, with the proof for this provided in Huberman (1982) and Ingersoll (1984).

**Theorem 2.3 (Constraint imposed on $\alpha_N$ by no arbitrage for case with bounded eigenvalues; Huberman)**

If $\Sigma_N$ has bounded eigenvalues for large $N$, then the restriction in (2.12) requires the elements of the pricing-error vector $\alpha_N$ to become small for large $N$ in the following sense:

$$\alpha_N' \alpha_N \leq g_{1N}(\Sigma_N)(\alpha_N' \Sigma_N^{-1} \alpha_N) \leq \delta < \infty,$$

but without $\alpha_N$ being tied down to $\Sigma_N$.

**Remark 2.2** Notice that in the original formulation of the APT in (2.5), no factor is assumed to be observed; instead, the error term $r_t - \mu_N = B_N z_t + \epsilon_t$ has a latent factor structure through the common innovation $z_t$. In contrast, nowadays factor models are typically used assuming the existence of a given number of traded observed factors. This means that rather than the model in (2.5), one considers

$$r_t - \tau 1_N = \alpha_N + B_N(f_t - \tau 1_K) + \epsilon_t,$$

for the $K$-dimensional factors $f_t$, all of which are observed. Notice that (2.5) and (2.14) are equivalent if $f_t - \tau 1_K = \lambda + z_t$ and $\mu_N - \tau 1_N = \alpha_N + B_N \lambda$.

We now show that the APT model in the existing literature can be extended to the case where, as $N$ increases, some of the eigenvalues of $\Sigma_N$ are not bounded. This typically occurs when there are missing or mismeasured factors, as explained below in Section 2.3.5.

**Theorem 2.4 (Constraint imposed on $\alpha_N$ by no arbitrage for case with unbounded eigenvalues)**

Suppose that the vector of asset returns, $r_t$, satisfies Assumptions 2.1 and 2.2. Suppose that for some finite $1 \leq p < N$ the following three conditions hold: (i) $\sup_N g_{pN}(\Sigma_N) = \infty$; (ii) $\sup_N g_{p+1N}(\Sigma_N) \leq \delta < \infty$; and, (iii) $\inf_N g_{NN}(\Sigma_N) \geq \delta > 0$. Then, the APT restriction in (2.12) is satisfied by the pricing error $\alpha_N$, represented as

$$\alpha_N = A_N \lambda_{miss} + a_N,$$

where $A_N$ is an $N \times p$ matrix whose $j$th column equals $g_{jN}(\Sigma_N)v_{jN}(\Sigma_N)$, where $1 \leq j \leq p$, $v_{jN}(\Sigma_N)$ is the eigenvector of $\Sigma_N$ associated with the eigenvalue $g_{jN}(\Sigma_N)$, $\lambda_{miss}$ is some $p \times 1$ vector, and $a_N$ is some non-zero $N \times 1$ vector that satisfies

$$a_N' \Sigma_N^{-1} a_N \leq \delta < \infty.$$

**Remark 2.3** One can interpret the two components of $\alpha_N$ in (2.15) in a variety of ways. The first term, $A_N \lambda_{miss}$, could be associated with $p$ latent or missing pervasive factors, where $A_N$ are the factor loadings and $\lambda_{miss}$ are the risk premia for the missing factors.\textsuperscript{18} The second term, $a_N$, is the idiosyncratic part of the pricing error $\alpha_N$; for instance, $a_N$ could be interpreted as representing managerial skills or views of analysts. Under Assumptions 2.1 and 2.2, the expected excess return can be written as

$$E(r_t - \tau 1_N) = \mu_N - \tau 1_N = \alpha_N + B_N \lambda = (A_N \lambda_{miss} + a_N) + B_N \lambda.$$

Observe that, compared to equation (2.14), equation (2.17) contains the decomposition of the extra term $\alpha_N = (A_N \lambda_{miss} + a_N)$, which represents the effect of misspecification of the traded observed factors.

\textsuperscript{18}Pervasiveness of the latent missing factors with loadings $A_N$ follows from the assumption that $g_{NN}$ for $1 \leq j \leq p$ are diverging for large $N$. 
Figure 2.1 Weighted sum of the squared pricing errors

In this figure, we plot the weighted sum of the squared pricing errors, $\alpha_N^T \Sigma^{-1}_N \alpha_N$, as the number of assets $N$ increases for three different cases. In the first case, which is the one studied in the existing literature, the elements of the pricing-error vector, $\alpha_N$, become small as $N$ increases and the eigenvalues of $\Sigma_N$ are bounded; in this case, the weighted sum of the squared pricing errors is bounded, as shown by the dotted (red) line. In the second case, we allow for large pricing errors that are related to factors. In this case, some of the eigenvalues of $\Sigma_N$ are unbounded. As the dashed (blue) line shows, even in this case the weighted sum of the squared pricing errors is bounded, demonstrating that the pricing errors can be large without violating the APT restriction. The third case, illustrated by the dotted-dashed (green) line, shows that if the pricing errors were large but the eigenvalues of $\Sigma_N$ were incorrectly specified to be bounded, only then would the APT restriction be violated.

Remark 2.4 More importantly, Theorem 2.4 shows that the common perception that the pricing error $\alpha_N$ needs to be small in the APT is not accurate. In particular, if the maximum eigenvalue of the residual covariance matrix is not bounded, then the pricing errors can also be large without violating the APT restriction given in (2.12); this is illustrated in Figure 2.1. What Theorem 2.4 states is that if the maximum eigenvalue of $\Sigma_N$ is asymptotically unbounded, then the contribution of the pricing error to the portfolio return could also be large, but for this to satisfy the no-arbitrage condition, any portfolio earning this high return would not be well diversified and would be bearing idiosyncratic risk. To see this, observe that by Chamberlain and Rothschild (1983, Theorem 4), under the assumptions made for deriving Theorem 2.4, the covariance matrix of residuals, $\Sigma_N$, has the following approximate p-factor structure:

$$\Sigma_N = A_N A_N^T + C_N,$$

(2.18)

where $C_N$ is a $N \times N$ positive semi-definite matrix with bounded eigenvalues.\(^{19}\) Therefore, the residual

\(^{19}\)Under the assumptions of Chamberlain and Rothschild (1983, Theorem 4), $C_N$ will be non-singular for a sufficiently large $N$, which implies that $\Sigma_N$ is also non-singular for a sufficiently large $N$. 
variance of the return on any portfolio weights \( w_N \) is given by
\[
w_N^\prime \Sigma_N w_N = w_N^\prime A_N A_N^\prime w_N + w_N^\prime C_N w_N.
\] (2.19)

Whereas the second term, \( w_N^\prime C_N w_N \), goes to zero for well-spread portfolios (that is, for portfolios with \( w_N^\prime w_N \to 0 \)), there is no guarantee that the same occurs for the first term, \( w_N^\prime A_N A_N^\prime w_N \).

Ingersoll (1984, p. 1026) defines the setting of Theorem 2.3 as one with bounded residual variation; we label the setting of Theorem 2.4 as one with unbounded residual variation.

### 2.3.5 Different forms of model misspecification

In this section, we study the different forms of model misspecification of the true model. Throughout this section, we assume that the true model is given by (2.20), where the subscript “0” indicates the true value of a parameter (and we do not use the subscript “N” for the true value of parameters in order to limit the number of subscripts):
\[
r_{t} - r_f 1_{N} = a_0 + B_0 (f_{0,t} - r_f 1_{K_0}) + \varepsilon_t,
\] (2.20)

for \( K_0 \) factors, of which \( 0 \leq K \leq K_0 \) are observed. In the above equation, \( f_{0,t} \) denotes all the common factors affecting returns, of which \( f_t \) are the factors that are observed, \( B_0 \) are the true loadings on these factors, and \( a_0 \) are the true firm-specific components of expected returns that satisfy the APT restriction. The true mean and covariance matrix of \( f_{0,t} \) are denoted by \( \lambda_0 \) and \( \Omega_0 \), respectively. We describe four forms of model misspecification that are captured by the framework we have described in the paper.\(^{20}\)

The first is related to the “beta” component of returns. The second, third, and fourth are related to the “alpha” component of returns, arising from the presence of a pricing error unrelated to factors (that is, \( a_0 \neq 0 \)), missing factors (that is, for \( K < K_0 \)), and mismeasured factors. For expositional ease, we will assume that for all cases other than Case 2 below, \( a_0 = 0 \).

**Case 1: Incorrect means or covariances for the factors**

We start by considering the situation in which all \( K_0 \) factors are observed, implying that \( K = K_0 \) and that there is no error in measuring these factors because our objective in this case is to study misspecification in the “beta” component of returns; that is, misspecification in the assumed distribution of the factors, namely \( \lambda \neq \lambda_0 \) or \( \Omega \neq \Omega_0 \) because investors have incorrect views.\(^{21}\) This implies that the portfolio weights will also be incorrectly specified, even in population, despite the fact that the model in (2.20) can be correctly estimated using OLS to obtain consistent estimates of \( B_0 \) and the residual covariance matrix, \( \Sigma_0 \).

**Case 2: Pure pricing errors (unrelated to factors)**

Next, consider the case in which all factors are observed without error but expected returns also depend on non-factor-related characteristics, given by \( a_0 \neq 0 \) in (2.20). Thus, in this case, misspecification will arise if the investor erroneously sets \( a_N = 0 \). In this setting, the pricing error \( a_0 \) does not influence the variance-covariance matrix of returns, in contrast to the next case.

\(^{20}\) Another form of misspecification is associated with the phenomenon of “spurious” factors; namely, when the chosen factors have a small covariance with the returns of the test assets. For estimation of risk premia in the presence of spurious factors, see Bryzgalova (2016b) and references therein. Although the presence of spurious factors affects identification and inference about risk premia (unless the factors are traded), Saissoy et al. (2016) show that regardless of the effect on risk premia, inference on the expected (excess) returns, and hence, portfolio construction, is not affected.

\(^{21}\) For example, one could imagine a world in which the true risk premia are conditionally time varying but the investor models them as constant through time.
Case 3: Missing factors

Suppose now that of the $K_0$ factors, only $K$ are observed and $p = K_0 - K > 0$ are missing, and suppose for simplicity that the observed and missing factors are uncorrelated. For simplicity, we also assume that $a_0 = 0$. Then, the model in (2.20) can be rewritten as

$$r_t - r_f 1_N = A_N (f_{\text{miss},t} - r_f 1_p) + B_N (f_t - r_f 1_K) + \varepsilon_t,$$

where $f_{\text{miss},t}$ are the missing factors,

$$f_{0,t} = \begin{pmatrix} f_{\text{miss},t} \\ f_t \end{pmatrix} \quad \text{and} \quad B_0 = \begin{pmatrix} A_N & B_N \end{pmatrix}.$$  

Rewriting $f_{\text{miss},t} - r_f 1_p = \lambda_{\text{miss}} + \mathbf{z}_{\text{miss},t}$, it follows that

$$r_t - r_f 1_N = A_N \lambda_{\text{miss}} + B_N (f_t - r_f 1_K) + (\varepsilon_t + A_N \mathbf{z}_{\text{miss},t}),$$

where $E[\mathbf{z}_{\text{miss},t}] = 0$ and $E[\mathbf{z}_{\text{miss},t} \mathbf{z}_{\text{miss},t}'] = I_p$ to achieve identification.\(^{22}\)

Case 4: Mismeasured factors

Finally, consider the case in which all $K_0$ factors are measured with error. In particular, the observed factors satisfy $f_t = f_{t,t} + \eta_t$, where the measurement error $\eta_t$ has mean $E(\eta_t) = \mu_\eta$ and covariance matrix $E((\eta_t - \mu_\eta)(\eta_t - \mu_\eta)') = \Sigma_\eta$.\(^{23}\) As before, for simplicity we also assume that $a_0 = 0$. Then, (2.20) can be rewritten as

$$r_t - r_f 1_N = -B_0 \mu_\eta + B_N (f_t - r_f 1_K) + (\varepsilon_t - B_N (\eta_t - \mu_\eta)).$$

The econometric problem in estimating (2.24) is akin to an errors-in-variables problem: the residual is correlated with the observed factors through the measurement error, $\eta_t$. For econometric identification one typically sets $\mu_\eta = 0$. Also observe that there is a link between $\alpha_N$ and $B_N$ (in fact, $B_N$ appears in three terms of the model in (2.24)); this link can be used to test for the presence of mismeasurement and can also be exploited in the estimation.

2.4. Mitigating Model Misspecification

We study the weights and returns for the family of mean-variance portfolios. In particular, we study: (1) the mean-variance efficient portfolio when a risk-free asset is available, $w_M^\text{mve}$; (2) the global-minimum-variance portfolio when a risk-free asset is not available, $w_M^\text{gmv}$; and (3) the mean-variance efficient portfolios in the absence of a risk-free asset, $w_N^\text{mve}$, which are the Markowitz frontier portfolios that have the smallest variance for a given target mean. These three portfolio are displayed in Figure 2.2.

For each of these three portfolios, we treat model misspecification in three steps. First, we show how the weights and returns of these portfolios can be decomposed into two components: a component that depends on the risk premia (beta) and a component that depends on the pricing error (alpha). Second, we show how misspecification in the beta component of returns can be mitigated. In particular, we demonstrate that the mean-variance portfolio weights are dominated by the alpha portfolio weights as the number of assets increases asymptotically. Given the secondary role played by the beta portfolio, we show that it can be replaced, without any loss of efficiency, by a class of benchmark portfolios that by construction are

---

\(^{22}\)There are two possible cases for the set of missing factors, $f_{\text{miss},t}$. Some of them, for instance $f_{\text{miss},1,t}$ could possibly be pervasive while others, $f_{\text{miss},2,t}$, are non-pervasive, where in turn we can split the columns of $A_N = (A_{N_1}, A_{N_2})$ and analogously for $\lambda_{\text{miss}}$ and $\mathbf{z}_{\text{miss},t}$. In particular, using our definition of regularity, one then obtains that as $N \rightarrow \infty$ that $\lim_{N \rightarrow \infty} (A_{N_1}(E[\varepsilon_t \varepsilon_t'])^{-1} A_{N_1}) = \infty$ and $\lim_{N \rightarrow \infty} (A_{N_2}(E[\varepsilon_t \varepsilon_t'])^{-1} A_{N_2}) \leq \delta < \infty$, where $E[\varepsilon_t \varepsilon_t']$ has bounded maximum eigenvalue and $p_1 + p_2 = p$.

\(^{23}\)Of course, it is possible that some of the factors are measured without any error; in that case the means and variances of the components of $\eta_t$ associated with these factors are zero.
FIGURE 2.2 Mean-variance and minimum-variance frontier portfolios

In this figure, we plot three kinds of mean-variance portfolios one can study: (1) the global-minimum-variance portfolio, \( w_N^\text{mv} \); (2) the mean-variance efficient portfolio when a risk-free asset is available, \( w_N^\text{MV} \), which lies on the Capital Market Line; and, (3) the mean-variance efficient frontier portfolios in the absence of a risk-free asset. The figure also shows the tangency portfolio, \( w_N^\text{tan} \), which is a special case of the mean-variance portfolios in which all the wealth is invested in risky assets, with nothing invested in the risk-free asset.

In the rest of this section, we provide the details of these three steps for the mean-variance efficient portfolio in the presence of a risk-free asset. To conserve space, the analysis of the global-minimum-variance portfolio and the frontier portfolios is relegated to the online appendix.

### 2.4.1 Decomposing the mean-variance portfolio

The mean-variance efficient portfolio in the presence of a risk-free asset is defined by the solution to the following optimization problem:

\[
\begin{align*}
    w_N^\text{MV} &= \arg\max_{w_N} \left( w_N' \mu_N + (1 - w_N' 1_N) \gamma - \frac{1}{2} w_N' \Sigma_N w_N \right),
\end{align*}
\]  

(2.25)

where \( 0 < \gamma < \infty \) is the coefficient of risk aversion, \( w_N^\text{MV} = (w_1^\text{MV}, \ldots, w_N^\text{MV})' \) is the vector of portfolio weights in the \( N \) risky assets, and, the investment in the risk-free asset is given by \( 1 - 1_N' w_N^\text{MV} \). Alternatively, if one wished to formulate the above problem in terms of a constraint that required the portfolio to achieve a
target mean of $\mu^*$, one needs to set
\[
\gamma = \frac{(\mu_N - \tau_f 1_N)\Sigma^{-1}_N (\mu_N - \tau_f 1_N)}{\mu^* - \tau_f}.
\] (2.26)

The solution to the optimization problem in (2.25) is \[^24\]
\[
w^\text{mv}_N = \frac{1}{\gamma} \Sigma^{-1}_N (\mu_N - \tau_f 1_N).
\] (2.28)

By standard arguments, the return on portfolio $w^\text{mv}_N$ has a conditional mean, standard deviation, and Sharpe ratio given by the following three expressions:
\[
\mu^\text{mv} - \tau_f = \gamma^{-1} \left( (\mu_N - \tau_f 1_N)\Sigma^{-1}_N (\mu_N - \tau_f 1_N) \right),
\] (2.29)
\[
\sigma^\text{mv} = \gamma^{-1} \left( (\mu_N - \tau_f 1_N)\Sigma^{-1}_N (\mu_N - \tau_f 1_N) \right)^{1/2}, \quad \text{and}
\] (2.30)
\[
\text{SR}^\text{mv} = \left( (\mu_N - \tau_f 1_N)\Sigma^{-1}_N (\mu_N - \tau_f 1_N) \right)^{1/2}.
\] (2.31)

The following theorem, which is valid for any finite $N$, establishes the relations that exist across the mean-variance portfolio, $w^\text{mv}_N$, and the two portfolios that depend on the alpha and beta components of returns, $w^\alpha_N$ and $w^\beta_N$, respectively. The mean-variance portfolio and its decomposition into the “alpha” and “beta” portfolios, is displayed in Figure 2.3.

**Theorem 2.5 (Decomposing weights of mean-variance portfolio)** Suppose that the vector of asset returns, $r_t$, satisfies Assumptions 2.1 and 2.2. Then for any finite $N > K$ and $\mu^* > \tau_f$, the mean-variance portfolio weights satisfy the following decomposition:
\[
w^\text{mv}_N = \phi^\alpha w^\alpha_N + \phi^\beta w^\beta_N,
\] (2.32)
where
\[
w^\alpha_N = \frac{1}{\gamma^\alpha} \Sigma^{-1}_N \hat{\alpha}_N,
\] (2.33)
\[
w^\beta_N = \frac{1}{\gamma^\beta} \Sigma^{-1}_N B_N \hat{\lambda},
\] (2.34)
with $\gamma^\alpha = \frac{\phi^\alpha \Sigma^{-1}_N \hat{\alpha}_N}{\mu^* - \tau_f}$, $\gamma^\beta = \gamma - \gamma^\alpha = \frac{\lambda^* \Sigma^{-1}_N B_N \hat{\lambda}}{\mu^* - \tau_f}$, $\phi^\alpha = \frac{\phi^\alpha \Sigma^{-1}_N \hat{\alpha}_N}{\gamma}$, $\phi^\beta = \frac{\phi^\beta \Sigma^{-1}_N B_N \hat{\lambda}}{\gamma}$, and $\gamma$ defined in (2.26).\[^{25}\] Furthermore, the portfolios $w^\alpha_N$ and $w^\beta_N$ satisfy the orthogonality condition,
\[
(w^\alpha_N)\Sigma_N w^\beta_N = (w^\alpha_N)\Sigma_N w^\beta_N = 0.
\] (2.35)
Moreover, $w^\beta_N$ is the minimum-variance portfolio that is orthogonal to $w^\alpha_N$ and vice versa. Finally, we have two-fund separation: the inefficient portfolios $w^\alpha_N$ and $w^\beta_N$ can generate all the portfolios on the efficient mean-variance frontier of risky assets.

---

\[^{24}\] A special case of the mean-variance portfolio is the tangency portfolio, which has zero wealth invested in the risk-free asset:
\[
w^\text{mv}_N = \frac{w^\text{mv}_N}{1_N w^\text{mv}_N} = \frac{\Sigma^{-1}_N (\mu_N - \tau_f 1_N)}{1_N \Sigma^{-1}_N (\mu_N - \tau_f 1_N)}.
\] (2.27)

\[^{25}\] One can interpret $\gamma^\alpha$ as the ratio of the share of the contribution of $w^\alpha_N$ to the expected return on the mean-variance portfolio with unit risk aversion, $(\mu_N - \tau_f 1_N)\Sigma^{-1}_N (\mu_N - \tau_f 1_N)$, over the target excess mean return, $\mu^* - \tau_f$. The role of the $\phi^\alpha$ and $\phi^\beta$ coefficients is to ensure that the $w^\alpha_N$ and $w^\beta_N$ portfolios achieve the same target mean return as $w^\text{mv}_N$, which is $\mu^*$. 

Figure 2.3 Decomposition of the mean-variance portfolio

In this figure, we plot the mean-variance portfolio in the presence of risk-free asset, $\mathbf{w}^\beta_N$, and its decomposition into two inefficient portfolios: one that depends only on the pricing errors, $\mathbf{w}^\alpha_N$, and another that depends only on the factor exposure and their premia, $\mathbf{w}^\gamma_N$.

Remark 2.5 Note that the portfolio $\mathbf{w}^\alpha_N$, defined in (2.33), depends only on the pricing error but not on the risk premia, $\lambda$, or the factor-covariance matrix, $\Omega$, which is why we label this portfolio the “alpha” portfolio. Analogously, the portfolio strategy $\mathbf{w}^\beta_N$, defined in (2.34), depends on factor exposures and their risk premia, but not on the pricing errors, $\alpha_N$. For practical portfolio construction and risk management, what this implies is that if the alpha portfolio is constructed using the expression in (2.33), then it will not have any exposure to factor risk because $\mathbf{w}^\alpha_N' \mathbf{B}_N = 0$.

Moreover, as illustrated in Figure 2.4, the weights of the $\mathbf{w}^\beta_N$ portfolio are typically small and positive, while the weights of the $\mathbf{w}^\alpha_N$ portfolio are large and take both positive and negative values implying that the alpha portfolio has both long and short positions.

Remark 2.6 The orthogonality condition in (2.35) above says that the two portfolios $\mathbf{w}^\alpha_N$ and $\mathbf{w}^\beta_N$ are uncorrelated, both conditionally on the factors and also unconditionally. In addition to $\mathbf{w}^\alpha_N$ and $\mathbf{w}^\beta_N$ being orthogonal to each other, if one searched for the minimum-variance portfolio that is orthogonal to $\mathbf{w}^\alpha_N$, the resulting portfolio would be $\mathbf{w}^\beta_N$, and vice versa. That is, even though the $\mathbf{w}^\alpha_N$ and $\mathbf{w}^\beta_N$ portfolios are obtained simply by relying on the APT decomposition of the total mean return, these portfolios can also be characterized as being the result of an optimization, which is described in Lemma 2.2 and that extends Roll (1980) to the case where, in addition to investing in risky assets, one can also invest in a risk-free asset. This mutual optimality property of the $\mathbf{w}^\alpha_N$ and $\mathbf{w}^\beta_N$ portfolios drives the two-fund separation result: the alpha and beta portfolios, both of which are inefficient, span the entire efficient frontier of risky assets.

We now characterize the returns of the two components of the mean-variance portfolio.
2.4. MITIGATING MODEL MISSPECIFICATION

Figure 2.4 Typical weights of the $w^*_N$ and $w^\beta_N$ portfolios

In this bar chart, we plot the typical weights of the $w^*_N$ (gray bars) and $w^\beta_N$ portfolios (black bars) for the case in which the number of assets is $N = 20$. The figure shows that the weights of the $w^*_N$ portfolio are small and positive. In contrast, the weights of the $w^\beta_N$ portfolio are large and take both positive and negative values.

Theorem 2.6 (Decomposing returns of mean-variance portfolio) Suppose that the vector of asset returns, $\mathbf{r}$, satisfies Assumptions 2.1 and 2.2, and $\alpha_N \neq 0$. Then, for any finite $N > K$, and assuming $\mu^* > r_f$, the returns of the inefficient portfolios $w^*_N$ and $w^\beta_N$ have a mean, volatility (standard deviation), and Sharpe ratio that have the same quadratic form as the corresponding expressions for the efficient $w^0_N$ given in (2.29), (2.30), and (2.31):

$$\mu^* - r_f = \frac{1}{\gamma^\alpha} \alpha_N \Sigma_N^{-1} \alpha_N = \mu^* - r_f; \quad \mu^\beta - r_f = \frac{1}{\gamma^\beta} \lambda^\beta \mathbf{B}_N^T \mathbf{V}_N^{-1} \mathbf{B}_N \lambda = \mu^* - r_f; \quad (2.36)$$

$$\sigma^\alpha = \frac{1}{\gamma^\alpha} (\alpha_N \Sigma_N^{-1} \alpha_N)^{1/2}; \quad \sigma^\beta = \frac{1}{\gamma^\beta} (\lambda \mathbf{B}_N^T \mathbf{V}_N^{-1} \mathbf{B}_N \lambda)^{1/2}; \quad (2.37)$$

$$\text{SR}^\alpha = (\alpha_N \Sigma_N^{-1} \alpha_N)^{1/2}; \quad \text{SR}^\beta = (\lambda \mathbf{B}_N^T \mathbf{V}_N^{-1} \mathbf{B}_N \lambda)^{1/2}; \quad (2.38)$$

with the Sharpe ratios satisfying: $0 \leq \text{SR}^\alpha < \infty; \quad 0 \leq \text{SR}^\beta < \infty; \quad$ and

$$(\text{SR}^{\text{SR}})^2 = (\text{SR}^\alpha)^2 + (\text{SR}^\beta)^2. \quad (2.39)$$

Remark 2.7 An important insight is that the quantity on the left-hand side of the APT restriction in (2.8), $\alpha_N \Sigma_N^{-1} \alpha_N$, is exactly the same as the square of $\text{SR}^\alpha$ in (2.38). Thus, the APT restriction in (2.8), which is typically interpreted as a bound on the pricing errors, can instead be interpreted as a bound on the Sharpe ratio of the $w^*_N$ portfolio. The decomposition of the square of the Sharpe ratio of the mean-variance portfolio in (2.39) is obtained also in Treynor and Black (1973) for the case of the single index model with a diagonal covariance matrix for the residuals. Gibbons et al. (1989b, p. 115) also recognize that
\[ \hat{\alpha}_N \Sigma_N^{-1} \alpha_N = (SR^{\alpha})^2 - (SR^{\beta})^2 \] but do not interpret the left-hand side as the square of the Sharpe ratio of a portfolio, in particular, the alpha portfolio. In Lemma 2.1, we provide the general conditions that are needed for decomposing the squared Sharpe ratio of any portfolio that can be written as the sum of two orthogonal components.\textsuperscript{26}

### 2.4.2 Mitigating misspecification in the beta component of returns

To treat misspecification arising from the beta component of returns, we study the mean-variance portfolio weights for the case in which the number of assets is asymptotically large.

For this analysis, we need to extend Definition 2.2 of a regular economy in two different dimensions. The earlier definition of regularity was applied to \( B_N \) and \( \Sigma_N \). One, we extend the definition to any arbitrary matrix of dimension \( N \times K \), such as \( D_N \), and an arbitrary positive-definite \( N \times N \) matrix, \( C_N \). Two, we impose that all the eigenvalues are diverging at precisely the same rate.\textsuperscript{27}

**Definition 2.7 (C\(_N\)-regularity)** A matrix \( D_N \) is \( C_N \)-regular if there exists an increasing function of \( N \), \( f(N) \), such that for any \( 1 \leq j \leq K \), the eigenvalues \( g_{jK}(\frac{1}{f(N)}D_N^{-1}C_N^{-1}D_N) \to \delta_j > 0 \), where \( \delta_j \) is some finite positive constant.

We now state our result about the asymptotic properties of mean-variance portfolio weights.

**Theorem 2.8 (Weights of alpha, beta, and mean-variance portfolios for large \( N \))** Suppose that the vector of asset returns, \( r_t \), satisfies Assumptions 2.1 and 2.2 and \( \alpha_N \neq 0 \). Suppose also that \( A_N \), \( B_N \), and \( 1_N \) are \( C_N \)-regular with the same scaling factor \( f(N) \), and \( \alpha_N \) and \( \beta_N \) are not asymptotically collinear.\textsuperscript{28}

As \( N \to \infty \), then:

(i) \( 0 \leq \phi^\alpha \leq 1, \ 0 \leq \phi^\beta \leq 1 \), and, element-by-element,

\[
\frac{w^{\beta}_{N,i}}{w^{\alpha}_{N,i}} \to 0.
\]

(ii) The sum of the squared components of the mean-variance portfolio vectors \( \mathbf{w}^{\alpha}_N \mathbf{w}^{\alpha}_N \) is always bounded, whereas \( \mathbf{w}^{\beta}_N \mathbf{w}^{\beta}_N \) always converges to zero.

(iii) The sum of the components of the mean-variance portfolio vectors \( |1_N \mathbf{w}^{\alpha}_N| \) can diverge to infinity, whereas \( |1_N \mathbf{w}^{\beta}_N| \) is always bounded.

(iv) The vector of weights for the mean-variance portfolio are asymptotically equivalent, element-by-element, to the weights of \( \phi^\alpha \mathbf{w}^{\alpha}_N \):

\[
\mathbf{w}^{\text{MV}}_{N,i} = (1 - \phi^\beta)w^{\alpha}_{N,i} + \phi^\beta w^{\beta}_{N,i} \sim (1 - \phi^\beta)w^{\alpha}_{N,i} = \phi^\alpha w^{\alpha}_{N,i},
\]

where the symbol \( \sim \) denotes asymptotic equivalence.\textsuperscript{29}

\textsuperscript{26}The bound on the square of the Sharpe ratio of the alpha portfolio can also be seen as providing a theoretical rationalization for the no-good-deal bound in Cochrane and Saa-Queipo (2001). In our discussion of the related literature in Section 2.2, we also explain how imposing the APT restriction is analogous to the approach adopted in Garlappi et al. (2007), in which one accounts for parameter uncertainty in portfolio choice using the minmax approach originally proposed in Gilboa and Schmeidler (1989).

\textsuperscript{27}A special case of the definition below is the notion of pervasive factors defined in Connor and Korajczyk (1986, Assumption 6) and Connor et al. (2010, p. 85). Both papers use \( f(N) = N \), but Connor and Korajczyk require all the eigenvalues to diverge at least at that rate, whereas Connor et al. use a definition similar to ours in the sense that it requires all eigenvalues to diverge at precisely the same rate.

\textsuperscript{28}By asymptotic collinearity we mean that either \( A_N' \Sigma_N A_N \to 0 \) or \( B_N' \Sigma_N B_N \to 0 \) or both, as \( N \) diverges, depending on whether the number of unobserved factors \( p \leq K \), \( p \geq K \) or \( p = K \), where \( M_C = I_N - C(C' C)^{-1}C' \) is the matrix that spans the space orthogonal to any full-column-rank matrix \( C \). When \( p \leq K \), a sufficient condition for this is \( A_N = B_N \delta + C_N \) for some constant \( K \times p \) matrix \( \delta \) and some residual matrix \( C_N \) satisfying \( C_N' C_N \to 0 \).

\textsuperscript{29}We say that \( a_n \sim b_n \), if \( a_n/b_n \to 1 \) as \( n \to \infty \).
2.4. MITIGATING MODEL MISSPECIFICATION

Figure 2.5  Relative average magnitude of weights of \( w_N^\alpha \) and \( w_N^\beta \) portfolios

In this figure, we plot three quantities as the number of assets, \( N \), increases. The three quantities are: (1) the average magnitude of the weights of the \( w_N^\alpha \) portfolio, given by the dotted (red) line at the top of the figure; (2) the average magnitude of the weights of the \( w_N^\beta \) portfolio, given by the dashed (blue) line at the bottom of the figure; and (3) the ratio of the average magnitude of the weights of the \( w_N^\alpha \) portfolio to the corresponding weights of the \( w_N^\beta \) portfolio, given by the dotted-dashed (green) line. The figure shows that as \( N \) increases, the average magnitude of the weights of the \( w_N^\alpha \) portfolio declines faster than the average magnitude of the weights of the \( w_N^\beta \) portfolio.

Remark 2.8  The \( w_N^\alpha \) portfolio dominates the \( w_N^\beta \) portfolio across all three norms considered in the theorem above; this dominance is illustrated in Figure 2.5. The notion of diversification used in part (ii) of the theorem is the sum of the squares, which is the same notion adopted in Chamberlain (1983b). Because \( w_N^\beta \) \( \geq \) \( \sup_i |w_N^{\alpha i}| \), it follows that the \( w_N^\beta \) portfolio is diversified according to the sup norm criterion, which is the norm used in Green and Hollifield (1992). In contrast, the \( w_N^\alpha \) portfolio is not necessarily diversified according to the squared norm.\(^{30}\) Part (iii) of the theorem studies how the total investment in risky assets is allocated between the \( w_N^\alpha \) and \( w_N^\beta \) portfolios. The result in the theorem shows that \( 1_N w_N^\alpha \) could be greater than 1 and it could be growing without bound as \( N \) increases, implying that it may be optimal to lever up unboundedly the investment in the \( w_N^\alpha \) portfolio.\(^{31}\) On the other hand, the investment in the \( w_N^\beta \) portfolio is bounded, and hence, is associated with a finite amount of leverage.

In order to get a better understanding of the result in (2.40), in the corollary below we look at a special case where \( f(N) = N \). We consider only part (i) of the theorem, because the other parts of the theorem are

\(^{30}\)For example, there could be a finite number of assets with a sufficiently large alpha, in which case the weights of these assets will not go to zero. Alternatively, even if none of the assets has a particularly large alpha, the weights of the \( w_N^\alpha \) portfolio can go to zero at a sufficiently slow rate, as slow as \( 1/\sqrt{N_1} \), as shown in Corollary 2.8.1.

\(^{31}\)To see that \( 1_N w_N^\alpha \) can diverge, consider the following example in which \( \alpha_N = 1_N / \sqrt{N} \), \( \Sigma_N = \sigma^2 I_N \), with \( I_N \) the \( N \times N \) identity matrix, and there is a single factor with \( \beta \), with iid distribution having mean 1 and variance \( \sigma_\beta^2 > 0 \). Then, \( 1_N w_N^\alpha \sim \sqrt{N} \sigma_\beta^2 / (1 - \sigma_\beta^2) \) which goes to infinity with \( N \).
unchanged under the special case. Then, we provide the intuition for the result in the theorem above and the corollary.

**Corollary 2.8.1 (Weights of alpha and beta portfolios for large N: Special case)** Suppose that the assumptions of Theorem 2.8 are satisfied and that the row sums of $A_N, B_N$ and $C_N^{-1}$ are uniformly bounded.\(^{32}\) Suppose also that $f(N) = N$. Then, as $N \to \infty$:

- for the case of bounded-residual variation, the absolute value of the components of the mean-variance portfolio vectors $w_N^\alpha$ and $w_N^\beta$ decrease at most at the rate\(^{33}\)

$$|w_{N,i}^\alpha| = O\left|\left[1,\Sigma_N^{-1} a_N\right] + \frac{1}{N^2}\right|$$

and

$$|w_{N,i}^\beta| = O\left(\frac{1}{N}\right),$$

and

$$|w_{N,i}^\alpha| = O\left(\frac{1}{N}\right), \quad \text{and} \quad |w_{N,i}^\beta| = O\left(\frac{1}{N}\right).$$

(2.42)

- for the case of unbounded-residual variation they decrease at most at the rate

$$|w_{N,i}^\alpha| = O\left(\frac{1}{N}\right), \quad \text{and} \quad |w_{N,i}^\beta| = O\left(\frac{1}{N}\right).$$

(2.43)

To understand the intuition for the results about the dominance of the $w_N^\alpha$ portfolio weights, recall that the objective of mean-variance portfolio optimization is to maximize the portfolio Sharpe ratio, which entails increasing the mean of the portfolio return and/or reducing the volatility of the portfolio return. There are two sources of risk: factor exposure and idiosyncratic exposure. The factor exposure of the $w_N^\alpha$ portfolio is zero—irrespective of the rate at which the weights decrease—because of the orthogonality of $w_N^\alpha$ to $B_N$. Regarding exposure to idiosyncratic risk, the elements of $w_N^\alpha$ cannot decrease faster than $1/N^2$ because then the idiosyncratic risk of the portfolio goes to zero; however, the idiosyncratic risk of the alpha portfolio coincides with its Sharpe ratio, implying that the Sharpe ratio would also go to zero. On the other hand, the APT restriction does not allow the rate at which the weights decrease to be slower than $1/N^2$. To understand this, consider the simple case in which $\Sigma_N$ is the identity matrix. Recall that for the sum of $N$ positive terms to be bounded, it suffices that each term declines not slower than $1/N$. The square-root rate follows from the fact that in our case each term is in fact the square of $w_{N,i}^\alpha$. Thus, the rate of $1/N^2$ strikes just the correct balance between optimizing the risk and return of the $w_N^\alpha$ portfolio.

Let us now look at the $w_N^\beta$ portfolio. If the weights decrease at any rate slower than $1/N$, then the systematic exposure explodes because the factors are pervasive (see Definition 2.2). On the other hand, if the weights decrease faster than $1/N$, then the portfolio risk declines to zero, leading to a Sharpe ratio of zero because the expression for the Sharpe ratio is exactly the same as the one for the risk of the portfolio. So, the rate of $1/N$ strikes the correct balance between optimizing the risk and return of the $w_N^\beta$ portfolio.\(^{34}\) Notice that the rate $1/N$ makes the $w_N^\beta$ portfolio well diversified, even with respect to idiosyncratic exposure, enhancing its Sharpe ratio even further.

Above, we have shown that the mean-variance portfolio weights are dominated by the alpha portfolio weights as the number of assets increases asymptotically. Given the secondary role played by the beta portfolio, we show below that under certain conditions it can be replaced, without any loss of efficiency, by a class of benchmark portfolios that by construction are independent of the distribution of the observed factors, $\lambda$ and $\Omega$, and hence, immune to beta mispecification. This demonstrates that one can construct optimal portfolios that are independent of risk premia.

---

\(^{32}\)Given an $N \times M$ matrix $D$, we say its row sums are uniformly bounded when $\sup_{1 \leq j \leq N} \sum_{i=1}^M |d_{ij}| \leq \delta < \infty$, for some arbitrary $\delta$.

\(^{33}\)For any finite dimensional non-negative $a_N$ and $b_N$, $a_N = O(b_N)$ means that $a_N/b_N \leq \delta < \infty$, for any constant $\delta > 0$.

\(^{34}\)We can see the above argument also in the expression of the portfolio weight: $w_N^\beta = V_N^{-1}(B_N \lambda) = (B_N \Omega B_N^{-1} + \Sigma_N^{-1})(B_N \lambda)$. Because $B_N$ appears twice in the denominator of $w_N^\beta$, it causes its faster decay to zero. On the other hand, $w_N^\alpha = \Sigma_N^{-1} \delta_N$. However, only the $A_N \lambda_{\text{miss}}$ part of $\delta_N$ can appear in the denominator of $w_N^\alpha$, implying that $w_N^\alpha$ decays to zero slowly whenever $a_N \neq 0_N$. 
2.4. MITIGATING MODEL MISSPECIFICATION

Theorem 2.9 (Weight and Sharpe ratio of mean-variance portfolio for large $N$) Suppose that the vector of asset returns, $r_t$, satisfies Assumptions 2.1 and 2.2 and $\alpha_N \neq 0$. Suppose further that the investor holds a well-diversified benchmark portfolio $w_{bench}^N$ satisfying the following properties:

$$ (w_{bench}^N)'\alpha_N \to 0, \quad B_N w_{bench}^N \to c_{bench}^N, \quad (w_{bench}^N)'\Sigma_N w_{bench}^N \to 0, \quad (2.44) $$

where $c_{bench}$ is a $K \times 1$ vector of constants, different from the zero vector, satisfying $X'c_{bench} \neq 0$. Let $SR_{bench}^*$ be the Sharpe ratio corresponding to the benchmark portfolio weights $w_{bench}^N$ modified to achieve the target mean $\mu^*$; that is, $w_{bench}^N = \frac{\mu^* - r_f}{(w_{bench}^N)'(\mu_N - r_f 1_N)} w_{bench}^N$. (i) If $c_{bench}$ is perfectly proportional to the vector $\Omega^{-1}\lambda$, then $w_{N,i}^{mv}$ is dominated by $w_{N,i}^{\alpha}$:

$$ w_{N,i}^{mv} \sim (1 - \phi_{bench}^*) w_{N,i}^\alpha + \phi_{bench}^* w_{N,i}^{bench}, \quad \phi_{bench}^* \sim (1 - \phi_{bench}^*) w_{N,i}^\alpha, \quad (2.45) $$

but both $SR^*$ and $SR_{bench}^*$ contribute to $SR_{mv}$,

$$ (SR_{mv})^2 \sim (SR^*)^2 + (SR_{bench}^*)^2. \quad (2.46) $$

Moreover, if $K = 1$, then $c_{bench}$ is always perfectly proportional to the vector $\Omega^{-1}\lambda$.

(ii) If $c_{bench}$ is not proportional to the vector $\Omega^{-1}\lambda$, then $w_{N,i}^{mv}$ is not asymptotically equivalent to $(1 - \phi_{bench}^*) w_{N,i}^\alpha + \phi_{bench}^* w_{N,i}^{bench}$ and

$$ (SR_{mv})^2 \sim (SR^*)^2 + (SR_{bench}^*)^2. \quad (2.47) $$

Remark 2.9 The first assumption in (2.44) implies that the benchmark portfolio is asymptotically orthogonal to $\alpha_N$. The second assumption rules out that the benchmark portfolio return is equal to the risk-free return in the limit. The third assumption requires that the benchmark portfolio be well diversified. Note that for the unbounded-variation case, the first assumption is satisfied whenever $(w_{bench}^N)'\alpha_N \to 0$ and $(w_{bench}^N)'1_N \to 0$, where the latter condition ensures that $w_{bench}^N$ diversifies away the contribution of the latent factors, $\alpha_N$, to $\Sigma_N$.

Remark 2.10 To construct a valid benchmark portfolio, one can rely on the insights of Treynor and Black (1973) and DeMiguel et al. (2006b). The results of Treynor and Black (1973) can be interpreted as saying that $w_{N,i}^0$ can be approximated by a portfolio that is similar to the market portfolio, $w_{N,k}^{bkt}$, suitably normalized to achieve a target mean of $\mu^*$. Alternatively, the empirical findings of DeMiguel et al. (2006b) suggest that one could hold an equally weighted portfolio suitably normalized, implying that $w_{bench}^N = \frac{\mu^* - r_f}{(1_N)'(\mu_N - r_f 1_N) N} 1_N$. In the theorem above, we formalize these two proposals by reporting the results for any arbitrary benchmark portfolio with target mean $\mu^*$. We show the condition under which a benchmark portfolio, combined with the alpha portfolio, will coincide asymptotically with the optimal mean-variance portfolio. This condition is always satisfied when there is only a single factor, that is, $K = 1$.

Remark 2.11 The assumptions in (2.44) imply that the return on the benchmark portfolio is asymptotically equivalent to the return on the portfolio of factors with weight $c_{bench}$; that is, $(w_{bench}^N)'(r_t - r_f 1_N) = (c_{bench})'(r_t - r_f 1_N) + o_p(1)$. Therefore, asymptotic optimality of the benchmark portfolio requires that $c_{bench}$ equals the mean-variance portfolio constructed using the $K$ factors. This choice guarantees that the benchmark portfolio achieves the largest possible Sharpe ratio, as stated in part (i) of the theorem. Figure 2.6 shows that as the number of risky assets increases, the ratio of $(SR_{mv})^2$ to $(SR^*)^2 + (SR_{bench}^*)^2$ quickly approaches 1.

Remark 2.12 In striking contrast to the portfolio weights $w_{N,i}^0$ and $w_{N,i}^\alpha$, in which the components of $w_{N,i}^\alpha$ asymptotically dominate those of $w_{N,i}^0$, their Sharpe ratios are, in general, of the same order of magnitude even for large $N$. The reason is that the (excess) return on the portfolio $w_{N,i}^\alpha$ is $(\lambda_i + z_i)'B_N w_{N,i}^0 + \varepsilon_i w_{N,i}^\alpha$. Thus, even though the variance of the idiosyncratic component of the return, $\varepsilon_i w_{N,i}^\alpha$, is diversified away as we increase the number of assets, as long as some of the factors are pervasive, the term $B_N w_{N,i}^\alpha$ will not be diversified away. Hence, the excess mean return and Sharpe ratio of $w_{N,i}^\alpha$ will not go to zero.
Figure 2.6 Sharpe ratio of \( w_N^{\text{opt}} \) and \( w_N^{\text{bench}} \) relative to that of \( w_N^{\text{opt}} \) portfolio

This figure shows that as the number of assets, \( N \), increases, the sum of the squares of the Sharpe ratios of the \( w_N^{\text{opt}} \) and \( w_N^{\text{bench}} \) portfolios converges to the square of the Sharpe ratio of the mean-variance portfolio.

Having addressed the problem of misspecification arising from the beta component of returns, we now consider misspecification in the alpha component of returns.

### 2.4.3 Mitigating misspecification in the alpha component of returns

In this section, we explain how to treat misspecification arising from the alpha component of returns.

Section 2.3.5 explains how the vector of pricing errors, \( \alpha_N \), captures model misspecification arising from missing factors and asset-specific characteristics unrelated to factors. The analysis in the previous section, Section 2.4.2, demonstrates that as the number of assets increases, the weights of the \( w_N^{\text{opt}} \) portfolio dominate the weights of the \( w_N^{\beta} \) portfolio. That is, the vector of pricing errors, \( \alpha_N \), plays a dominant role in the choice of optimal mean-variance portfolio weights and their returns. Therefore, it is vital to estimate \( \alpha_N \) precisely. The APT restriction, reproduced below,

\[
\alpha_N' \Sigma_N^{-1} \alpha_N \leq \delta < \infty,
\]

provides exactly the condition that must be imposed in the estimation of the factor model generating returns. In principle, the constraint in (2.48) binds when \( N \to \infty \) because the theory does not specify a particular value for \( \delta \) (though, from the result in (2.39) we know that \( \delta \) is less than the square of the Sharpe ratio of the market portfolio, if the market is an efficient portfolio). Following the common practice in applied-econometrics and statistics of using asymptotic standard errors even though the sample size is finite, we impose the APT restriction when estimating the model, which, of course, is always for a finite number of assets.

The way in which we impose the APT restriction given in (2.48) depends on whether we are in the case of bounded or unbounded residual variation. We propose a multi-step procedure to determine in which
case we are. In the first step, one estimates the parameters of the factor model conditional on the factor realizations without imposing the APT restriction. Having obtained consistent estimates of the parameters, the next step is to analyze the possibility of pervasive missing factors by studying the eigenvalues associated with the estimated $\Sigma_N$, where the $\hat{}$ denotes an estimated quantity. This part uses conventional principal-component analysis of $\Sigma_N$, and it allows one to determine the number of latent pervasive factors, $p$; see, for example, Anderson (1984). In the next two subsections, we provide the details of how to estimate the model when $p = 0$ (bounded-variation case) and when $p > 0$ (unbounded-variation case), after imposing the APT constraint. To conserve space, the case where the factors themselves are measured with error is discussed in Appendix 2.4.12.

Estimation for the case of bounded residual variation ($p = 0$)

In the bounded-residual-variation case (that is, $\alpha_N = a_N$), the true unconditional means and covariances of returns satisfy

$$E(r_t - \gamma_1 1_N) = \mu_0 - \gamma_1 1_N = \alpha_0 + B_0 \lambda_0, \quad \text{Var}(r_t) = V_0 = B_0 \Omega_0 B_0' + \Sigma_0,$$

where $\lambda_0 = E(f_t) - \gamma_1 1_K$ is the vector of risk premia, $\Omega_0 = \text{Var}(f_t)$ is the covariance matrix of the factors assuming, without loss of generality, stationarity of the $K$ factors, $f_t$ and that the factors are traded. As before, the subscript "0" indicates the true value of a parameter and we do not use the subscript "N" for the true value of parameters in order to limit the number of subscripts.

In order to identify $\lambda_0$ and $\Omega_0$, one needs also to consider the information stemming from the sample observations for $f_t$. Although our argument applies to virtually any estimation procedure, we will illustrate it with respect to the (pseudo) Gaussian ML estimator. This is a natural estimator for our model when the first two moments of asset returns are specified correctly, although distributional assumptions (such as normality) are not required; hence, the use of pseudo ML.

The (pseudo) ML estimator, based on the unconditional joint distribution of $\begin{pmatrix} r_t \\ f_t \end{pmatrix}$ and assuming i.i.d. residuals for simplicity, is

$$L(\theta) = -\frac{1}{2} \log(\text{det}(\Sigma_N))$$

$$-\frac{1}{2T} \sum_{t=1}^{T} \left( r_t - \gamma_1 1_N - \alpha_N - B_N(f_t - \gamma_1 1_K) \right)' \Sigma_N^{-1} \left( r_t - \gamma_1 1_N - \alpha_N - B_N(f_t - \gamma_1 1_K) \right)$$

$$-\frac{1}{2} \log(\text{det}(\Omega)) - \frac{1}{2T} \sum_{t=1}^{T} \left( f_t - \gamma_1 1_K - \lambda \right)' \Omega^{-1} \left( f_t - \gamma_1 1_K - \lambda \right),$$

where $\theta = (\alpha_N', \text{vec}(B_N)', \text{vec}(\Sigma_N)', \lambda', \text{vec}(\Omega))'$. Therefore, the ML estimators for $\alpha_0, B_0, \Sigma_0$ coincide with the OLS estimators, conditional on the realization of the factors. On the other hand, the ML estimators for $\lambda_0$ and $\Omega_0$ are the sample mean and sample covariance of the factors $f_t$.

However, because the APT restriction is not guaranteed to hold, one should consider the ML-constrained (MLC) estimator:

$$\tilde{\theta}_{MLC} = \arg\max_{\theta} L(\theta) \text{ such that } \alpha_N' \Sigma_N^{-1} \alpha_N \leq \delta.$$
Because the parameter \( \alpha_0 \) is constrained only by the APT restriction, imposing this constraint may lead \( \theta_{\text{MLC}} \) to be a more precise estimator of the true parameter values compared to the unconstrained estimator, \( \theta_{\text{ML}} \). The theory does not tell us what \( \delta \) should be in (2.52). As discussed earlier, we know that the upper bound of \( \delta \) must be less than the square of the Sharpe ratio of the mean-variance efficient portfolio; in our empirical application, we choose \( \delta \) using cross-validation techniques.

To impose the APT restriction one can consider a penalized log-likelihood function as follows in Theorem 2.10.

**Theorem 2.10** (Parameter identification by imposing asset-pricing restriction: Bounded-variation case) Suppose that the vector of asset returns, \( r_t \), satisfies Assumption 2.1. Given any \( \kappa > 0 \),

\[
\hat{\theta}_{\text{MLC}} = \arg\max_{\theta} \left\{ L(\theta) - \kappa(\alpha'_N \Sigma^{-1}_N \alpha_N - \delta) \right\},
\]

where \( L(\theta) \) is defined in (2.51). If \( \left( \sum_{t=1}^T f_t f'_t \right) \) is nonsingular, then \( \hat{\theta}_{\text{MLC}} = (\hat{\alpha}'_N, \text{vec}(\hat{\mathbf{B}}_{\text{MLC}}), \text{vec}(\hat{\Sigma}_{\text{MLC}}))' \), \( \text{vech}(\hat{\mathbf{B}}_{\text{MLC}}) \), \( \text{vech}(\hat{\Sigma}_{\text{MLC}}) \)' exists, where:

\[
\hat{\alpha}_N_{\text{MLC}} = \frac{1}{1 + \kappa} \left[ \bar{r} - r_f 1_N - \hat{\mathbf{B}}_{\text{MLC}} (\bar{f} - r_f 1_K) \right],
\]

\[
\hat{\mathbf{B}}_{N_{\text{MLC}}} = \left( \sum_{t=1}^T f_t f'_t \right) \left( \sum_{t=1}^T f_t f'_t \right)^{-1}, \quad \text{and}
\]

\[
\hat{\Sigma}_{N_{\text{MLC}}} = \frac{1}{T} \sum_{t=1}^T (\bar{r}_t - \hat{\mathbf{B}}_{N_{\text{MLC}}} \bar{f}_t) (\bar{r}_t - \hat{\mathbf{B}}_{N_{\text{MLC}}} \bar{f}_t)',
\]

where \( \bar{f} = T^{-1} \sum_{t=1}^T f_t, \bar{r}_N = T^{-1} \sum_{t=1}^T r_t, \bar{f}_t = f_t - r_f 1_K - \frac{1}{(1 + \kappa)} (\bar{f} - r_f 1_K), \bar{r}_t = r_t - r_f 1_N - \frac{1}{(1 + \kappa)} (\bar{r}_N - r_f 1_N) \), and the MLC estimators \( \hat{\lambda}_{\text{MLC}} \) and \( \text{vech}(\hat{\Omega}_{\text{MLC}}) \)' coincide with the sample mean and covariance of the factors \( f_t \).

**Remark 2.13** The constrained estimator \( \hat{\alpha}_N_{\text{MLC}} \) turns out to be precisely the ridge estimator for \( \alpha_0 \), because \( \kappa > 0 \).\(^{38}\) The estimators of \( \hat{\mathbf{B}}_{N_{\text{MLC}}} \) and \( \hat{\Sigma}_{N_{\text{MLC}}} \) are also functions of \( \kappa \) because of the APT constraint, in contrast to \( \hat{\lambda}_{\text{MLC}} \) and \( \hat{\Omega}_{\text{MLC}} \), which are simply the sample mean and sample covariance of the factors \( f_t \), because the APT constraint does not affect the distribution of the factors \( f_t \).

**Estimation for the case of unbounded residual variation with missing factors (p > 0)**

The second case of alpha misspecification is of unbounded residual variation, which arises when there are \( p > 0 \) missing pervasive factors. For the case in which the pricing error is unrelated to the missing factors, \( a_N \) is zero in (2.15), we get that \( \alpha_N = A_N \lambda_{\text{miss}} \) and \( \Sigma_N = A_N A'_N + C_N \), where \( \lambda_{\text{miss}} \) is the risk premia corresponding to the missing factors, and \( C_N \) is an \( N \times N \) positive-definite matrix with bounded eigenvalues that represents the covariance matrix of the pure idiosyncratic component of the error returns. Observe that \( \alpha_N \) is a component of the expected return, \( \mu_N \); likewise, \( \Sigma_N \) is a component of the return-covariance matrix, \( \Sigma_N \). Hence, \( A_N \) appears in both the mean and covariance matrix of returns. MacKinlay and Pástor (2000) use this insight to improve the precision of the estimated \( A_N \) parameters, which, in turn, substantially

\(^{38}\)One approach is to solve for the Karush-Kuhn-Tucker multiplier \( \hat{\lambda}_{\text{MLC}} \). An iterative procedure is required to solve for \( \hat{\lambda}_{\text{MLC}} \) and \( \hat{\mathbf{B}}_{N_{\text{MLC}}} \) jointly; then the estimators \( \hat{\alpha}_{N_{\text{MLC}}} \) and \( \hat{\Sigma}_{N_{\text{MLC}}} \) follow. Instead, we choose \( \kappa \) by cross-validation, which is computationally simpler. The formulae for the estimators of the other parameters remain the same. The cross-validation approach is closer to a lasso (least absolute shrinkage and selection operator) formulation in which one considers the penalized log-likelihood \( L(\theta) - \kappa(\alpha'_N \Sigma^{-1}_N \alpha_N) \); however, in contrast to lasso, our constraint is quadratic.
improves the performance of the estimated portfolio.\textsuperscript{39} Importantly, using the Sherman-Morrison-Woodbury formula, it follows that

$$\alpha' \Sigma_N^{-1} \alpha_N = \lambda_{\text{miss}}^{-1} A_N \Sigma_N^{-1} A_N \lambda_{\text{miss}} = \lambda_{\text{miss}}^{-1} (I_p + \alpha' N C_N^{-1} A_N)^{-1} (A' N C_N^{-1} A_N) \lambda_{\text{miss}}.$$  \hspace{1cm} (2.57)

Thus, $\alpha' \Sigma_N^{-1} \alpha_N$ converges to $\lambda_{\text{miss}}^{-1} \lambda_{\text{miss}}$ as $N \to \infty$ because $(I_p + \alpha' N C_N^{-1} A_N)^{-1} (A' N C_N^{-1} A_N)$ converges to the identity matrix, given that the missing factors are pervasive, implying that $(A' N C_N^{-1} A_N)$ is increasing without bound. This means that the APT restriction is always satisfied for the case of only missing pervasive factors (that is, the case in which $a_N = 0$), once we recognize that $\Sigma_N$ contains the loadings of the missing factors, $A_N$.

However, for the general unbounded-variation case where the pricing error consists of both missing factors and a component that is unrelated to factors, $\alpha_N = A_N \lambda_{\text{miss}} + a_N$, the APT restriction is not automatically satisfied. Therefore, when estimating the model, we need to impose the additional constraint: $a'_N a_N \leq \delta < \infty$ for any $N$. Under the same assumptions as above concerning the $K$ observed factors $f_t$, the true unconditional means and covariances of returns now satisfy the equations below, where $C_0$ has bounded maximum eigenvalue, in contrast to $\Sigma_0$:

$$E(r_t - r_f 1_N) = \mu_0 - r_f 1_N = A_0 \lambda_{\text{miss}0} + a_0 + B_0 \lambda_0, \quad \text{Var}(r_t) = V_0 = B_0 \Omega B'_0 + A_0 A'_0 + C_0. \hspace{1cm} (2.58)$$

As in the previous case of bounded variation, the joint log-likelihood function $L(\theta)$ can be decomposed as follows:

$$L(\theta) = -\frac{1}{2} \log(\det(A_N A'_N + C_N)) - \frac{1}{2T} \sum_{t=1}^{T} \left( r_t - r_f 1_N - A_N \lambda_{\text{miss}} - a_N - B_N (f_t - r_f 1_K) \right)'$$
$$\times (A_N A'_N + C_N)^{-1} \left( r_t - r_f 1_N - A_N \lambda_{\text{miss}} - a_N - B_N (f_t - r_f 1_K) \right)$$
$$- \frac{1}{2} \log(\det(\Omega)) - \frac{1}{2T} \sum_{t=1}^{T} \left( f_t - r_f 1_K - \lambda \right)' \Omega^{-1} \left( f_t - r_f 1_K - \lambda \right). \hspace{1cm} (2.59)$$

$$L(\theta) = -\frac{1}{2} \log(\det(\Omega)) - \frac{1}{2T} \sum_{t=1}^{T} \left( f_t - r_f 1_K - \lambda \right)' \Omega^{-1} \left( f_t - r_f 1_K - \lambda \right). \hspace{1cm} (2.60)$$

Without loss of generality, one can assume that the missing factors are uncorrelated with each other and have unit variance, achieving identification of $A_0$.\textsuperscript{40} However, $\lambda_{\text{miss}0}$ and $\alpha_0$ cannot be identified separately unless the APT restriction is imposed, as shown in Theorem 2.11.

**Theorem 2.11 (Unbounded-variation case)** Suppose that the vector of asset returns, $r_t$, satisfies Assumption 2.1. Given any $\kappa \geq 0$, then

$$\hat{\theta}_{\text{MLC}} = \arg\max_{\theta} \left\{ L(\theta) - \kappa (a'_N \Sigma_N^{-1} a_N - \delta) \right\}, \hspace{1cm} (2.62)$$

where $L(\theta)$ is defined in (2.61), and $\hat{\theta}_{\text{MLC}} = (\hat{a}_{N,\text{MCL}}', \hat{\lambda}_{\text{miss},\text{MCL}}', \text{vec}(\hat{A}_{N,\text{MCL}})', \text{vec}(\hat{B}_{N,\text{MCL}})', \text{vec}(\hat{C}_{N,\text{MCL}})')', \hat{\lambda}_{\text{MCL}}'$, $\text{vech}(\hat{f}_{\text{MCL}})$).

(i) For $\kappa > 0$, if the APT restriction holds exactly, that is, $a'_N \Sigma_N^{-1} a_N = \delta$, and $\Sigma_{ff} - \tilde{f} \tilde{f}'$ is nonsingular, then

$$\text{vec}(\hat{B}_{N,\text{MCL}}) = \left( \Sigma_{ff} \otimes I_N \right)^{-1} \left( \tilde{f} \tilde{f}' \otimes (2G_N - G_N G_N) \right)^{-1} \text{vec} \left( \Theta_{ff} - (2G_N - G_N G_N) \tilde{f} \tilde{f}' \right), \hspace{1cm} (2.63)$$

$$\hat{a}_{N,\text{MCL}} = \frac{1}{\kappa + 1} \left( \tilde{f} - \hat{B}_{N,\text{MCL}} \hat{f} - \hat{A}_{N,\text{MCL}} \hat{\lambda}_{\text{miss},\text{MCL}} \right), \hspace{1cm} (2.64)$$

\textsuperscript{39}Note that because we are interpreting the missing factors as unobserved, without loss of generality, one can assume that $A_N \tilde{A}_N$ represents the contribution of the missing factors to the residual variance $\Sigma_N$ because the missing factors are assumed to be uncorrelated and have unit variance, leaving the risk premia $\lambda_{\text{miss},\text{MCL}}$ as free parameters to be estimated. MacKinlay and Pästor (2003) consider a different identification assumption. For $p = 1$, they estimate $\alpha_N$ without distinguishing between $A_N$ and $\lambda_{\text{miss},\text{MCL}}$, implying that the contribution of the single missing factor to the return variance equals $\alpha_N a'_N / (\text{SR}^R)^2$, where $\text{SR}^R$ is the Sharpe ratio of the missing factor.

\textsuperscript{40}We are implicitly assuming that observed and missing factors are mutually uncorrelated.
in which $\Sigma_{N, MLC} = \hat{\mathbf{A}}_{N, MLC} \hat{\mathbf{A}}'_{N, MLC} + \hat{\mathbf{C}}_{N, MLC}$, $\Sigma_{ff} = \frac{1}{T} \sum_{t=1}^{T} \tilde{f}_t \tilde{f}_t'$, $\Sigma_{f} = \frac{1}{T} \sum_{t=1}^{T} \tilde{f}_t r_t$, $\bar{r}_t = (r_t - r_f 1_N)$, $\bar{r}_f = (f_t - r_f 1_K)$, and

$$G_N = \frac{1}{\kappa + 1} \mathbf{1}_N + \frac{\kappa}{(\kappa + 1)} \hat{\mathbf{A}}_{N, MLC} \hat{\mathbf{A}}'_{N, MLC} \hat{\Sigma}_{N, MLC}^{-1} \hat{\mathbf{A}}'_{N, MLC} \hat{\Sigma}_{N, MLC}^{-1}.$$

Note that $\hat{\mathbf{A}}_{N, MLC}$ and $\hat{\mathbf{C}}_{N, MLC}$ do not admit a closed-form solution and, as before, $\hat{\lambda}_{MLC}$ and $\hat{\Omega}_{MLC}$ coincide with the sample mean and sample covariance of the factors $f_t$.

(ii) For $\kappa = 0$ and the APT restriction not holding exactly, one can identify only $\alpha_0 = \mathbf{A}_0 \lambda_{miss0} + \mathbf{a}_0$ but not the two components separately, and one obtains

$$\hat{\alpha}_{N, MLC} = \bar{r}_N - \hat{\mathbf{B}}_{N, MLC} \bar{f}, \quad (2.66)$$

and the expression for $\text{vec} (\hat{\mathbf{B}}_{N, MLC})$ can be obtained by setting $\kappa = 0$ in the terms that appear in (2.63). The expressions for $\hat{\lambda}_{MLC}$ and $\hat{\Omega}_{MLC}$ are unchanged, and, as before, the expressions for the estimators of $\hat{\mathbf{A}}_{N, MLC}$ and $\hat{\mathbf{C}}_{N, MLC}$ do not admit a closed-form solution.

2.5. Numerical Illustration

In this section, our goal is to illustrate the improvement in out-of-sample portfolio performance that results from our new theoretical insights and to identify the circumstances when these improvements will be large and when they will be small; our objective is not to run a horse race across the large number of portfolio strategies studied in the literature—many of which could exploit the insights from our theoretical analysis. In particular, all portfolio strategies that rely on estimating expected returns could benefit from distinguishing between the alpha and beta component of returns, using asymptotic analysis to mitigate misspecification in the beta component of returns, and imposing the restriction arising from the extended APT to mitigate misspecification in the alpha component of returns.43

This section is divided into three parts. In the first part, we explain the design of our simulation experiment. In the second part, we demonstrate the improvement in portfolio performance resulting from treating misspecification in the beta component of returns, which is done using the insights from the asymptotic analysis of portfolio weights, as reported in Section 2.4.2. In the third part, we demonstrate the gains from treating misspecification in the alpha component of returns, which is done by imposing the APT restriction, as explained in Section 2.4.3.

2.5.1 Simulation design and performance evaluation

The design of our simulation analysis is similar to that in MacKinlay and Pástor (2000). We consider the case in which the number of assets is $N = 100$ and the investor assumes, and therefore estimates, a single-factor model:

$$r_t = \alpha_N + \beta_N f_t + \varepsilon_t. \quad (2.67)$$

Throughout the exercise, we assume that the risk-free interest rate is 0 and that the observed factor $f_t$ is IID and has Gaussian distribution. For the “base case” of our simulation exercise, we assume that the observed factor has a monthly mean equal to $\lambda = \frac{5}{12 \times 100}$ and monthly variance equal to $\Omega = \left(\frac{15}{\sqrt{12 \times 100}}\right)^2$; both $\lambda$ and $\Omega$ are scalars in our numerical illustration, because the investor assumes that there is only a single factor.

We consider two environments, one with mispricing leading to bounded residual variation ($\alpha_N = \alpha_N$) and the other with a single missing factor leading to unbounded residual variation ($\alpha_N = \lambda_N \lambda_{miss}$ with

---

43The online appendix shows how our theoretical insights can be applied to mitigate misspecification also in the minimum-variance portfolio studied in, for example, Jagannathan and Ma (2003); Ledoit and Wolf (2004), and DeMiguel et al. (2009a). By using the insights in Sections 2.5.2 and 2.5.3 for mitigating misspecification in expected returns, one can improve further the performance of these strategies that ignore expected returns and focus on mitigating misspecification in just the covariance matrix of returns.
Both $a_N$ and $A_N$ are generated from an IID multivariate Gaussian distribution with mean $0_N$ and covariance matrix equal to $\sigma^2_N I_N$, with $\sigma_N = \frac{1}{\sqrt{12} \times 100}$ which, in order to be conservative and consistent with the empirical data for individual stocks, is one quarter of the value used in MacKinlay and Pástor (2000). In both cases, $\varepsilon_t$ is IID with a multivariate Gaussian distribution with a monthly mean of $0_N$. In the bounded-variation case, the monthly covariance matrix of $\varepsilon_t$ is $\Sigma_N = \sigma^2_N I_N$; in contrast, in the unbounded-variation case, the monthly covariance matrix is $\Sigma_N = \Lambda_N \Lambda_N' + \sigma^2_N I_N$. In both cases, $\sigma = \frac{20}{\sqrt{12} \times 100}$. Observe that to ensure identification, just as in MacKinlay and Pástor (2000), we set the variance of the missing factor equal to one, which implies that the risk premium $\lambda_{\text{miss}}$ coincides with the Sharpe ratio for the missing factor; we set $\lambda_{\text{miss}} = \frac{0.72}{\sqrt{12}}$.

We measure the performance of a particular portfolio strategy using its out-of-sample Sharpe ratio, which is computed as follows. First, using the above parameter values, we simulate $M = 100$ Monte Carlo paths of length $T = 300$ months. For each path, we estimate the moments of asset returns using a rolling window of 120 monthly observations, based on which we construct the portfolio weight for each of the portfolio strategies described below. We then compute the return of the portfolio strategy based on the realized return in the 121st month following the estimation window. We repeat this using the series of realized returns for the next 179 months, we construct the Sharpe ratio for this particular path of the Monte Carlo simulation. We repeat this for each Monte Carlo path and report the average Sharpe ratio across all the paths.

### 2.5.2 Model misspecification in the beta component of returns

In this section, we study the effect on portfolio performance of model misspecification in the beta component of returns, in particular, misspecification of the risk premium, $\lambda$, which is given by the mean of the factor in excess of the risk-free rate.

In this analysis, we use the data-generating model described above, in which over each 120-month period the sample excess mean return on the single factor is $\bar{\lambda} = \frac{1}{120} \sum_{t=1}^{120} f_t$. In our experiment, we assume that the investor has the incorrect view that $\lambda$ is either half or double of its sample estimator, $\hat{\lambda}$. Then, we compare two versions of the MLC strategy, which is the ML-based constrained portfolio based on the sample mean and factor covariance matrix of $\Omega^T \beta_N \beta_N' + \Sigma_N$ with the APT constraint in which $\delta$ is obtained using ten-fold cross validation. The first version of this strategy, MLC$_{\text{views}}$, is based on the investor's (incorrect) view of the factor risk premia, $\lambda$. The second version of the strategy, MLC$_{\text{robust}}$, is based on our asymptotic analysis in Theorem 2.9, which is robust to misspecification of $\lambda$. Recall that in Theorem 2.9 we had shown that, under certain conditions, one could replace the beta portfolio by a benchmark portfolio without loss of performance; here, the benchmark portfolio that we use is the equally weighted portfolio, which always satisfies the conditions for the theorem in a one-factor environment. Comparing these two strategies, both of which impose the APT constraint, allows us to understand the benefits arising from mitigating misspecification arising from the beta component of returns. The results of this comparison are given in Table 2.1.

In Panel A of Table 2.1, we consider the case in which the pricing error has bounded variation; that is, $\alpha_N = a_N$. The first row of this table gives the base case, in which there are $N = 100$ risky assets and the investor has the correct view about $\lambda$, which implies that in this case MLC$_{\text{views}} = $ MLC$_{\text{robust}}$, which for our simulated data is 1.96 p.a. In contrast, the Sharpe ratio of the equally weighted (EW) portfolio, which is not optimizing, is only 0.48 p.a. and the Sharpe ratio of the mean-variance (MV) portfolio is only 0.02 p.a. because this strategy relies on estimates of both the sample covariance matrix and the sample mean, and it is well-known that this strategy performs poorly, especially for large $N$. The global-minimum-variance portfolio based on the sample-covariance matrix, which ignores expected returns entirely, has a Sharpe ratio

---

42 The multi-step procedure outlined above in Section 2.4.3, based on principal-component analysis, will identify the presence of the missing factor.

43 As we show below, using a larger value of $\sigma$ would strengthen our results.

44 Note that the MLC$_{\text{robust}}$ portfolio is using the equally weighted portfolio only to approximate its beta component, while for the EW strategy the equally weighting is for the entire portfolio.
Table 2.1  Out-of-Sample Sharpe Ratios with Beta Misspecification

This table reports the annualized Sharpe ratios averaged across $M = 100$ Monte Carlo simulations for the following strategies: MV, the mean-variance portfolio based on the sample mean and sample covariance matrix; EW, the equal-weighted portfolio; MLCviews, the ML-based constrained mean-variance portfolio based on the (incorrect) views of the investor regarding $\lambda$; MLCrobust, the ML-based constrained mean-variance portfolio that is robust to model misspecification based on the theoretical results from our asymptotic analysis. The $t$-statistic in the penultimate column is for the difference in the Sharpe ratio of the MLCrobust portfolio and the equally weighted (EW) portfolio, and the $t$-statistic in the last column is for the difference in the Sharpe ratio of the MLCrobust portfolio and the MLCviews portfolio.

<table>
<thead>
<tr>
<th></th>
<th>MV</th>
<th>EW</th>
<th>GMV</th>
<th>MLCviews</th>
<th>MLCrobust</th>
<th>t-stat wrt EW</th>
<th>t-stat wrt MLCviews</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Pricing errors unrelated to factors ($\alpha_N = \alpha_N^*$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base case ($\lambda = \hat{\lambda}, N = 100$)</td>
<td>0.02</td>
<td>0.48</td>
<td>0.14</td>
<td>1.96</td>
<td>1.96</td>
<td>15.74</td>
<td>—</td>
</tr>
<tr>
<td>Low $\lambda$ (half of base case, $N = 100$)</td>
<td>0.02</td>
<td>0.48</td>
<td>0.14</td>
<td>0.99</td>
<td>1.96</td>
<td>16.47</td>
<td>8.60</td>
</tr>
<tr>
<td>High $\lambda$ (double of base case, $N = 100$)</td>
<td>0.02</td>
<td>0.48</td>
<td>0.14</td>
<td>1.72</td>
<td>1.96</td>
<td>16.47</td>
<td>3.28</td>
</tr>
<tr>
<td>Low $\lambda$ (half of base case, $N = 5$)</td>
<td>0.45</td>
<td>0.42</td>
<td>0.40</td>
<td>0.29</td>
<td>0.42</td>
<td>0.01</td>
<td>2.53</td>
</tr>
<tr>
<td>High $\lambda$ (double of base case, $N = 5$)</td>
<td>0.45</td>
<td>0.42</td>
<td>0.40</td>
<td>0.30</td>
<td>0.42</td>
<td>0.01</td>
<td>2.32</td>
</tr>
<tr>
<td><strong>Panel B: Pricing errors related to factors ($\alpha_N = A_N \lambda_{miss}$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base case ($\lambda = \hat{\lambda}, N = 100$)</td>
<td>0.00</td>
<td>0.48</td>
<td>0.06</td>
<td>0.54</td>
<td>0.54</td>
<td>2.75</td>
<td>—</td>
</tr>
<tr>
<td>Low $\lambda$ (half of base case, $N = 100$)</td>
<td>0.00</td>
<td>0.48</td>
<td>0.06</td>
<td>0.50</td>
<td>0.54</td>
<td>2.57</td>
<td>2.53</td>
</tr>
<tr>
<td>High $\lambda$ (double of base case, $N = 100$)</td>
<td>0.00</td>
<td>0.48</td>
<td>0.06</td>
<td>0.51</td>
<td>0.54</td>
<td>2.57</td>
<td>2.23</td>
</tr>
<tr>
<td>Low $\lambda$ (half of base case, $N = 5$)</td>
<td>0.26</td>
<td>0.42</td>
<td>0.36</td>
<td>0.41</td>
<td>0.33</td>
<td>-2.48</td>
<td>-2.35</td>
</tr>
<tr>
<td>High $\lambda$ (double of base case, $N = 5$)</td>
<td>0.26</td>
<td>0.42</td>
<td>0.36</td>
<td>0.47</td>
<td>0.33</td>
<td>-2.48</td>
<td>-4.69</td>
</tr>
</tbody>
</table>

of 0.14 p.a., which is better than that of the MV portfolio, but substantially smaller than that of the EW and MLC strategies.

Now we study the effect of having an incorrect view about the market risk premium, $\lambda$. We see that, for the case of $N = 100$ reported in the second and third rows of Panel A of Table 2.1, the Sharpe ratio from using the strategy based on the incorrect views regarding $\lambda$, MLCviews, is significantly smaller than that from MLCrobust, which is immune to misspecification in the beta component of returns. For example, if the incorrect view is half of the MLE estimate of $\lambda$, the Sharpe ratio of MLCviews is 0.99 p.a. instead of the 1.96 p.a. achieved by the MLCrobust strategy. Moreover, we see that to achieve these gains based on our “large $N$” analysis, it is sufficient to have $N = 100$ risky assets.

To identify the circumstances when the improvement is small from using the strategy that mitigates misspecification in the beta component of return, consider the case where there are only $N = 5$ risky assets. In this case, the asymptotic results of Theorem 2.9 do not apply, and we see from the last two rows of Panel A of Table 2.1 that the strategy that accounts for beta misspecification achieves a Sharpe ratio of only 0.42 p.a., which is the same as that for the equally weighted (EW) portfolio, and smaller than the Sharpe ratio of 0.45 p.a. for the MV portfolio, whose performance improves as $N$ decreases.

The insights that emerge from analyzing Panel B of Table 2.1, which considers the case where the pricing error has unbounded variation, $\alpha_N = A_N \lambda_{miss}$ with $a_N = 0_N$, are similar to those from Panel A. The only difference is that the gains from using the MLCrobust strategy relative to MLCviews are smaller in Panel B than in Panel A. The reason for this is that, as explained in Theorem 2.8 and Corollary 2.8.1, for the case with $a_N = 0_N$ the weights $w_N^\lambda$ do not dominate $w_N^\lambda$, instead, they behave just like the $w_N^\lambda$ weights, so there are no benefits from our asymptotic results. In practice, of course, one would encounter the presence of both $a_N$ and $A_N$. 
2.5. NUMERICAL ILLUSTRATION

Table 2.2 Out-of-Sample Sharpe Ratios with Alpha Misspecification

This table reports the annualized Sharpe ratios averaged across $M = 100$ Monte Carlo simulations for the following strategies: MV, the mean-variance portfolio based on the sample mean and sample covariance matrix; EW, the equal-weighted portfolio; MLU, the ML-based unconstrained mean-variance portfolio based on the sample mean and covariance matrix implied by the factor model, $\beta_N\beta_N'\Omega + \sigma_N^2I_N$ but without the APT constraint; and MLC, the ML-based constrained mean-variance portfolio based on the sample mean and factor covariance matrix of $\beta_N\beta_N'\Omega + \Sigma_N$ with the APT constraint in which $\delta$ is obtained using ten-fold cross validation. The $t$-statistic in the penultimate column is for the difference in the Sharpe ratio of the MLC portfolio and the equally weighted (EW) portfolio, and the $t$-statistic in the last column is for the difference in the Sharpe ratio of the MLC portfolio and the MLU portfolio.

<table>
<thead>
<tr>
<th></th>
<th>MV</th>
<th>EW</th>
<th>GMV</th>
<th>MLU</th>
<th>MLC</th>
<th>$t$-stat</th>
<th>$t$-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>wrt EW</td>
<td>wrt MLU</td>
</tr>
<tr>
<td><strong>Panel A: Pricing errors unrelated to factors ($\alpha_N = a_N$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base case</td>
<td>0.02</td>
<td>0.48</td>
<td>0.14</td>
<td>1.41</td>
<td>1.96</td>
<td>15.74</td>
<td>7.77</td>
</tr>
<tr>
<td>Low $\lambda$ (half of base case)</td>
<td>-0.01</td>
<td>0.23</td>
<td>0.11</td>
<td>0.84</td>
<td>1.17</td>
<td>9.81</td>
<td>5.18</td>
</tr>
<tr>
<td>Low $\sigma_\epsilon$ (half of base case)</td>
<td>0.03</td>
<td>0.48</td>
<td>0.21</td>
<td>4.04</td>
<td>4.85</td>
<td>25.34</td>
<td>2.27</td>
</tr>
<tr>
<td>Low $\sigma_\alpha$ (half of base case)</td>
<td>-0.01</td>
<td>0.48</td>
<td>0.09</td>
<td>0.44</td>
<td>0.60</td>
<td>3.17</td>
<td>4.18</td>
</tr>
<tr>
<td><strong>Panel B: Pricing errors related to factors ($\alpha_N = A_N \lambda_{mix}$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base case</td>
<td>0.00</td>
<td>0.48</td>
<td>0.06</td>
<td>0.19</td>
<td>0.57</td>
<td>2.75</td>
<td>7.31</td>
</tr>
<tr>
<td>Low $\lambda$ (half of base case)</td>
<td>-0.03</td>
<td>0.23</td>
<td>0.02</td>
<td>0.17</td>
<td>0.28</td>
<td>1.38</td>
<td>1.88</td>
</tr>
<tr>
<td>Low $\sigma_\epsilon$ (half of base case)</td>
<td>-0.03</td>
<td>0.48</td>
<td>0.02</td>
<td>0.11</td>
<td>0.60</td>
<td>3.54</td>
<td>8.36</td>
</tr>
<tr>
<td>Low $\sigma_\alpha$ (half of base case)</td>
<td>0.01</td>
<td>0.48</td>
<td>0.06</td>
<td>0.26</td>
<td>0.56</td>
<td>2.40</td>
<td>6.27</td>
</tr>
</tbody>
</table>

2.5.3 Model misspecification in the alpha component of returns

In this section, we study the effect from mitigating model misspecification in only the alpha component of returns. To do this, we compare the Sharpe ratio of two strategies, one with the APT constraint (MLC) and the other unconstrained (MLU). In order to isolate the effect of the APT constraint, we do not use the asymptotic results of Theorem 2.9 to mitigate misspecification arising from the beta component of returns.

The MLC portfolio imposes the APT constraint, relies on the sample mean, and uses the covariance matrix $\Omega \beta_N \beta_N' + \Sigma_N$. The MLU portfolio, on the other hand, does not impose the APT constraint and relies on the sample mean but uses the covariance matrix $\Omega \beta_N \beta_N' + \sigma_N^2 I_N$. Note that in Panel A of Table 2.2, in which we consider the case in which the pricing error is unrelated to the factors $\alpha_N = a_N$, $\Sigma_N = \sigma_N^2 I_N$, so the only difference between MLU and MLC is that the latter includes the APT constraint. In Panel B, in which we consider the case in which the pricing error is related to factors $\alpha_N = A_N \lambda_{mix}$, $\Sigma_N$ differs from $\sigma_N^2 I_N$, implying that MLC is misspecified. However, MacKinlay and Pastor (2000) show that using the (misspecified) $\sigma_N^2 I_N$ instead of $\Sigma_N = A_N A_N' + \sigma_N^2 I_N$ *improves* performance out of sample. Thus, our comparison of MLU with MLC is more conservative: that is, if one were to use $\Sigma_N = A_N A_N' + \sigma_N^2 I_N$ for MLU, its difference from MLU would be even larger.

We start by looking at the row for the "base case" in Panel A of Table 2.2. We see that the MV portfolio perform poorly and, as has been shown in the literature, the EW portfolio achieves a higher Sharpe ratio because it does not suffer from estimation error. Examining the MLU portfolio, we see that it performs much better than the EW portfolio: its Sharpe ratio is higher than that of the EW portfolio—1.41 p.a. instead of 0.48 p.a. This is because the equally weighted portfolio fails to take advantage of the dispersion in alphas and earns only the average alpha, which is zero, while the MLU portfolio can fully exploit the presence of pricing errors. The MLC portfolio, which imposes the APT constraint, performs even better: it has a Sharpe ratio of 1.96 p.a., which is significantly higher than even that of the MLU strategy. The superior performance of the MLC portfolio relative to the MLU portfolio highlights the importance of imposing the APT restriction.
In addition to the "base case" of the simulations described above, we look at three variations in Table 2.2. The second row of Panel A considers the case in which the risk premium on the observed factor, \( \lambda \), is half the base-case value, which corresponds to a low-return environment. Not surprisingly, in this case the Sharpe ratios of all the portfolios decrease, but that of the MLC strategy decreases the least, and continues to be significantly higher than that of MLU as well as the other strategies. The third row of Panel A considers the case in which the residual risk is half of its base-case value. In this case, the Sharpe ratio of the EW portfolio does not change at all, but the Sharpe ratios of the other strategies, which rely on estimated return moments, improve substantially. One might wonder under what conditions the EW strategy will outperform the MLC strategy: this happens when the alphas are small in absolute value. In the last row of Panel A, we consider the case in which the dispersion of alphas is only half the base-case value. In this case, the Sharpe ratio of the EW strategy is still the same as its base-case value, but now the Sharpe ratio of MLC is significantly smaller, though still significantly greater than that of the EW and MLU portfolios.

Panel B of Table 2.2 shows that the insights described above are similar for the case in which the pricing errors are related to factors. Note that the gains from using the MLC strategy are smaller in Panel B than in Panel A. This is partly because, as explained in Section 2.4.3, for the case in which the pricing error has only unbounded variation, \( \lambda_N \) appears in both the mean and covariance matrix of returns, and hence, imposing the APT restriction does not have any benefit because this restriction is satisfied automatically.

2.6. Conclusion

In this paper, we have provided a rigorous foundation and characterization, based on the APT, for alpha and beta portfolios, where the "alpha" portfolio is one that depends on pricing errors and the "beta" portfolio depends on factor risk premia. We then show how these properties can be exploited to mitigate the effects of model misspecification for portfolio choice.

Our first result is to explain that one can extend the interpretation of the APT so that the alpha in it represents not just small pricing errors that are independent of factors but also large pricing errors arising from mismeasured or missing factors. We also show how the APT model can capture misspecification in the beta component of returns. We then use the mathematical structure underlying the APT to study the mitigation of model misspecification for the mean-variance portfolio.

Our key insight is that instead of treating model misspecification directly in the mean-variance portfolios, it is better to first decompose mean-variance portfolios into two components, a "beta" portfolio and an "alpha" portfolio, and then to treat misspecification in these two components using different methods. Misspecification in the beta component of returns is treated by utilizing the property that the weights of the alpha portfolio dominate the corresponding weights in the beta portfolio as the number of assets increases asymptotically. Misspecification in the alpha component of returns is treated by imposing the APT restriction on the weighted sum of squares of the pricing errors when estimating the return-generating model. We use simulations to demonstrate that these theoretical findings can be exploited to achieve an improvement in out-of-sample portfolio performance that is both economically and statistically significant.
2.A. Proofs for Theorems

Note that Theorems 2.1 and 2.3 are derived in both Huberman (1982) and Ingersoll (1984); therefore, we do not include the proofs for these theorems. The proofs for all other theorems in the main text of the manuscript are given below, starting with some preliminary lemmas.

2.A.1 Lemma on decomposition of the Sharpe ratio

Lemma 2.1 Consider the portfolio weights \( w_N = w_{N,1} + w_{N,2} \) such that \( w_{N,1} \) is orthogonal to \( w_{N,2} \):
\[
w_{N,1} V_N w_{N,2} = 0.
\]

Then, defining \( \text{SR}_1 = \frac{w'_N (\mu_N - r_f 1_N)}{(w'_N V_N w_{N,1})^{1/2}} \) and letting \( \text{SR} \) denote the Sharpe ratio of the portfolio \( w_N \), we always have
\[
\text{SR}^2 = \left( \frac{w'_N (\mu_N - r_f 1_N)}{w'_N V_N w_{N,1}} \right)^2 \leq (\text{SR}_1)^2 + (\text{SR}_2)^2.
\]

Finally, equality holds if and only if:
\[
\frac{w'_{N,1} (\mu_N - r_f 1_N)}{w'_{N,1} V_N w_{N,1}} = w'_{N,2} (\mu_N - r_f 1_N) / w'_{N,2} V_N w_{N,2}.
\]

Proof. Defining for simplicity
\[
\mu_i - r_f = w'_{N,i} (\mu_N - r_f 1_N) \quad \text{and} \quad \sigma_i^2 = w'_{N,i} V_N w_{N,i},
\]
we have
\[
\text{SR}^2 = \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right)^2 \frac{\sigma_1^2}{w'_N V_N w_{N,1}} + \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right)^2 \frac{\sigma_2^2}{w'_N V_N w_{N,2}} + 2 \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right) \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right) \frac{\sigma_1^2 \sigma_2^2}{w'_N V_N w_{N,1} w'_N V_N w_{N,2}}
\]
(2.A.1)
\[
= \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right)^2 + \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right)^2 + \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right) \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right) \left( \frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2} \right)
\]
(2.A.2)
\[
= \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right)^2 + \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right)^2 + \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right) \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right) \left( \frac{\sigma_1^2}{\sigma_2^2} \right).
\]
(2.A.3)

Using the orthogonality of \( w_{N,1} \) and \( w_{N,2} \), we have \( w'_N V_N w_{N,1} = w'_{N,1} V_N w_{N,1} + w'_{N,2} V_N w_{N,2} = \sigma_1^2 + \sigma_2^2 \), so that the term in square brackets can be rewritten as
\[
- \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right)^2 \frac{\sigma_2^2}{w'_N V_N w_{N,1}} - \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right)^2 \frac{\sigma_1^2}{w'_N V_N w_{N,2}} + 2 \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right) \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right) \frac{\sigma_1^2 \sigma_2^2}{w'_N V_N w_{N,1} w'_N V_N w_{N,2}}
\]
(2.A.4)
\[
= \frac{1}{w'_N V_N w_{N,1}} \left( - \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right)^2 \frac{\sigma_1^2}{\sigma_2^2} - \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right)^2 \frac{\sigma_1^2}{\sigma_2^2} + 2 \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right) \frac{\sigma_1^2}{\sigma_2^2} \left( \frac{\mu_1 - r_f}{\sigma_2^2} \right) \frac{\sigma_1^2}{\sigma_2^2} \right)
\]
(2.A.5)
\[
= - \frac{1}{w'_N V_N w_{N,1}} \left( \frac{\mu_1 - r_f}{\sigma_1^2} - \mu_2 - r_f \frac{\sigma_1^2}{\sigma_2^2} \right)^2.
\]
(2.A.6)

Hence,
\[
\text{SR}^2 = \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right)^2 + \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right)^2 - \frac{1}{w'_N V_N w_{N,1} w'_N V_N w_{N,2}} \left( \left( \frac{\mu_1 - r_f}{\sigma_1^2} - \mu_2 - r_f \frac{\sigma_1^2}{\sigma_2^2} \right)^2 \right)
\]
(2.A.7)
\[
\leq \left( \frac{\mu_1 - r_f}{\sigma_1^2} \right)^2 + \left( \frac{\mu_2 - r_f}{\sigma_2^2} \right)^2 = (\text{SR}_1)^2 + (\text{SR}_2)^2.
\]
(2.A.8)

Equality holds if and only if
\[
\left( \frac{\mu_1 - r_f}{\sigma_1^2} - \mu_2 - r_f \frac{\sigma_1^2}{\sigma_2^2} \right)^2 = 0,
\]
which, in turn, can be rearranged as
\[
\frac{(\mu_1 - r_f)}{\sigma_1^2} = \frac{(\mu_2 - r_f)}{\sigma_2^2}.
\]
2.A.2 Extension of Roll (1980)

Roll (1980) shows that in the absence of a risk-free rate, for any inefficient portfolio one can identify the subspace of portfolios that are orthogonal to this portfolio with minimum variance. That is, corresponding to any inefficient portfolio, the number of zero-beta portfolios is infinite—one for each level of target mean. If the portfolio is efficient, then the subspace shrinks to a single point; that is, there is a unique zero-beta portfolio. In order to interpret our findings, we extend the result in Roll (1980) to the case in which investors can invest also in a risk-free asset.

**Lemma 2.2 (Extension of Roll (1980) to the case with a risk-free asset)** Let \( w_N^* \) be any, possibly inefficient, portfolio. Let \( w_N^e \) be the portfolio that satisfies

\[
\min \frac{1}{2} (w_N^e)'V_N w_N^e \quad \text{s.t.} \quad (w_N^e)'V_N w_N^e = 0
\]  

(2.A.9)

and

\[
\mu_N^i w_N^e + (1 - 1_N^i w_N^e) \gamma_j = \mu^*,
\]  

(2.A.10)

for a given target mean \( \mu^* \). Then,

\[
w_N^e = \begin{pmatrix} w_N^e, V_N^{-1}(\mu_N - \gamma_j 1_N) \end{pmatrix} \begin{pmatrix} (\sigma^2)^2 & \mu^* - \gamma_j \\ \mu^* - \gamma_j & (\text{SR}^{mv})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^* - \gamma_j \end{pmatrix},
\]  

(2.A.11)

where \( \begin{pmatrix} w_N^e, V_N^{-1}(\mu_N - \gamma_j 1_N) \end{pmatrix} \) is the \( N \times 2 \) matrix obtained by joining the \( N \times 1 \) vector of portfolio weights \( w_N^e \) with the \( N \times 1 \) vector \( V_N^{-1}(\mu_N - \gamma_j 1_N) \).

**Proof.** We adapt Roll’s (1980) proof of the main theorem. The Lagrangian for our problem is

\[
L(w_N^e, \lambda_1, \lambda_2) = (w_N^e)'V_N w_N^e - \lambda_1 ((w_N^e)'V_N w_N^e) - \lambda_2 (\mu_N^i w_N^e + (1 - 1_N^i w_N^e) \gamma_j - \mu^*),
\]  

with first-order conditions

\[
2V_N w_N^e = \begin{pmatrix} V_N w_N^e, (\mu_N - \gamma_j 1_N) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.
\]

Pre-multiplying both sides by \( 2^{-1} \begin{pmatrix} V_N w_N^e, (\mu_N - \gamma_j 1_N) \end{pmatrix} V_N^{-1} \) gives

\[
\begin{pmatrix} 0 \\ \mu^* - \gamma_j \end{pmatrix} = 2^{-1} \begin{pmatrix} (\sigma^2)^2 & \mu^* - \gamma_j \\ \mu^* - \gamma_j & (\text{SR}^{mv})^2 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.
\]

Substituting out for \( \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \) concludes the proof.

**Remark 2.14** When \( w_N^e \) is efficient, then \( w_N^e = 0 \), which implies that the zero-beta portfolio to \( w_N^e \) is the portfolio that invests 100% in the risk-free asset. In fact substituting \( w_N^e = \gamma^{-1} V_N^{-1}(\mu_N - \gamma_j 1_N) \) into the first constraint gives \( 0 = (w_N^e)'V_N w_N^e = \gamma^{-1}(\mu_N - \gamma_j 1_N)'w_N^e = \gamma^{-1}(\mu_N - \gamma_j) \), where the last equality is due to the second constraint. Therefore one obtains \( \mu_e = \gamma_j \) which, by no-arbitrage, implies \( w_N^e = 0 \).

Recall the well-known result that the entire efficient frontier of risky assets can be generated from holding any two efficient portfolios. However, one can show that the efficient frontier of risky assets can also be generated by holding two inefficient portfolios, as long as one is the minimum-variance orthogonal portfolio of the other. This leads to the following result:

**Corollary 2.2.1 (Extension of Corollary 3 of Roll (1980) to the case with a risk-free asset)** There is a weighted average of, possibly inefficient, portfolio \( w_N^e \) with a corresponding minimum-variance orthogonal portfolio \( w_N^e \) that produces an efficient portfolio.

**Remark 2.15** The above theorem implies that the subspace of minimum-variance portfolios orthogonal to \( w^* \) is given by the two lines described by the expression below:

\[
\mu^* = \gamma_j \pm \sigma^* \sqrt{(\text{SR}^{mv})^2 - (\text{SR}^2)^2}.
\]  

(2.A.12)

Notice from the equation above and the dashed and dotted lines in Figure 2.3 that the slopes of the two lines are smaller (in absolute value) than the slopes of the capital market lines. For portfolios that are efficient, the subspace shrinks to a single point, which is the risk-free rate of return, as one can see from setting the Sharpe ratio of portfolio \( w_N^e \) equal to the Sharpe ratio of the mean-variance portfolio \( w_N^{mv} \) in the equation above.

Huang and Litzenberger (1988) show that, depending on the level of the risk-free rate relative to the mean of the global minimum-variance portfolio, the capital market line could be sloping up or down.
2.A.3 Equivalent representations for the portfolio \( w_N^\alpha \)

The portfolio \( w_N^\alpha \) in (2.33) has five equivalent representations, which we will use throughout the paper to gain insights for the portfolio weights when the number of assets is large. These five representations are given in the next lemma.

**Lemma 2.3 (Equivalent representations for \( w_N^\alpha \))**

\[
\begin{align*}
\quad w_N^\alpha &= \frac{1}{\gamma^\alpha} \Sigma_N^{-1} \alpha_N \\
&= \frac{1}{\gamma^\alpha} V_N^{-1} \alpha_N \\
&= \frac{1}{\gamma^\alpha} \tilde{V}_N \alpha_N \\
&= \frac{1}{\gamma^\alpha} \tilde{V}_N (\mu_N - r_f 1_N) \\
&= \frac{1}{\gamma^\alpha} \tilde{V}_N \alpha_N.
\end{align*}
\]

where

\[
\tilde{V}_N = \left[ \Sigma_N^{-1} - \Sigma_N^{-1} B_N (B_N' \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1} \right].
\]

**Proof.** The first equality is the definition of the \( w_N^\alpha \) portfolio in (2.33); the second equality follows from the orthogonality of the projection in (2.11); the third and fourth equalities follow from the definition of \( V_N \) in (2.6), the definition of \( \tilde{V}_N \) (2.18), and the fact that

\[
\tilde{V}_N B_N = 0.
\]

The fifth equality follows from \( \mu_N - r_f 1_N = \alpha_N + B_N \lambda \).

**Remark 2.1** It is useful to discuss the relation between \( V_N^{-1} \) and \( \tilde{V}_N \). Note that the Sherman-Morrison-Woodbury formula implies

\[
V_N^{-1} = \left[ \Sigma_N^{-1} - \Sigma_N^{-1} B_N (\Omega^{-1} + B' N \Sigma_N^{-1} B_N)^{-1} B' N \Sigma_N^{-1} \right].
\]

Setting \( \Omega^{-1} = 0 \) in the expression above then leads to the expression for \( \tilde{V}_N \) in (2.18); that is, when \( \Omega \) takes arbitrarily large value, then \( V_N^{-1} \) tends toward \( \tilde{V}_N \). To understand the intuition for this, observe that mean-variance optimization penalizes assets with large variances. As the elements of \( \Omega \) become arbitrarily large, all assets will have a large variance, and therefore, a low Sharpe ratio. The only way to avoid this outcome is to diversify the exposure of the portfolio to common factors; that is, the magnitude of the elements of \( V_N^{-1} B_N \) to be as small as possible. This is automatically achieved when \( \Omega^{-1} = 0 \) in the expression of \( V_N^{-1} \), which leads to \( \tilde{V}_N \), guaranteeing that (2.19) is satisfied.

2.A.4 APT restriction in terms of projection errors

In the lemma below, we show that the APT restriction can be expressed as a function of either the projection errors \( \hat{\alpha}_N \) or their (element by element) limit \( \alpha_N \).

**Lemma 2.4 (Equivalence of APT constraint in terms of \( \hat{\alpha}_N \) and \( \alpha_N \))**

\[
\hat{\alpha}_N \Sigma_N^{-1} \alpha_N \leq \delta < \infty \text{ implies } \alpha_N \Sigma_N^{-1} \alpha_N \leq \delta < \infty.
\]

**Proof.** Note that from (2.9) and (2.10) we have

\[
\begin{align*}
\hat{\alpha}_N &= (\mu_N - r_f 1_N) - B_N \hat{\lambda} \\
&= (I_N - B_N (B_N' \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1})(\alpha_N + B_N \lambda) \\
&= (I_N - B_N (B_N' \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1})\alpha_N.
\end{align*}
\]
Because
\[
\Sigma_N^{-1} \tilde{\alpha}_N = \Sigma_N^{-1} (I_N - B_N (B_N' \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1}) \alpha_N = \tilde{V}_N \alpha_N,
\]
it follows that
\[
\tilde{\alpha}_N' \Sigma_N^{-1} \tilde{\alpha}_N = \tilde{\alpha}_N' \Sigma_N^{-1} \Sigma_N \tilde{\alpha}_N
\] (2.25)
\[
= \tilde{\alpha}_N' \tilde{V}_N \Sigma_N \tilde{V}_N \alpha_N
\] (2.26)
\[
= \tilde{\alpha}_N' \tilde{V}_N \alpha_N,
\] (2.27)
where \( \tilde{V}_N \) is defined in (2.18). Therefore, the condition in (2.28) below,
\[
\tilde{\alpha}_N' \Sigma_N^{-1} \tilde{\alpha}_N \leq \delta < \infty,
\] (2.28)
implies from (2.27) that
\[
\tilde{\alpha}_N' \Sigma_N^{-1} \tilde{\alpha}_N = \alpha_N' \tilde{V}_N \alpha_N = \alpha_N' \Sigma_N^{-1} \alpha_N - \alpha_N' \Sigma_N^{-1} B_N (B_N' \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1} \alpha_N \leq \delta < \infty.
\] (2.29)

Remark 2.1 Because \( \alpha_N' \tilde{V}_N \alpha_N \geq 0 \), therefore (2.29) implies that
\[
0 \leq \alpha_N' \Sigma_N^{-1} B_N (B_N' \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1} \alpha_N \leq \alpha_N' \Sigma_N^{-1} \alpha_N \leq \delta < \infty,
\] (2.30)
implying that (2.12), as well as the equation below, hold by no-arbitrage:
\[
\alpha_N' \Sigma_N^{-1} B_N (B_N' \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1} \alpha_N \leq \delta < \infty.
\] (2.31)

\subsection{2.A.5 Proof of Theorem 2.4}

By Chamberlain and Rothschild (1983, Theorem 4) the residual covariance matrix satisfies
\[
\Sigma_N = A_N A_N' + C_N,
\]
where \( C_N \) is a positive definite matrix with eigenvalues uniformly bounded by \( g_{p+1} N(\Sigma_N) \). By the Sherman-Morrison-Woodbury decomposition,
\[
\Sigma_N^{-1} = C_N^{-1} - C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1}.
\]
Therefore, by substitution,
\[
\alpha_N' \Sigma_N^{-1} \alpha_N = \alpha_N' C_N^{-1} \alpha_N - \alpha_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} \alpha_N = \alpha_N' C_N^{-1} \alpha_N
\] (2.32)
\[
= (A_N \lambda_{\text{miss}} + a_N) C_N^{-1} (A_N \lambda_{\text{miss}} + a_N)
\] (2.33)
\[
- (A_N \lambda_{\text{miss}} + a_N)' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} (A_N \lambda_{\text{miss}} + a_N)
\] (2.34)
\[
= \lambda_{\text{miss}}' A_N C_N^{-1} A_N \lambda_{\text{miss}} - \lambda_{\text{miss}}' A_N C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N \lambda_{\text{miss}}
\] (2.35)
\[
+ a_N C_N^{-1} a_N - a_N C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N
\] (2.36)
\[
+ 2a_N C_N^{-1} A_N \lambda_{\text{miss}} - 2a_N C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N \lambda_{\text{miss}}.
\] (2.37)

We now show that \( \alpha_N' \Sigma_N^{-1} \alpha_N \) is bounded even as \( N \) diverges. We look each of the term on the right-hand side of the last equality sign, one by one. Thus,
\[
\lambda_{\text{miss}}' A_N C_N^{-1} A_N \lambda_{\text{miss}} - \lambda_{\text{miss}}' A_N C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N \lambda_{\text{miss}} = \lambda_{\text{miss}}' (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N \lambda_{\text{miss}}
\] (2.38)
\[
= \lambda_{\text{miss}}' (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N \lambda_{\text{miss}} \leq \lambda_{\text{miss}}' \lambda_{\text{miss}}.
\] (2.39)
\[
\lambda_{\text{miss}}' I_p + A_N' C_N^{-1} A_N \leq \lambda_{\text{miss}}' I_p + A_N' C_N^{-1} A_N \lambda_{\text{miss}} \leq \lambda_{\text{miss}}' \lambda_{\text{miss}}.
\] (2.40)
because $I_p - (I_p + A_N^{-1} C_N^{-1} A_N)^{-1} A_N^{-1} C_N^{-1} A_N$ is positive semidefinite. Next, for the third term,

$$a_N^{-1} C_N^{-1} a_N \leq a_N^{-1} a_N g_{iN}^{-1}(C_N).$$

Now, the $j$th element of $a_N^{-1} C_N^{-1} A_N$, obtained by considering the $j$th column of $A_N$, for every $1 \leq j \leq p$, satisfies

$$|a_N^{-1} C_N^{-1} a_N g_{jN} v_{jN}| \leq g_{jN}^{-1}(a_N^{-1} C_N^{-1} a_N) \leq g_{jN}^{-1/2} g_{iN}^{-1}(C_N)(a_N^{-1} a_N)^{1/2},$$

(recalling that $v_{jN} v_{jN} = 1$, where for simplicity we set $v_{jN} = v_{jN}(\Sigma_N), g_{jN} = g_{jN}(\Sigma_N)$. Moreover, the $(i, j)$th element for every $1 \leq i, j \leq p$, of $(A_N^{-1} C_N^{-1} A_N)$ is equal to $g_{iN} g_{jN} v_{iN} C_N^{-1} v_{jN}$. Therefore, assuming without loss of generality that $g_{iN} = \max\{g_{iN}, \ldots, g_{pN}\}$ for $N$ large enough, then $(I_p + A_N^{-1} C_N^{-1} A_N)^{-1}$ decreases at rate $g_{iN}^{-1}$. On the other hand, for the same reason, the elements of the vector $A_N^{-1} C_N^{-1} a_N$ diverge at most at rate $g_{iN}^{-1}$. Thus, the fourth term satisfies:

$$|a_N^{-1} C_N^{-1} A_N (I_p + A_N^{-1} C_N^{-1} A_N)^{-1} A_N^{-1} C_N^{-1} a_N| \leq \delta g_{iN}^{-1/2} g_{jN}^{-1} g_{iN}^{-1} = \delta. \quad (2.4.1)$$

Concerning the last two terms, it turns out that their difference converges to zero. In fact,

$$|2a_N^{-1} C_N^{-1} A_N \lambda - 2a_N^{-1} C_N^{-1} A_N (I_p + A_N^{-1} C_N^{-1} A_N)^{-1} A_N^{-1} C_N^{-1} A_N \lambda| \quad (2.4.2)$$

$$= 2a_N^{-1} C_N^{-1} A_N (I_p + A_N^{-1} C_N^{-1} A_N)^{-1} \lambda \quad (2.4.3)$$

$$\leq \delta g_{iN}^{-1} \lambda \quad (2.4.4)$$

$$\leq \delta g_{iN}^{-1} \lambda \quad (2.4.5)$$

### 2.4.6 Proof of Theorem 2.6

Given that $\mu_N - \gamma 1_N = \alpha_N + B_N \lambda$,

$$w_N^{\alpha} = \frac{1}{\gamma} V_N^{-1}(\mu_N - \gamma 1_N) \quad (2.4.6)$$

$$= \frac{1}{\gamma} V_N^{-1} \alpha_N + \frac{1}{\gamma} V_N^{-1} B_N \lambda \quad (2.4.7)$$

$$= \frac{1}{\gamma} V_N^{-1} \alpha_N + \frac{1}{\gamma} V_N^{-1} B_N \lambda \quad (2.4.8)$$

Thus,

$$w_N^{\alpha} = \phi^\alpha w_N^{\alpha} \quad (2.4.9)$$

Moreover,

$$w_N^{\alpha'} V_N w_N^{\beta} = \frac{1}{\gamma} w_N^{\alpha'} V_N V_N^{-1} B_N \lambda = \frac{1}{\gamma} w_N^{\alpha'} B_N \lambda = \frac{1}{\gamma \gamma} \alpha_N \tilde{V}_N B_N \lambda = 0, \quad (2.5.0)$$

because $\tilde{V}_N$ is orthogonal to $B_N$. Similarly,

$$w_N^{\alpha'} \Sigma_N w_N^{\beta} = \frac{1}{\gamma \gamma} w_N^{\alpha'} \Sigma_N V_N^{-1} B_N \lambda \quad (2.5.1)$$

$$= \frac{1}{\gamma \gamma} w_N^{\alpha'} B_N (I_K - (\Omega^{-1} + B_N \Sigma_N^{-1} B_N)^{-1} B_N \Sigma_N^{-1} B_N) \lambda \quad (2.5.2)$$

$$= \frac{1}{\gamma \gamma} \alpha_N \tilde{V}_N B_N (I_K - (\Omega^{-1} + B_N \Sigma_N^{-1} B_N)^{-1} B_N \Sigma_N^{-1} B_N) \lambda \quad (2.5.3)$$

$$= 0. \quad (2.5.4)$$

We now show that $w_N^{\alpha}$ and $w_N^{\beta}$ are the minimum-variance portfolios orthogonal to one another. This is accomplished by showing that these portfolio weights satisfy the result of Lemma 2.2. In particular, we need to verify that $w_N^{\alpha}$ satisfies

$$w_N^{\alpha} = (w_N^{\alpha}, V_N^{-1}(\mu_N - \gamma 1_N)) \left( \sigma^2^{\alpha} \begin{pmatrix} \mu^\alpha - \gamma f \\ \mu^\alpha - \gamma f \end{pmatrix} \right)^{-1} \begin{pmatrix} 0 \\ \sigma^2 \end{pmatrix}. \quad (2.5.5)$$
Simple calculations leads to
\[
\begin{pmatrix}
\frac{(\sigma^\alpha)^2}{\mu^\alpha - r_f} & \frac{\sigma^\alpha}{\mu^\alpha - r_f} (\text{SR}^\text{mv})^2 \\
\mu^\alpha - r_f & (\text{SR}^\text{mv})^2
\end{pmatrix}^{-1}
\begin{pmatrix}
0 \\
\mu^\beta - r_f
\end{pmatrix}
= \frac{1}{(\text{SR}^\text{mv})^2} \begin{pmatrix}
\frac{\delta_N^\alpha V_N \alpha_N}{\gamma^\alpha} - \frac{(\alpha_N^\beta V_N \alpha_N)^2}{(\gamma^\alpha)^2} \\
\frac{\delta_N^\alpha V_N \alpha_N}{\gamma^\alpha}
\end{pmatrix}
\begin{pmatrix}
\frac{\delta_N^\alpha V_N \alpha_N}{\gamma^\alpha} \\
\frac{\alpha_N^\beta V_N \alpha_N}{\gamma^\alpha}
\end{pmatrix}
\begin{pmatrix}
\frac{1}{\gamma^\beta} \\
1
\end{pmatrix}
\frac{\hat{\beta} B_N^\alpha V_N^{-1} B_N \hat{\lambda}}{\gamma^\beta}
\tag{2.56}
\end{align}
\]

Given that
\[
\begin{pmatrix}
\Sigma_N^{-1} \hat{\alpha}_N, V_N^{-1}(\hat{\alpha}_N + B_N \hat{\lambda}) \\
\frac{\delta_N^\alpha V_N \alpha_N}{\gamma^\alpha}
\end{pmatrix}
\begin{pmatrix}
\frac{\delta_N^\alpha V_N \alpha_N}{\gamma^\alpha} \\
\frac{\alpha_N^\beta V_N \alpha_N}{\gamma^\alpha}
\end{pmatrix}
= \frac{1}{(\gamma^\alpha)^2} V_N^{-1} B_N \hat{\lambda}
\tag{2.59}
\end{align}
\]

and
\[
\begin{pmatrix}
\delta_N^\alpha V_N \alpha_N \\
\delta_N^\alpha V_N \alpha_N
\end{pmatrix}
\begin{pmatrix}
\delta_N^\alpha V_N \alpha_N \\
\delta_N^\alpha V_N \alpha_N
\end{pmatrix}
= \frac{1}{(\gamma^\alpha)^2} \frac{\delta_N^\alpha V_N \alpha_N}{(\gamma^\alpha)^2} (\hat{\beta} B_N^\alpha V_N^{-1} B_N \hat{\lambda})
\tag{2.60}
\end{align}
\]

one finally obtains
\[
\begin{pmatrix}
w^\alpha_N, V_N^{-1}(\mu_N - r_f 1_N)
\end{pmatrix}
\begin{pmatrix}
(\sigma^\alpha)^2 & \mu^\alpha - r_f \\
\mu^\alpha - r_f & (\text{SR}^\text{mv})^2
\end{pmatrix}^{-1}
\begin{pmatrix}
0 \\
\mu^\beta - r_f
\end{pmatrix}
= \frac{1}{(\gamma^\alpha)^2} \frac{\delta_N^\alpha V_N \alpha_N}{(\gamma^\alpha)^2} V_N^{-1} B_N \hat{\lambda}
\tag{2.61}
\end{align}
\]

Along the same lines, it follows that \(w^\beta_N\) satisfies
\[
\begin{pmatrix}
w^\beta_N, V_N^{-1}(\mu_N - r_f 1_N)
\end{pmatrix}
\begin{pmatrix}
(\sigma^\beta)^2 & \mu^\beta - r_f \\
\mu^\beta - r_f & (\text{SR}^\text{mv})^2
\end{pmatrix}^{-1}
\begin{pmatrix}
0 \\
\mu^\alpha - r_f
\end{pmatrix}
\tag{2.62}
\end{align}
\]

2.7 Proof of Theorem 2.6

Because \(\tilde{V}_N V_N \tilde{V}_N = \tilde{V}_N \Sigma_N \tilde{V}_N = \tilde{V}_N\), it follows that
\[
\begin{align}
\mu^\alpha &= w^\alpha_N \mu_N + (1 - w^\alpha_N) r_f \\
&= \frac{1}{\gamma^\alpha} \frac{\delta_N^\alpha V_N \alpha_N}{\gamma^\alpha} + r_f \\
&= \frac{1}{\gamma^\alpha} \frac{\alpha_N^\beta V_N \alpha_N}{\gamma^\alpha} + r_f \\
&= \frac{1}{\gamma^\alpha} \frac{\delta_N^\alpha V_N \alpha_N}{\gamma^\alpha} + r_f \\
&= \frac{1}{(\gamma^\alpha)^2} \frac{\delta_N^\alpha V_N \alpha_N}{(\gamma^\alpha)^2} V_N^{-1} V_N \tilde{V}_N \tilde{V}_N \alpha_N = \frac{1}{(\gamma^\alpha)^2} \frac{\delta_N^\alpha V_N \alpha_N}{(\gamma^\alpha)^2} \tilde{V}_N \tilde{V}_N \alpha_N.
\end{align}
\]

Then use \(\frac{\delta_N^\alpha V_N \alpha_N}{\gamma^\alpha} = \frac{\delta_N^\alpha V_N \alpha_N}{\gamma^\alpha} \Sigma_N^{-1} \tilde{V}_N\).

Because \(B_N \Sigma_N^{-1} \tilde{V}_N = 0\), one gets
\[
\begin{align}
\mu^\beta &= w^\beta_N \mu_N + (1 - w^\alpha_N) r_f \\
&= \frac{1}{\gamma^\beta} \frac{\alpha_N^\beta V_N \alpha_N}{\gamma^\beta} + r_f, \\
&= \frac{1}{\gamma^\beta} \frac{\delta_N^\beta V_N \alpha_N}{\gamma^\beta} + r_f, \\
&= \frac{1}{(\gamma^\beta)^2} \frac{\alpha_N^\beta V_N \alpha_N}{(\gamma^\beta)^2} V_N^{-1} V_N \tilde{V}_N \tilde{V}_N \alpha_N = \frac{1}{(\gamma^\beta)^2} \frac{\alpha_N^\beta V_N \alpha_N}{(\gamma^\beta)^2} \tilde{V}_N \tilde{V}_N \alpha_N.
\end{align}
\]

(2.63)
2.A. PROOFS FOR THEOREMS

The boundedness of \( \tilde{\alpha}_N \Sigma_N^{-1} \tilde{\alpha}_N \) follows from Theorem 2.1 (APT). Next,

\[
B_N' \Sigma_N^{-1} B_N = B_N' (\Sigma_N^{-1} - \Sigma_N^{-1} B_N(\Omega^{-1} + B_N \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1} ) B_N \tag{2.A.72}
\]
\[
= B_N' \Sigma_N^{-1} B_N(\Omega^{-1} + B_N \Sigma_N^{-1} B_N)^{-1} \Omega^{-1}, \tag{2.A.73}
\]
\[
= (\Omega^{-1} + \Omega^{-1}) (\Omega^{-1} + B_N \Sigma_N^{-1} B_N)^{-1} \Omega^{-1}, \tag{2.A.74}
\]
\[
= \Omega^{-1} - \Omega^{-1} (\Omega^{-1} + B_N \Sigma_N^{-1} B_N)^{-1} \Omega^{-1}. \tag{2.A.75}
\]

Therefore the positive-definite matrix \( B_N' \Sigma_N^{-1} B_N \) is bounded above by the constant matrix \( \Omega^{-1} \), implying boundedness of the former. Premultiplying and postmultiplying the above expression by \( \lambda \) yields the result.

Finally,

\[
(\text{SR}^N)^2 = (\mu_N - \gamma N)' \Sigma_N^{-1} (\mu_N - \gamma N) \tag{2.A.76}
\]
\[
= (\tilde{\alpha}_N + B_N \lambda)' \Sigma_N^{-1} (\tilde{\alpha}_N + B_N \lambda) \tag{2.A.77}
\]
\[
= \tilde{\alpha}_N \lambda' \lambda + \lambda' B_N \Sigma_N^{-1} B_N \lambda \tag{2.A.78}
\]
\[
= \tilde{\alpha}_N \Sigma_N^{-1} \tilde{\alpha}_N + \lambda' B_N \Sigma_N^{-1} B_N \lambda \tag{2.A.79}
\]
\[
= (\text{SR}^N)^2 + (\text{SR}^N)^2. \tag{2.A.80}
\]

where the third equality follows from orthogonality.

2.A.8 Proof of Theorem 2.8

Regarding part (i), by orthogonality \( (\mu_N - \gamma N)' \Sigma_N^{-1} (\mu_N - \gamma N) = \tilde{\alpha}_N \Sigma_N^{-1} \tilde{\alpha}_N + \lambda' B_N \Sigma_N^{-1} B_N \lambda \). This implies \( 0 \leq \phi^\alpha \leq 0, 0 \leq \phi^\beta \leq 1 \).

We now consider the unbounded variation case. The bounded variance case will follow by setting \( A_N = 0 \). Recall that now \( \Sigma_N = A_N A_N' + C_N \) and \( \alpha_N = A_N \lambda_{\text{miss}} + a_N \). Consider first \( w_N^x \), where its \( i \)th component satisfies

\[
w_{N,i}^x = 1_{N,i} w_N^\alpha - \frac{1}{\gamma^\alpha} 1_{N,i} \Sigma_N^{-1} \alpha_N - \frac{1}{\gamma^\alpha} 1_{N,i} \Sigma_N^{-1} B_N(B_N' \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1} \alpha_N, \tag{2.A.81}
\]

where \( 1_{N,i} \) is an \( N \)-dimensional vector in which the \( i \)th element is one and the rest of the elements are zero. We deal with the two terms on the right-hand side of \( w_{N,i}^x \), separately. By the Sherman-Morrison-Woodbury formula \( \Sigma_N^{-1} = C_N^{-1} - C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} \), obtaining

\[
1_{N,i} \Sigma_N^{-1} \alpha_N = 1_{N,i} C_N^{-1} \alpha_N - 1_{N,i} C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} \alpha_N \tag{2.A.82}
\]
\[
= 1_{N,i} C_N^{-1} A_N \lambda_{\text{miss}} - 1_{N,i} C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N \lambda_{\text{miss}} \tag{2.A.83}
\]
\[
+ 1_{N,i} C_N^{-1} a_N - 1_{N,i} C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} a_N \tag{2.A.84}
\]
\[
= 1_{N,i} C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} \lambda_{\text{miss}} \tag{2.A.85}
\]
\[
+ 1_{N,i} C_N^{-1} a_N - 1_{N,i} C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} a_N. \tag{2.A.86}
\]

By Holder’s inequality, taking the norm and using the relation between norm and maximum eigenvalue, one obtains

\[
|1_{N,i} \Sigma_N^{-1} \alpha_N| = O\left(\|\lambda_{\text{miss}}\| \frac{1_{N,i} C_N^{-1} A_N}{f(N)} + 1_{N,i} C_N^{-1} a_N \right) + \|a_N\| \frac{1_{N,i} C_N^{-1} A_N}{f(N)}. \tag{2.A.87}
\]

Along the same lines

\[
1_{N,i} \Sigma_N^{-1} B_N = 1_{N,i} C_N^{-1} B_N - 1_{N,i} C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} B_N, \tag{2.A.88}
\]
\[
B_N' \Sigma_N^{-1} B_N = B_N C_N^{-1} B_N - B_N C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} B_N, \tag{2.A.89}
\]
\[
B_N' \Sigma_N^{-1} \alpha_N = B_N' C_N^{-1} \alpha_N - B_N C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} \alpha_N \tag{2.A.90}
\]
\[
= B_N C_N^{-1} A_N \lambda_{\text{miss}} - B_N C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N \lambda_{\text{miss}} \tag{2.A.91}
\]
\[
+ B_N C_N^{-1} a_N - B_N C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} a_N \tag{2.A.92}
\]
\[
= B_N C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} \lambda_{\text{miss}} \tag{2.A.93}
\]
\[
+ B_N C_N^{-1} a_N - B_N C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} a_N. \tag{2.A.94}
\]
Therefore, using the same arguments as above, one obtains

\[
|\beta' N^{-1} B_N (\beta' N^{-1} B_N)^{-1} B_N' N^{-1} \alpha_N| = \mathcal{O}\left(\frac{\|\beta' N^{-1} A_N\| + \|\beta' N^{-1} B_N\|}{f^{\frac{1}{2}}(N)}\right),
\]

because, under our assumptions, the eigenvalues of \(\beta' N^{-1} A_N\) and \(\beta' N^{-1} B_N\) have the same behavior. In particular, the first term \(\beta' N^{-1} B_N\) is \(\mathcal{O}(\|\beta' N^{-1} A_N\| + \|\beta' N^{-1} B_N\|)\), the second term \(B_N' N^{-1} B_N\) is \(\mathcal{O}(f(N))\), and the third term \(B_N' N^{-1} \alpha_N\) is \(\mathcal{O}(f^{\frac{1}{2}}(N))\).

For the \(w_N\) portfolio, its ith component satisfies

\[
w_{N,i} = \frac{1}{\gamma'} \beta' N^{-1} B_N \lambda' - \frac{1}{\gamma'} \beta' N^{-1} B_N (\Omega^{-1} + B_N' N^{-1} B_N)^{-1} B_N' N^{-1} B_N \lambda
\]

\[
= \frac{1}{\gamma'} \beta' N^{-1} B_N (\Omega^{-1} + B_N' N^{-1} B_N)^{-1} \Omega^{-1} \lambda
\]

(2.95)

(2.96)

and using the above formula for \(\beta' N^{-1} B_N\) and \(B_N' N^{-1} B_N\) concludes, where we use \(\lambda' \rightarrow \lambda' \neq 0\).

Regarding part (ii), \(w_{N,i}^aw_{N,i}^b < \|\beta' N^{-1} B_N\| (\alpha' N^{-1} \alpha_N)/\gamma' < \delta < \infty\). Moreover, \(\lambda' B_N' V_N^{-1} V_N^{-1} B_N \lambda' = \lambda' / \Omega^{-1}(\Omega^{-1} + B_N' N^{-1} B_N)^{-1} B_N' N^{-1} B_N \lambda'\). This implies \(\|w_{N,i}^a\|^2 \leq \|\beta' N^{-1} B_N\|^2\). This means that the squared norm of \(w_{N,i}^a\) goes to zero.

Part (iii) follows from \(\beta' N^{-1} B_N \leq \frac{1}{\gamma'} (\beta' N^{-1} \beta_N)^{\frac{1}{2}}\) with \(\beta' N^{-1} \beta_N \rightarrow 0\). In fact \(\beta' N^{-1} \beta_N \geq \gamma' (\gamma' + 1) N \geq 0\). On the other hand, \(\beta' N^{-1} B_N \leq \frac{1}{\gamma'} (\beta' N^{-1} \beta_N)^{\frac{1}{2}}\). By assumption, \(\alpha' N^{-1} \alpha_N \leq \delta < \infty\).

Part (iv) follows from part (i) once we show that the coefficients \(\phi'\) and \(\phi''\) are bounded away from zero. By assumption, the limit of \(\alpha' N^{-1} \alpha_N \leq \delta < \infty\). On the other hand, \(\beta' N^{-1} B_N \leq \lambda' / \Omega^{-1} \lambda' > 0\). Therefore the normalizing constants \(\phi'\) and \(\phi''\) remain bounded away from zero also asymptotically and can be ignored for the purpose of evaluating the limiting behavior of the weights.

**Remark 2.2** For the bounded-residual-variation case, the result in (2.40) follows from the fact that, under Assumptions 2.1 and 2.2, the absolute value of the components of the mean-variance portfolio vector decreases at most at the rate:

\[
|w_{N,i}| = \mathcal{O}\left(\frac{|\beta' N^{-1} \alpha_N| + \|\beta' N^{-1} B_N\|}{f^{\frac{1}{2}}(N)}\right),
\]

(2.97)

\[
|w_{N,i}^\delta| = \mathcal{O}\left(\frac{\|\beta' N^{-1} B_N\|}{f(N)}\right),
\]

(2.98)

From equations (2.97) and (2.98), we see that \(w_{N,i}^\delta\) can dominate \(w_{N,i}^\beta\) as the number of assets increases. In particular, \(w_{N,i}^\delta\) dominates \(w_{N,i}^\beta\) when the pricing-error term, \(\beta' N^{-1} \alpha_N\), goes to zero slowly as the number of assets increases. The weights \(w_{N,i}^\beta\), dominate \(w_{N,i}^\delta\) also because the second term on the right-hand side of (2.97), \(\|\beta' N^{-1} B_N\| / f^{\frac{1}{2}}(N)\), dominates the term on the right-hand side of (2.98), \(\|\beta' N^{-1} B_N\| / f(N)\); this dominance arises because the two terms have different denominators: \(f^{\frac{1}{2}}(N)\) instead of \(f(N)\).

Recall that the APT bounds the pricing error from above; that is, \(\alpha' N^{-1} \alpha_N \leq \delta < \infty\). However, the APT is silent about whether \(\alpha' N^{-1} \alpha_N \leq \delta < \infty\). When this expression is bounded away from zero, one can show that the ratio \(w_{N,i}^\beta / w_{N,i}^\delta\) always decreases at a rate that is equal or faster than \(1/f^{\frac{1}{2}}(N)\).

**Remark 2.3** For the unbounded-residual-variation case (that is, where \(\alpha_N = A_N \lambda_{miss} + a_N\) and \(\Sigma_N = A_N A_N' + C_N\)), the result in (2.40) follows from the fact that, under Assumptions 2.1 and 2.2, the absolute value of the components of the mean-variance portfolio vector decreases at most at the rate

\[
|w_{N,i}^\beta| = \mathcal{O}\left(\frac{\|\beta' N^{-1} A_N\| + \|\beta' N^{-1} B_N\|}{f^{\frac{1}{2}}(N)}\right),
\]

(2.99)

\[
|w_{N,i}^\delta| = \mathcal{O}\left(\frac{\|\beta' N^{-1} A_N\|}{f(N)}\right).
\]

(2.100)
Recall that \( \lambda_{\text{miss}} \) can be interpreted as the risk premia on the unobserved factors with loadings \( \Lambda_N \), whereas the vector \( \alpha_N \) represents the pure pricing error that is not associated with a factor structure. The \( \alpha_N \) component dominates the behavior of the portfolio weights \( w_N^\alpha \) in (2.1.99), whereas the risk premia \( \lambda_{\text{miss}} \) component declines to zero faster. In general, the portfolio weight \( w_j^\beta \) in (2.1.100) declines at the same, fast rate as the risk premia \( \lambda_{\text{miss}} \) component of \( w_N^\beta \). When \( \alpha_N \) is non-zero then, as before, the \( w_N^\alpha \) portfolio dominates the \( w_N^\beta \) portfolio across all three norms considered in the theorem above.

**Remark 2.4** Equation (2.41) shows that the portfolio \( w_N^\alpha \) dominates \( w_N^\beta \) for large \( N \). Observe that \( w_N^\alpha \) is functionally independent of the factor risk premia, \( \lambda \), and the factor covariance matrix, \( \Omega \), making it robust to misspecification in the beta component of returns by construction. In contrast, portfolio \( w_N^\beta \) depends on both \( \lambda \) and \( \Omega \).

### 2.1.9 Proof of Theorem 2.9

Define \( w_N^\text{bench} \) to be the benchmark portfolio constructed imposing the target mean, \( \mu^* \); that is: \( w_N^{\text{bench}} = \frac{\mu^* - r_f}{\sum (\mu^* - r_f)} \). \( w_N^{\text{bench}} \) and \( \gamma^{\text{bench}} = \frac{\gamma}{\mu^* - r_f} \), and \( \gamma \) given in (2.26).

Consider first the case when \( c^{\text{bench}} = \delta \Omega^{-1} \), for some scalar \( \delta \neq 0 \). Then \( B_N^\beta w_N^{\text{bench}} \rightarrow \frac{\delta (\mu^* - r_f)}{\Omega^{-1} \lambda} \). Moreover, because \( (w_N^{\text{bench}})^\top V_N w_N^{\text{bench}} \rightarrow (c^{\text{bench}})^\top \Omega^{\text{bench}} = \delta^2 \lambda \Omega^{-1} \lambda \), then \( (w_N^{\text{bench}})^\top V_N w_N^{\text{bench}} \rightarrow \frac{(\mu^* - r_f)^2}{\lambda \Omega^{-1} \lambda} \), yielding

\[
\lim_{N \to \infty} (SR_{\text{bench}}^\beta)^2 \rightarrow (SR_{\infty}^\beta)^2 = \left( \frac{\mu^* - r_f}{\lambda \Omega^{-1} \lambda + \alpha_N^\top \Sigma_N^{-1} \alpha_N} \right)^2.
\]

Therefore,

\[
\phi^{\text{bench}} \sim \phi^\beta \sim \frac{\lambda \Omega^{-1} \lambda}{\lambda \Omega^{-1} \lambda + \alpha_N^\top \Sigma_N^{-1} \alpha_N},
\]

and Lemma 1 is satisfied because

\[
\frac{\phi^{\text{bench}} (w_N^{\text{bench}})^\top (\mu - r_f 1_N)}{(\phi^{\text{bench}})^2 (w_N^{\text{bench}})^\top V_N w_N^{\text{bench}}} \rightarrow \frac{\lambda \Omega^{-1} \lambda + \alpha_N^\top \Sigma_N^{-1} \alpha_N}{\lambda \Omega^{-1} \lambda} \frac{(\mu^* - r_f)^2}{(\mu^* - r_f)^2/\lambda \Omega^{-1} \lambda} = \frac{\mu^* - r_f}{\mu^* - r_f}
\]

and

\[
\frac{\phi^\alpha (w_N^\alpha)^\top (\mu - r_f 1_N)}{(\phi^\alpha)^2 (w_N^\alpha)^\top V_N w_N^\alpha} \rightarrow \frac{(\mu^* - r_f)^2}{\mu^* - r_f}
\]

implying \( (SR_{\infty}^\beta)^2 \sim (SR_{\infty}^\beta)^2 + (SR_{\text{bench}}^\beta)^2 \).

If \( K = 1 \), then \( c^{\text{bench}} = \delta \Omega^{-1} \lambda = \delta \lambda \) is always satisfied, and any benchmark portfolio \( w_N^{\text{bench}} \) satisfying (2.44) can be combined with \( w_N^\alpha \) to obtain mean-variance efficiency. Now consider the case when \( c^{\text{bench}} \) is not proportional to \( \lambda \). Then,

\[
B_N^\beta w_N^{\text{bench}} \rightarrow \left( \frac{(\mu^* - r_f)}{\phi^{\text{bench}} \lambda} \right) c^{\text{bench}} \text{ and } (w_N^{\text{bench}})^\top V_N w_N^{\text{bench}} \rightarrow \left( \frac{(\mu^* - r_f)^2}{\phi^{\text{bench}} \lambda} \right) (c^{\text{bench}})^\top \Omega^{\text{bench}},
\]

yielding \( (SR_{\text{bench}}^\beta)^2 \rightarrow \frac{(c^{\text{bench}})^\top \Omega^{\text{bench}}}{(c^{\text{bench}})^\top \Omega^{\text{bench}}} = (SR_{\text{fin}}^\beta)^2 \). However, note that Lemma 1 does not hold, implying that

\[
(SR_{\infty}^\alpha)^2 + (SR_{\infty}^{\text{bench}})^2 = (\alpha_N^\top \Sigma_N^{-1} \alpha_N) + \frac{((c^{\text{bench}})^\top \lambda)^2}{(c^{\text{bench}})^\top \Omega^{\text{bench}}} < (\alpha_N^\top \Sigma_N^{-1} \alpha_N) + \lambda \Omega^{-1} \lambda = (SR_{\infty}^\text{mv})^2,
\]

because

\[
\frac{(c^{\text{bench}})^\top \Omega^{\text{bench}}}{(c^{\text{bench}})^\top \Omega^{\text{bench}}} = \frac{(c^{\text{bench}})^\top \Omega^{-1} \lambda}{(c^{\text{bench}})^\top \Omega^{\text{bench}}} \leq \frac{(c^{\text{bench}})^\top \Omega^{-1} \lambda}{(c^{\text{bench}})^\top \Omega^{\text{bench}}} = \lambda \Omega^{-1} \lambda.
\]

The strict inequality is implied whenever \( \Omega^{\text{bench}} \) and \( \Omega^{-1} \lambda \) are not proportional, which is implied by \( c^{\text{bench}} \) being not being proportional to \( \Omega^{-1} \lambda \), implied by our assumption.
Remark 2.5 For portfolio weights $w_N^α$ and $w_N^β$ that satisfy the assumptions of Theorem 2.8 we have the following results: for $N \to \infty$, the asymptotic means of the excess returns on the portfolios $w_N^α$ and $w_N^β$,

$$
\mu_∞^α - r_f = \lim_{N \to \infty} (\mu_∞^α - r_f) = \lim_{N \to \infty} \frac{1}{\gamma_α} V_N α_N, \quad \mu_∞^β - r_f = \lim_{N \to \infty} (\mu_∞^β - r_f) = \frac{1}{\gamma_β} \lambda_0 \Omega^{-1} \lambda
$$

(2.107)

satisfy $0 < (\mu_∞^α - r_f) < \infty$ and $0 < (\mu_∞^β - r_f) < \infty$. Furthermore, the asymptotic variances and squared Sharpe ratios of the returns on the portfolios $w_N^α$ and $w_N^β$ are given by the same expressions as the means and satisfy the same properties.

We now interpret the inequality $\mu_∞^α - r_f > 0$ in terms of the asymptotic variance of the portfolio. Recall that $\mu_∞^α$ represents not just the asymptotic mean return but also the asymptotic variance of the portfolio, which is equal to the limit of the idiosyncratic variance of the portfolio because for any $N$,

$$\text{var}(r_t^N w_N^α) = \text{var}(ε_t^N w_N^α) = \alpha_N^\prime V_N α_N.$$

(2.108)

We will have exact equality, $\mu_∞^α - r_f = 0$, in two cases: one, where each element of $α_N$ is zero; that is, the case of exact pricing (see Chamberlain (1983b, Corollary 1)), and two, where $α_N$ is asymptotically collinear with $B_N$; see footnote 28 for the definition of asymptotically collinearity.

2.6.10 Proof of Theorem 2.10

The formulae for $α_{N\text{, MLC}}, \hat{B}_{N\text{, MLC}}$ and $\hat{Σ}_{N\text{, MLC}}$ follow from solving the first-order conditions. For $\hat{λ}_{\text{MLC}}$ and $\hat{Π}_{\text{MLC}}$, one obtains precisely the sample mean and sample covariance matrix of $f_t$.

2.6.11 Proof of Theorem 2.11

Differentiating the penalized log-likelihood with respect to $λ_{\text{miss}}$ and $a_N$, one obtains the following $K + N$ equations (after some algebra):

$$
\left( A_N' Σ_N^{-1} I_N \right) \left( r_N - r_f 1_N - B_N(\bar{f} - r_f 1_K) \right) = \left( A_N' Σ_N^{-1} A_N \quad A_N' Σ_N^{-1} \begin{pmatrix} A_N \\ (1 + κ)I_N \end{pmatrix} \right) \left( \hat{λ}_{\text{miss}} \quad \hat{a}_N \right),
$$

(2.109)

where, for simplicity, we set $\hat{λ}_{\text{miss}} = \hat{λ}_{\text{miss, MLC}}, \hat{a}_N = \hat{a}_{N\text{, MLC}}$ and recall that $Σ_N = A_N A_N' + C_N$. It is straightforward to see that, because of the APT restriction, $λ_{\text{miss}}$ and $a_0$ can now be identified separately, as long as $κ > 0$. In fact, the above system of linear equations can be solved because the matrix pre-multiplying $λ_{\text{miss}}$ and $a_N$ is non-singular for every $κ > 0$, leading to the closed-form solution,

$$\hat{λ}_{\text{miss}} = (A_N' Σ_N^{-1} A_N)^{-1} A_N' Σ_N^{-1} \left( r_N - r_f 1_N - B_N(\bar{f} - r_f 1_K) \right),$$

(2.110)

and

$$\hat{a}_N = \frac{1}{κ + 1} \left( r_N - r_f 1_N - B_N(\bar{f} - r_f 1_K) - A_N \hat{λ}_{\text{miss}} \right).$$

(2.111)

Turning now to the first-order condition with respect to the generic $(a, b)$th element of $B_N$, denoted by $B_{ab}$, one obtains,

$$\frac{1}{T} \sum_{t=1}^T g_t Σ_N^{-1}( - \frac{∂B_N}{∂B_{ab}} \bar{f}_t + G_N \frac{∂B_N}{∂B_{ab}} \bar{f}_t ) = 0, \text{ with } 1 ≤ a ≤ N, 1 ≤ b ≤ K,$$

(2.112)

setting, for simplicity,

$$g_t = \left( \bar{r}_t - G_N \bar{r}_N - \bar{B}_N \bar{f}_t + G_N \bar{B}_N \bar{f} \right) \text{ and } g = \frac{1}{T} \sum_{t=1}^T g_t,$$

with

$$G_N = \frac{1}{(κ + 1)} I_N + \frac{κ}{(κ + 1)} A_N (A_N' Σ_N^{-1} A_N)^{-1} A_N' Σ_N^{-1},$$

$$\left( A_N' Σ_N^{-1} I_N \right) \left( r_N - r_f 1_N - B_N(\bar{f} - r_f 1_K) \right) = \left( A_N' Σ_N^{-1} A_N \quad A_N' Σ_N^{-1} \begin{pmatrix} A_N \\ (1 + κ)I_N \end{pmatrix} \right) \left( \hat{λ}_{\text{miss}} \quad \hat{a}_N \right).$$

(2.109)
and
\begin{align}
\tilde{r}_t &= (r_t - r_1 1_N), \\
\tilde{f}_t &= (f_t - r_1 1_K), \\
\bar{r}_N &= (\bar{r}_N - r_1 1_N), \\
\bar{f} &= (\bar{f} - r_1 1_K). 
\end{align}
(2.113)

Taking the vec operator for both sides of the first-order condition above gives
\[
\frac{1}{T} \sum_{t=1}^{T} (\tilde{f}_t' \otimes g'_t (\Sigma_N^{-1}) \text{vec}(\frac{\partial B_N}{\partial B_{ab}})) = (\tilde{f}' \otimes g' (\Sigma_N^{-1}) G_N) \text{vec}(\frac{\partial B_N}{\partial B_{ab}}),
\]
with \(1 \leq a \leq N, 1 \leq b \leq K\),

which can be more succinctly rewritten as
\[
\frac{1}{T} \sum_{t=1}^{T} \tilde{f}_t g'_t = \tilde{f}' g' (\Sigma_N^{-1}) G_N \Sigma_N.
\]

Next, recalling that \(\Sigma_{rf} = \frac{1}{T} \sum_{t=1}^{T} \tilde{r}_t \tilde{f}_t\)' and \(\Sigma_{rf} = \frac{1}{T} \sum_{t=1}^{T} \tilde{r}_t \tilde{f}_t\), with \(\Sigma_{rf} = \Sigma_{rf}'\), one obtains \(\Sigma_{rf} G_N \Sigma_N = \frac{1}{(\kappa + 1)} I_N + \frac{\kappa}{(\kappa + 1)} A_N (\Sigma_N^{-1} A_N^{-1})^{-1} A_N' = G_N\) and rearranging the above first-order condition gives
\begin{align}
\Sigma_{rf} - \tilde{f}' \tilde{r}'_N G_N' - \Sigma_{rf} \Sigma_{rf}' G_N' = \tilde{f}' \tilde{r}'_N (2 G_N' - G_N G_N') - \Sigma_{rf} \Sigma_{rf}' G_N' + \tilde{f}' \tilde{r}'_N (2 G_N' - G_N G_N') = 0. 
\end{align}
(2.115)

Transposing both sides, taking the vec, and solving for vec(\(B_N\)) gives
\[
\text{vec}(\tilde{B}_N) = \left( (\Sigma_{rf} \otimes I_N) - (\tilde{f}' \otimes (2 G_N - G_N G_N)) \right)^{-1} \text{vec}(\Sigma_{rf} - (2 G_N - G_N G_N) \tilde{r}_N \tilde{f}').
\]
(2.116)

We need to show that a solution for \(\tilde{B}_N\) exists. This requires one to establish that the matrix \( (\Sigma_{rf} \otimes I_N) - (\tilde{f}' \otimes (2 G_N - G_N G_N)) \) is invertible. This matrix can be written as
\[
(\Sigma_{rf} \otimes I_N) - (\tilde{f}' \otimes (2 G_N - G_N G_N)) = \left( (\Sigma_{rf} - \tilde{f}') \otimes I_N \right) + (\tilde{f}' \otimes (I_N - (2 G_N - G_N G_N))).
\]
(2.117)

The first matrix on the right hand side is non-singular. One then just needs to show that the second matrix is positive semi-definite. This follows because, \(I_N - (2 G_N - G_N G_N) = (I_N - G_N) (I_N - G_N)\), and we show below that \((I_N - G_N)\) is positive semi-definite.

\begin{align}
I_N - G_N &= I_N - \frac{1}{(\kappa + 1)} I_N - \left( \frac{\kappa}{(\kappa + 1)} A_N (A_N^t \Sigma_N^{-1} A_N)^{-1} A_N^t \Sigma_N^{-1} \right) \\
&= \left( \frac{\kappa}{(\kappa + 1)} I_N - A_N (A_N^t \Sigma_N^{-1} A_N)^{-1} A_N^t \Sigma_N^{-1} \right) \\
&= \left( \frac{\kappa}{(\kappa + 1)} I_N - A_N (A_N^t \Sigma_N^{-1} A_N)^{-1} A_N^t \Sigma_N^{-1} \right) \\
&= \left( \frac{\kappa}{(\kappa + 1)} I_N - I_N - (I_N - 1/2) A_N (A_N^t \Sigma_N^{-1} A_N)^{-1} A_N^t \Sigma_N^{-1} \right) \\
&= \frac{1}{1 + \kappa} (\Sigma_N - I_N - (I_N - 1/2) A_N (A_N^t \Sigma_N^{-1} A_N)^{-1} A_N^t \Sigma_N^{-1} \right).
\end{align}
(2.118)

The right-hand side is the product of positive-definite matrices, namely \(\Sigma_N\) and \(\Sigma_N^{-1/2}\), and of the matrix \(I_N - (I_N - 1/2) A_N (A_N^t \Sigma_N^{-1} A_N)^{-1} A_N^t \Sigma_N^{-1/2}\), which is a projection matrix orthogonal to \(\Sigma_N^{-1/2} A_N\), and therefore, positive semi-definite. Therefore, plugging \(B_N\) into \(\hat{\lambda}_{miss}\) and \(\hat{a}_N\), one obtains that
\[
\hat{\lambda}_{miss} = \hat{\lambda}_{miss}(A_N, C_N, \kappa), \quad \hat{a}_N = \hat{a}_N(A_N, C_N, \kappa),
\]
because \(\tilde{B}_N = \tilde{B}_N(A_N, C_N, \kappa)\). Substituting them into \(L(\theta) = \kappa (A_N^t \Sigma_N^{-1} A_N - \delta)\), gives the concentrated likelihood function, which is a function of only \(A_N\) and \(C_N\). Notice that all the estimates depend ultimately on \(\kappa\), which can be chosen by cross-validation methods.

Consider now the case where \(\kappa = 0\). The first-order conditions (2.109) become
\[
\left( A_N^t \Sigma_N^{-1} I_N^t \right) (\tilde{r}_N - r_1 1_N - B_N (\tilde{f} - r_1 1_K)) = \left( A_N^t \Sigma_N^{-1} I_N^t \right) (\tilde{r}_N, I_N) (\hat{\lambda}_{miss} \hat{a}_N).
\]
(2.121)
One can clearly obtain a unique solution for \((A_N, I_N)(\hat{\lambda}_{\text{miss}}^* \hat{\alpha}_N) = A_N \hat{\lambda}_{\text{miss}}^* + \hat{\alpha}_N\). However, to solve for \(\hat{\lambda}_{\text{miss}}^*\) and \(\hat{\alpha}_N\) separately, one needs to invert the matrix

\[
\begin{pmatrix}
A_N' \Sigma_N^{-1} \\
I_N
\end{pmatrix}
\begin{pmatrix}
A_N, I_N
\end{pmatrix}
= 
\begin{pmatrix}
A_N' \Sigma_N^{-1} A_N & A_N' \Sigma_N^{-1} I_N \\
A_N & I_N
\end{pmatrix}
\]

which is not possible because it is of dimension \((N + K) \times (N + K)\) but of rank \(N\), as the left-hand side shows that it is obtained from the product of two matrices of dimension \((N + K) \times N\). All the other parameters are identified separately, along with \(\alpha_0 = A_0 m_0 + a_0\), and their expressions follow from differentiating \(L(\theta)\) and solving the resulting first-order conditions.

2.12 Estimation for unbounded residual variation case with mismeasured factors

In this section, we consider the estimation of the model for the case in which all factors are observed but are measured with error: \(f_t = f_{0,t} + \eta_t\), where \(\eta_t\) has mean \(\mu_{\eta 0}\) and covariance matrix \(\Sigma_{\eta 0}\). Then, \(\alpha_0 = -B_0 \mu_{\eta 0}\) and \(\Sigma_0 = B_0 \Sigma_{\eta 0} B_0' + C_0\), where \(C_0\) is an \(N \times N\) positive-definite matrix with bounded eigenvalues that represents the covariance matrix of the pure idiosyncratic component of the error returns and where the subscript "0" indicates the true value of a parameter (and we do not use the subscript "N" for the true value of parameters in order to limit the number of subscripts).46

The case of mismeasured factors turns out to be a particular case of the so-called multivariate errors-in-variables model; see Fuller (1987, Chapter 4) for a classical reference.47 Although several approaches for estimation of multivariate errors-in-variables models exist (see, for example, Fuller (1987)) in analogy with the other forms of misspecification discussed above, we illustrate here ML estimation. Every estimation method of errors-in-variables models requires some further information, beyond a sample of observations of \(r_t\) and \(f_t\). For ML estimation we assume that we have the availability of a preliminary estimate, \(S_{C_0}\), of the \((N + K) \times (N + K)\) matrix \(S_{C_0} = \begin{pmatrix} C_0 & 0 \\ 0 & \Sigma_0 \end{pmatrix}\), with certain characteristics described in the statement of the theorem below and in Amemiya and Fuller (1984, Theorem 3).

To derive the log-likelihood function for estimation, it is convenient to adopt a slightly different parameterization from Case 4 in Section 2.3.5, rewriting the model as

\[
\begin{pmatrix}
\alpha_0 \\
B_0 & I_K
\end{pmatrix}
\begin{pmatrix}
f_{0,t} + \mu_{\eta 0} - r_f 1_K \\
r_t - r_f 1_N
\end{pmatrix}
+ \begin{pmatrix}
e_t \\
\eta_t - \mu_{\eta 0}
\end{pmatrix}
\]

with \(\alpha_0 = -B_0 \mu_{\eta 0}\). Then setting

\[
Z_t = \begin{pmatrix}
r_t - r_f 1_N \\
\frac{f_t}{f_t - r_f 1_K}
\end{pmatrix}
\]

one obtains the \((N + K) \times (N + K)\) matrix:

\[
\text{cov}(Z_t) = \Sigma_{Z_0} = \begin{pmatrix} B_0 & B_0 \\ I_K & I_K \end{pmatrix} \Omega_0 \begin{pmatrix} B_0 & B_0 \\ I_K & I_K \end{pmatrix} + \Sigma_{C_0}; \quad E(Z_t) = \mu_{Z_0} = \begin{pmatrix} \alpha_0 \\ 0_K \end{pmatrix} + \begin{pmatrix} B_0 & B_0 \\ I_K & I_K \end{pmatrix} \lambda_0,
\]

where we set \(\lambda_0 = E(f_{0,t} + \mu_{\eta 0} - r_f 1_K)\) and \(\Omega_0 = \text{Cov}(f_{0,t})\). We have implicitly assumed that \(\varepsilon_t, \eta_t\) and \(f_{0,t}\) are uncorrelated for any \(t, s, l\) in the evaluation of \(\Sigma_{Z_0}\). Therefore, when \((f_{0,t}, \varepsilon_t, \eta_t)'\) are iid and jointly normal, with a block-diagonal covariance matrix,48 the log likelihood (here adjusted for the degrees

\[\begin{align*}
\alpha_0' \Sigma_{\eta 0}^{-1} \alpha_0 &= \mu_{\eta 0}' C_0' B_0 \left( \Sigma_{\eta 0}^{-1} + B_0 C_0' B_0 \right)^{-1} \Sigma_{\eta 0}^{-1} \mu_{\eta 0}. \\
\intertext{Therefore, \(\alpha_0' \Sigma_{\eta 0}^{-1} \alpha_0\) converges to \(\mu_{\eta 0}' C_0' B_0 \left( \Sigma_{\eta 0}^{-1} + B_0 C_0' B_0 \right)^{-1} \Sigma_{\eta 0}^{-1} \mu_{\eta 0}\) implying that the APT restriction is always satisfied in this case, and therefore, there is no need to impose it in the estimation.}
\end{align*}\]

46Observes that \(B_0\) appears in both the intercept and residual covariance matrix of returns. Recognizing this allows one to improve the precision of the estimated \(B_0\), as discussed in the previous section. Using the Sherman-Morrison-Woodbury formula, we see that

\[
\alpha_0' \Sigma_{\eta 0}^{-1} \alpha_0 = \mu_{\eta 0}' C_0' B_0 \left( \Sigma_{\eta 0}^{-1} + B_0 C_0' B_0 \right)^{-1} \Sigma_{\eta 0}^{-1} \mu_{\eta 0}.
\]

47Typically, errors-in-variables models are classified as being either functional or structural, depending on whether the \(f_{0,t}\) are assumed constant or random variables with a known distribution. Here we adopt the latter interpretation.

48Mutual independence of \(\varepsilon_t\) with \(\eta_t\) is not necessary but is assumed here for the sake of simplicity.
of freedom) of model (2.24), with respect to \( \mathbf{Z}_t \) as the observable vector, is

\[
L(\theta) = -\frac{T - 1}{2T} \log(\det(\Sigma_Z)) - \frac{T - 1}{2T^2} \sum_{t=1}^{T} (\mathbf{Z}_t - \mu_Z)' \Sigma_Z^{-1}(\mathbf{Z}_t - \mu_Z) \tag{2.1.25}
\]

\[
- \frac{d}{2T} \log(\det(\Sigma_{C_\eta})) - \frac{d}{2T} \text{tr}((\Sigma_{C_\eta})^{-1} \Lambda_{C_\eta}), \tag{2.1.26}
\]

setting \( \theta = (\mu'_\eta, \text{vech}(\Sigma_\eta)', \text{vech}(\mathbf{B}_N)', \text{vech}(\mathbf{C}_N)', \text{vech}(\Omega)', \Lambda')' \) with \( \mu_Z = \left( \frac{\mathbf{B}_N}{\mathbf{I}_K} \right)' \) and \( \Sigma_Z = \left( \begin{array}{c} \mathbf{B}_N' \\ \mathbf{I}_K \end{array} \right)' \left( \begin{array}{c} \mathbf{B}_N \\ \mathbf{I}_K \end{array} \right) + \Sigma_{C_\eta}. \) All the parameters' estimators are highly nonlinear, except for the estimator for \( \lambda_\eta, \) but their closed-form expression has been established, as described in the next theorem.\(^{49}\)

The following theorem, exploits known results from Fuller (1987, Section 4.1.1, p. 298) adapted to our framework, and hence, is stated without proof.

**Theorem 2.5 (Mismeasured-factors case)** Suppose that the vector of asset returns, \( \mathbf{r}_t, \) satisfies Assumption 2.1. Assume that \( \Sigma_{C_{\eta_0}} > 0 \) and that its estimator \( \hat{\Sigma}_{C_\eta} \) is an unbiased estimator for \( \Sigma_{C_{\eta_0}}, \) distributed as Wishart with parameter \( \Sigma_{C_{\eta_0}} \) and degrees of freedom \( d > 0, \) for some constant \( d \) and independent of \( \{\mathbf{r}_t, \mathbf{f}_t\}. \) Set \( \hat{\Sigma}_g = (T - 1)^{-1} \sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{Z}_t \) and let \( \hat{\gamma}_{1:N+K} \geq \cdots \geq \hat{\gamma}_{N+K,N+K} \) be the eigenvalues of \( \hat{\Sigma}_{C_\eta} \) with corresponding eigenvectors \( \hat{\mathbf{v}}_{1:N+K}, \) with \( 1 \leq i \leq N + K, \) the first \( K \) of which we collect as

\[
\hat{\mathbf{G}} = \left( \begin{array}{ccc} \hat{\gamma}_{1:N+K} & & \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\gamma}_{K,N+K} \end{array} \right) \text{ and } \hat{\mathbf{V}} = \left( \begin{array}{c} \hat{\mathbf{v}}_{1:N+K} \\ \vdots \\ \hat{\mathbf{v}}_{K,N+K} \end{array} \right).
\]

If \( \hat{\gamma}_{K,N+K} > 1, \) then \( \hat{\theta}_{MLC} = (\hat{\mu}'_{\eta,MLC}, \text{vech}(\hat{\Sigma}'_{\eta,MLC})', \text{vech}(\hat{\mathbf{B}}_{MLC})', \text{vech}(\hat{\mathbf{C}}_{MLC})', \text{vech}(\hat{\Omega}_{MLC})', \Lambda_{MLC})' \) satisfies

\[
\hat{\theta}_{MLC} = \text{argmax}_{\theta} L(\theta), \tag{2.1.27}
\]

where \( L(\theta) \) is defined in (2.1.26),

\[
\hat{\mathbf{B}}_{MLC} = \mathbf{P}_{NK} \mathbf{P}_{KK}^{-1}, \tag{2.1.28}
\]

\[
\hat{\mu}_{\eta,MLC} = -\left( \hat{\mathbf{B}}_{MLC} \hat{\mathbf{B}}_{MLC} \right)^{-1} \hat{\mathbf{B}}_{MLC} (T - 1)^{-1} \sum_{t=1}^{T} \mathbf{r}_t - \hat{\mathbf{r}}_{1:N} + \hat{\mathbf{B}}_{MLC} \hat{\lambda}_{MLC}, \tag{2.1.29}
\]

\[
\hat{\Sigma}_{C_\eta,MLC} = (T - 1 + \delta)^{-1} [(T - 1)(\hat{\Sigma}_g - \hat{\Sigma}_Z) + d \hat{\Sigma}_{C_\eta}], \tag{2.1.30}
\]

\[
\hat{\lambda}_{MLC} = T^{-1} \sum_{t=1}^{T} \mathbf{f}_t - \mathbf{r}_j \mathbf{I}_K, \tag{2.1.31}
\]

\[
\hat{\Omega}_{MLC} = \mathbf{P}_{KK} (\hat{\mathbf{G}} - \mathbf{I}_K) \mathbf{P}_{KK}^{-1}, \tag{2.1.32}
\]

setting \( \hat{\Sigma}_g = \left( \begin{array}{cc} \hat{\mathbf{B}}_{MLC}' \hat{\mathbf{G}}_{MLC} \hat{\mathbf{B}}_{MLC} & \hat{\mathbf{B}}_{MLC} \hat{\mathbf{I}}_K \\ \hat{\mathbf{I}}_K & \hat{\mathbf{G}}_{MLC} \hat{\mathbf{B}}_{MLC} \end{array} \right) \) and \( \mathbf{P} = \hat{\Sigma}_{C_\eta}^{-1} \hat{\mathbf{V}} = (\mathbf{P}_{NK}, \mathbf{P}_{KK})'. \) Note that the upper-left and the lower-right blocks of \( \hat{\Sigma}_{C_\eta,MLC} \) provide the estimators \( \hat{\mathbf{C}}_{MLC} \) and \( \hat{\Sigma}_{\eta,MLC}, \) respectively.

On the other hand, if \( \hat{\gamma}_{K,N+K} \leq 1, \) then \( \mathbf{P}_{KK} \) is singular and \( \hat{\mathbf{B}}_{MLC} \) is not well-defined any longer.

**Remark 2.6** The unobserved factors \( \mathbf{f}_t, \) which determine the asset returns, can be estimated as

\[
\hat{\mathbf{f}}_t = \left( T^{-1} \sum_{t=1}^{T} \mathbf{f}_t - \hat{\mu}_{\eta,MLC} \right) + \hat{\mathbf{H}} (\mathbf{Z}_t - T^{-1} \sum_{s=1}^{T} \mathbf{Z}_t),
\]

setting \( \hat{\mathbf{H}} = \left( \mathbf{0}, \mathbf{I}_K - \hat{\Sigma}_{\eta,\eta}^{-1} \mathbf{I}_N - \hat{\mathbf{B}}_{MLC} \right) \) with \( \hat{\Sigma}_{\eta,\eta} = \left( \mathbf{I}_N - \hat{\mathbf{B}}_{MLC} \right) \hat{\Sigma}_{C_\eta,MLC} \left( \mathbf{I}_N - \hat{\mathbf{B}}_{MLC} \right)' + \hat{\Sigma}_{\eta,\eta} = \hat{\Sigma}_{\eta,MLC} \hat{\mathbf{B}}_{MLC} \).\(^{49}\)

\(^{49}\)Distributional assumptions (nearly distributed \( \mathbf{f}_t, \mathbf{e}_t, \mathbf{n}_t \)'s and Wishart \( \hat{\Sigma}_{C_\eta} \) are not required for consistency nor for asymptotic normality although the latter are often made when inference for all parameters is desired. Alternative conditions are bounded fourth moment of \( \mathbf{f}_t, \mathbf{e}_t, \mathbf{n}_t \)'s and asymptotic normality of \( d^2 \left( \hat{\Sigma}_{C_\eta} - \Sigma_{C_{\eta_0}} \right); \) see Amemiya and Fuller (1984, comments to Theorem 3, p 507).
Remark 2.7 For portfolio construction purposes, having estimated the model's parameters, we obtain the mean-variance weights as

$$\hat{\mathbf{w}}_{\mathbf{N}} = \frac{1}{\gamma} \left( \mathbf{B}_{N,MLC} \hat{\mathbf{C}}_{N,MLC} \right)^{-1} \left( -\mathbf{B}_{N,MLC} \hat{\mathbf{C}}_{N,MLC} + \mathbf{B}_{N,MLC} \hat{\mathbf{C}}_{0,MLC} \right).$$

Remark 2.8 As an alternative approach to the one described in the theorem above, one could estimate $\mathbf{a}_0$, $\mathbf{B}_0$, and the residual variance, $\Sigma_0 = \mathbf{B}_0 \Sigma_{\text{error}} \mathbf{B}_0^T + \mathbf{C}_0$, using an instrumental-variable-type estimator, because of the correlation between the mis-measured factors, $f_t$ and the residual, $\left( \mathbf{z}_t - \mathbf{B}_0 (\eta_t - \mu_{\text{me}}) \right)$. However, doing so would not allow one to exploit the fact that $\mathbf{B}_0$ appears both in the intercept term of expected returns and in the residual-covariance matrix, implying a loss of precision in the estimates. Note that the MLEs derived above require the preliminary estimate $\hat{\Sigma}_0$ of $\Sigma_{\text{error}}$, which can be obtained by the aforementioned instrumental-variable estimator, by means of an independent sample of observations.

Remark 2.9 For expositional simplicity, we assumed above that all observed factors were measured with error. In practice, it is possible that some factors are measured without error and, in fact, the above estimation procedure holds also when $\Sigma_0$ is singular; see Fuller (1987, Section 4.1.1, p. 298).\footnote{Anderson (1984) provides a test for the null hypothesis that the rank of $\Sigma_{\text{error}}$ is less or equal to $K$ against the alternative that this rank is greater than $K$, without requiring one to identify which factor is affected by measurement error.}

Remark 2.10 In practice, all the various forms of misspecification discussed in Section 2.3.5 are likely to arise at the same time. Therefore, expected excess returns (conditional on the factors) are $\mathbf{\mu}_0 - \gamma_1 \mathbf{1}_N = \mathbf{a}_0 + \mathbf{A}_0 \lambda_{\text{miss}} - \mathbf{B}_{K,0} \mathbf{\mu}_{K,0} + \mathbf{B}_0 \lambda_0$, and the residual covariance matrix is $\Sigma_0 = \mathbf{C}_0 + \mathbf{B}_{K,0} \Sigma_{K,0} \mathbf{B}_{K,0}^T + \mathbf{A}_0 \mathbf{A}_0^T$. Using the same arguments as the ones for Theorem 2.11, the APT restriction allows one to separately identify $\mathbf{A}_0$ from $\lambda_{\text{miss}}$, and $\mathbf{\mu}_{K,0}$. Using a procedure similar to the one described in the remarks above, one can then identify the number of missing and/or mis-measured factors and then estimate the full model using ML that takes into account the various restrictions arising from misspecification.
Chapter 3

Detecting Spurious Factors using Cross-Sectional Regressions

Valentina Raponi, Imperial College London
Paolo Zaffaroni, Imperial College London
Abstract

This paper investigates the large cross-sectional properties of the standard two-pass methodology, when useless factors are included in the beta-pricing specification. When the number of time-series observations, $T$, is assumed to be fixed, and contrary to the conventional large-$T$ framework, we find that the simple two-pass OLS estimator of risk premia exhibits desirable asymptotic properties that can be used to detect useless factors. In particular, we derived correctly-sized $t$-ratios, $F$-tests and goodness-of-fit measures that allow us to implement a powerful statistical strategy to test for factors that can be potentially irrelevant for the analysis. The results hold also under the assumption of potential model misspecification. The validity of our results is assessed by means of simulation exercises.

Keywords: beta-pricing models; useless factors; two-pass cross-sectional regressions; risk premia; model misspecification; large-$N$ asymptotics.

JEL classification: C12, C13, G12.
3.1. Introduction

Linear factor models, thanks to their simplicity and ease of interpretation, can be considered as a reference point in most of financial empirical works. The search for risk factors that improve the pricing performance of various asset-pricing models has generated a growing literature in empirical asset pricing, see e.g., Lintner (1965), Fama and MacBeth (1973), Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Parker and Julliard (2005), Jagannathan and Wang (2007).

Despite their simplicity, however, such models have produced very puzzling results in the empirical applications, with a list of more than 300 factors claimed to be relevant in explaining the cross-sectional variation of stock returns (Harvey et al. (2016)). This apparent significance of such wide range of risk factors can be attributed to a lack of model identification, which could be caused by the inclusion of factors that only weakly correlate (or do not correlate at all) with asset returns (see e.g., Jagannathan and Wang (1998), Kan and Zhang (1999b) and Kan and Zhang (1999a)). When this situation occurs, we say that such factors are “useless” or “spurious”, in the sense that they are independent of all the asset returns used in the model (Kan and Zhang (1999b)). As a result, all the inference on risk premia parameters could be totally invalid.

The main goal of this paper is to provide a simple, yet formal, statistical methodology that bypasses such identification issues and allows us to conduct correctly-sized inference on beta-pricing models by detecting factors that can be potentially irrelevant for the analysis.

When testing linear asset pricing models, one of the primary questions of interest is whether the beta risk of a particular factor is priced, i.e., whether the estimated risk premium associated with a given factor is significantly different from zero. Black (1972) and Fama and MacBeth (1973) develop a two-pass methodology in which the beta of each asset with respect to a factor is estimated in a first-pass time series regression; the estimated betas are then used in a second-pass cross-sectional regression (CSR) to estimate the risk premium associated with that factor.

The properties of the test statistics and goodness-of-fit measures under the two-pass methodology are usually developed under the assumptions that the model is correctly specified and the factors are correctly identified (Shanken (1992)), using the standard setting of a large time-series length, $T$, and a fixed number of asset returns, $N$.

However, statistical inference can become very problematic if the pricing model includes useless factors (Kan and Zhang (1999b), Kleibergen (2009), Gozpodinov et al. (2014)). In this case, all the risk premia parameters are no longer strongly identified and standard inference techniques become unreliable (see e.g., Jagannathan and Wang (1998), Kan and Zhang (1999b), Kleibergen (2009), Kleibergen and Zhan (2015), Burnside (2015), Gozpodinov et al. (2014), Bryzgalova (2016a), Gozpodinov et al. (2017), Gozpodinov et al. (2018)). The reason is that the true betas of the assets with respect to the useless factor are all zero, and hence the “true” risk premium associated to that useless factor is no longer identified in the second pass. That is, “as the estimated betas approach zero, the absolute value of the estimated risk premium needs to go to infinity, instead of zero, in order to “explain” the cross-sectional difference in the expected returns.” (Kan and Zhang (1999b)). As a result, two-pass CSR OLS estimators, $t$-tests, $R^2$-squared and standard inference procedures become inadequate and identification failure could lead to the conclusion that such factor is important, even though totally spurious by nature. Even more seriously, under standard estimation procedures (such as maximum likelihood and one-step generalized method of moments), factors that are spurious are found to be selected with high probability, while factors that are useful are driven out of the model (see, e.g., Gozpodinov et al. (2018)). Moreover, when the number of asset becomes large, spuriousness of risk premia estimates is further aggravated (see Kleibergen (2009)).

Last but not least, the problem of potential model misspecification should be also taken into consideration when evaluating asset pricing models. Indeed, it is now widely documented that misspecification is a common feature of many asset-pricing models, and reliable statistical inference crucially depends on its robustness to potential model misspecification.

Misspecification can arise for several reasons. A linear beta-pricing model can be incorrectly specified when expected returns are not exactly linear in the factor betas. Even when expected returns are linear in their factor betas, the econometric specification could be wrong if some of the factors are missing (Pástor
(2000)) or mismeasured (Roll (1977)). All these cases are compatible with the extended Asset Pricing Theory (APT) of Raponi et al. (2018).

Kan and Zhang (1999b), Gospodinov et al. (2014), among others, show that by ignoring model misspecification, one can erroneously conclude that a risk factor is priced while, in fact, it does not contribute at all to the pricing ability of the model. Moreover, statistical inference under potential model misspecification can become even more problematic if the pricing model includes spurious factors (Kleibergen (2009)).

Therefore, given the importance of the two-pass methodology in testing asset pricing models and the potential problem of misspecifying factors, a relevant question is how we can detect useless factors in the two-pass methodology. This is the main objective of our paper.

Notice that all the above results are established under the standard setting of a large time-series length, \( T \), and a fixed number of asset returns, \( N \). In this context, several methods have been proposed in the literature to tackle the effect of useless factors. Kleibergen (2009) proposes alternative statistics based on the maximum likelihood estimator of Gibbons (1982), whose large sample distributions remain unaltered when the betas are close to zero and \( N \) increases. Gospodinov et al. (2014) introduce a sequential model selection device based on robust t-statistics, where individual factors are dropped sequentially. Burnside (2015) uses model diagnostics based on Kleibergen and Paap (2006)'s rank test. Bryzgalova (2016a) proposes a shrinkage-based framework for eliminating individual (or multiple) spurious factors, where estimation and model reduction is done in one step, rather than sequentially. Using an asymptotic framework where the number of assets/portfolios grows with the time span of the data, Anatolyev and Mikusheva (2018) introduce a new estimation procedure based on sample-splitting instrumental variables regression.

In this paper we want to provide a methodology to test for useless factors in the context of beta-pricing models, which is valid when \( T \) is fixed (and possibly very small), while the number of test assets \( N \) becomes large. As explained in Raponi et al. (2019), this sampling scheme is particularly relevant nowadays, when thousands of stocks are traded every day, but only short time-series are used in practice, to mitigate issues of structural breaks and to avoid modelling time-varying parameters explicitly. In their paper, Raponi et al. (2019) show that, when \( N \) is large and \( T \) is fixed, the standard two-pass CSR OLS can not be applied, as it leads to a first order bias problem. Based on the bias-adjusted Shanken estimator (Shanken (1992)), they provide a formal methodology for estimating risk premia and testing asset-pricing models under both the cases of correctly specified and misspecified models. However, the case of useless factors is ruled out from their analysis because, as we show below, in this case the Shanken estimator does not exhibit standard properties and, therefore, cannot be used.

What works in this case, instead, is the simple two-pass CSR OLS estimator. This is one of the main results of our paper. Unlike the classical large-\( T \) framework, we show that the two-pass CSR OLS estimator of the risk premium associated with a useless factor converges to zero as \( N \) goes to infinity and \( T \) is fixed. Interestingly, this result holds regardless of the degree of model misspecification. This allows us to derive correctly-sized \( t \)-test for the null hypothesis that the risk premium of the useless factor is equal to zero, together with valid \( F \)-test and goodness-of-fit measures. Also the risk premia associated to the relevant factors, although biased, can be properly adjusted and consistently estimated.

Our technical assumptions are relatively weak and easily verifiable. In particular, we allow for a very general form of weak cross-sectional dependence among asset returns, following the same assumptions of Raponi et al. (2019).

To give the main intuition of our results, we start our analysis with a very simple illustrative example, where we assume that the model is correctly specified and the true asset-pricing model contains only an asset-specific intercept. We refer to it as the base case. Then, we estimate the model using useless factors and show how the CSR OLS represents a very powerful tool to identify a truly useless factor. We then generalize this simple base case to the case of potential model misspecification and show how inference can be still performed, even when the exact pricing condition is violated (Kan and Zhang (1999b)). We then provide a formal methodology which is valid in the more realistic case in which the true model contains also some useful factors and we still include some irrelevant factors in the model. The CSR OLS estimator can be still applied in this case and inference can be conduct in the same way, even though the formal derivations become more challenging.
However, important issues can arise if some of the relevant factors are omitted from the estimated model. In this case, the risk premia estimator and its asymptotic covariance matrix will be functions of the omitted factors, which of course are not known in practice. This greatly complicates the analysis, rendering inference unfeasible in practice.

We illustrate the finite sample properties of our results using Monte Carlo simulations and show that our procedure is reliable for the time-series and cross-sectional dimensions typically encountered in empirical works.

The rest of the paper is organized as follows. Section 3.2 derives the limiting distribution of the two-pass OLS estimator in the presence of useless factors for the base model, under the assumption of both correct model specification (Section 3.2.1) and model misspecification (Section 3.2.2). Section 3.3 presents the asymptotic results for the more general case in which the true asset-pricing relation specifies a multifactor model with known useful factors. The more complicated case of omitted useful factors is presented in Section 3.4. In Section 3.5, we investigate the sampling performance of our results using simulation exercises. Section 3.6 concludes. Assumptions, lemmas and technical proofs are collected in Appendices 3.A, 3.B, and 3.C, respectively.

### 3.2. Two pass-regression with useless factors when \( N \) is large and \( T \) is fixed

In this section we derive the limiting distribution of the CSR OLS estimator for the simplest case where the true model contains only an asset specific intercept. We refer to this case as the base case. Although this setting could be very unrealistic in empirical applications, we use this case as an illustrative example to give the main intuition of our analysis. We assume that the researcher estimates the model using a useless factor \( (g) \) in the first pass time-series regression and derives the estimates of the betas of the \( N \) assets with respect to the useless factor, without knowing that \( g \) is useless. The objective is to show the consequences of this choice on the second-pass estimator of risk premia, when \( N \) is large and \( T \) is fixed.

In the classical setting of large \( T \), it is now well-documented that the CSR OLS estimator cannot be used when a useless factor is used in testing asset-pricing models (see, e.g., Kan and Zhang (1999b), Kleibergen and Zhan (2015), Gospodinov et al. (2014)). In particular, if we wish to test the null hypothesis that the risk premium associated to such useless factor is equal to zero, one obtains that the probability of rejecting the null hypothesis goes to one, instead of zero, as \( T \) goes to infinity (Kan and Zhang (1999b)).

Instead, when we consider the large-\( N \) and fixed-\( T \) setting, we show that the OLS estimator maintains standard asymptotic properties and, importantly, the risk premium estimator associated to the useless factor converges to zero, as one might expect. This suggests that a correctly-sized test for the null hypothesis that the risk premium of the useless factor is equal to zero can be performed. In particular, we derive the asymptotic distribution of the OLS estimator and show how its covariance matrix can be consistently estimated in order to perform a valid \( t \)-test and derive correct standard errors. We also derive the asymptotic distribution of the \( F \)-test, to test the null hypothesis that the risk premia associated to a set of \( K_g \) useless factors are jointly zero. Finally, we derive the limiting behavior of the \( R \)-squared of the model. Interestingly, in our base setting, we find that the \( R \)-squared of the model converges to zero as \( N \) goes to infinity, suggesting that it can be used as a valid measure of goodness-of-fit. This is in contrast with the standard large-\( T \) results, where the \( R \)-squared of the model does not have a standard distribution in the presence of spurious factors, except for some special cases (Kan and Zhang (1999b)).

In Section 3.2.1 we analyze the behavior of the OLS estimator under the assumption that the model is correctly specified (i.e. the assumption of exact pricing holds). We then generalize this result in Section 3.2.2, where we consider the base model under potential misspecification. In both the cases, we show that the CSR OLS estimator exhibits standard asymptotic properties that allows us to derive correct inference on all the model parameters.
3.2.1 Two-pass regression with useless factors under correct model specification: the base case

For illustrative purpose, let us assume that returns on asset $i$, at time $t$, are generated from the following model

$$R_{it} = \alpha_i + \epsilon_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T$$  \hspace{1cm} (3.1)

where $\alpha_i$ is an asset specific intercept and the $\epsilon_{it}$s are the model residuals assumed to be independent and identically distributed (i.i.d.) over time and such that $E[\epsilon_{it}] = \sigma_i^2$. In matrix notation, we can write the model above as

$$R_t = \alpha + \epsilon_t \quad t = 1, \ldots, T$$  \hspace{1cm} (3.2)

where $\alpha = [\alpha_1, \ldots, \alpha_N]'$, and $\epsilon_t = [\epsilon_{1t}, \ldots, \epsilon_{Nt}]'$. Let $\gamma_0$ be the zero-beta rate and let $1_N$ be an $N$-vector of ones. When the model is correctly specified, then the following assumption of exact pricing holds

$$E[R_t] = \gamma_0 1_N.$$  \hspace{1cm} (3.3)

Let $\tilde{R}_t = \frac{1}{T} \sum_{t=1}^{T} R_{it}$, $\tilde{R}_N = [\tilde{R}_1, \ldots, \tilde{R}_N]'$, and $\bar{\epsilon} = \frac{1}{T} \sum_{t=1}^{T} \epsilon_t$. Averaging (3.2) over time, imposing (3.3), and noting that $E[R_t] = \alpha$, yields

$$\tilde{R} = \gamma_0 1_N + 0_{N \times K} \gamma_1' + \bar{\epsilon}$$  \hspace{1cm} (3.4)

where $\Gamma' = [\gamma_0, \gamma_1']$, with $\gamma_1'$ representing a $K$-vector of ex-post factor risk premia. Moreover, $X = [1_N, 0_{N \times K}]$, with $0_{N \times K}$ denoting an $(N \times K)$ matrix of zeros.

Consider now a $K_g$-vector of useless factors, $g_t = [g_{1t}, \ldots, g_{K_g t}]'$, such that $\text{Cov}(R_{it}, g_t) = 0$, for all $i = 1, \ldots, N$, and let $G = [g_{1t}, \ldots, g_{T t}]'$ be the $T \times K_g$ matrix of useless factors. Now assume that, instead of the true model in (3.1), we estimate the following factor model

$$R_{it} = \bar{\alpha}_i + \bar{\beta}_1 g_{1t} + \cdots + \bar{\beta}_{K_g} g_{K_g t} + \bar{\epsilon}_{it} = \bar{\alpha}_i + \bar{\beta}' g_t + \bar{\epsilon}_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T$$  \hspace{1cm} (3.5)

where $\bar{\alpha}_i$ represents the asset specific intercept, $\bar{\beta}_i = [\bar{\beta}_{1i}, \ldots, \bar{\beta}_{K_g i}]'$ is a vector of multiple regression betas of asset $i$ with respect to the $K_g$ factors $g_t$, and $\bar{\epsilon}_{it}$ is the $i$-th return's idiosyncratic component satisfying the same assumptions of $\epsilon_{it}$ in (3.1). Using matrix notation, we can also write

$$R_t = \bar{\alpha}_i 1_T + G \bar{\beta}_i + \bar{\epsilon}_i \quad i = 1, \ldots, N,$$  \hspace{1cm} (3.6)

where $R_t = [R_{1t}, \ldots, R_{T t}]'$, $1_T$ is a $T$-vector of ones, and $\bar{\epsilon}_i = [\bar{\epsilon}_{1i}, \ldots, \bar{\epsilon}_{Ti}]'$. Defining $\tilde{G} = G - 1_T \bar{g}'$, with $\bar{g} = [\bar{g}_1, \ldots, \bar{g}_{K_g - 1}]' = \frac{1}{T} \sum_{t=1}^{T} g_t$, then the OLS estimator of $\bar{\beta}_i$ is

$$\hat{\beta}_i = 0_{K_g} + (\tilde{G}' \tilde{G})^{-1} \tilde{G}' \bar{\epsilon}_i \quad i = 1, \ldots, N$$  \hspace{1cm} (3.7)

Let $\hat{\beta} = [\hat{\beta}_1, \ldots, \hat{\beta}_N]'$. Then, in matrix form

$$\hat{B}_g = \epsilon' P_g,$$  \hspace{1cm} (3.8)

where $\epsilon = [\epsilon_1, \ldots, \epsilon_N]$ is a $(T \times N)$ matrix and $P_g = \tilde{G}(\tilde{G}' \tilde{G})^{-1}$. We now run a single CSR of the sample mean vector $\bar{R}$ on $\bar{X}_g = [1_N, \hat{B}_g]$ to estimate $\Gamma'$ in the second pass. That is

$$\bar{R} = \bar{X}_g \Gamma' + \eta'_g,$$  \hspace{1cm} (3.9)
3.2. TWO PASS-REGRESSION WITH USELESS FACTORS WHEN N IS LARGE AND T IS FIXED

where \( \eta \eta' = \left( \bar{\varepsilon} - (\bar{X}_{g} - X)' \Gamma' \right) \) and \( X = [1_N, 0_{N \times K}] \). If we use the identity matrix as the weighting matrix in the second-pass CSR, we obtain the following OLS estimator for the feasible representation in (3.9):

\[
\hat{\Gamma}_g = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \end{bmatrix} = (\bar{X}_{g}' \bar{X}_g)^{-1} \bar{X}_g' \bar{R}.
\]

(3.10)

The Shanken (1992) estimator, instead, has the following form

\[
\hat{\Gamma}^*_g = \left( \frac{\bar{X}_g' \bar{X}_g}{N} - \hat{\Lambda}_g \right)^{-1} \frac{\bar{X}_g' \bar{R}}{N}, \quad \text{with} \quad \hat{\Lambda}_g = \begin{bmatrix} 0 & 0_{K_g} \\ 0_{K_g} & \hat{\sigma}^2 (\hat{G}' \hat{G})^{-1} \end{bmatrix},
\]

\[
\hat{\sigma}^2 = \frac{1}{N(T - K_g - 1)} \text{tr}(\hat{\varepsilon}' \hat{\varepsilon}),
\]

(3.11)

(3.12)

where \( \text{tr}(\cdot) \) denotes the trace operator, and \( \hat{\varepsilon} \) is the \((T \times N)\) matrix of first-pass residuals.

In the large-\( N \) and fixed-\( T \) setting with no useless factors, Raponi et al. (2019) show that the CSR OLS estimator (\( \hat{\Gamma} \)) is biased and inconsistent and propose to use the Shanken estimator (\( \hat{\Gamma}^*_g \)) to restore consistency and conduct valid inference on risk premia parameters. Given this result, one might ask whether the Shanken estimator still possesses desirable properties even when useless factors are used in the regression. Unfortunately, the answer is negative. We show in the next proposition that, in the presence of useless factors, the risk premia estimator corresponding to the useless factors does not converge to zero, but rather to a (non standard) random variable, making all the inference unreliable.

**Proposition 3.1** Assume that the representation in (3.1) holds and that (3.3) is satisfied. Then, under Assumptions 3.1-3.5,

\[
\hat{\Gamma}^*_g \rightarrow_d \begin{bmatrix} \gamma_0 \\ Z^{-1} H \end{bmatrix}
\]

where \( Z \) is \( K_g \times K_g \) matrix of Normal random variables such that

\[
\sqrt{N} D = \sqrt{N} P_g (\hat{\varepsilon}' - \hat{\sigma}^2 I_T) P_g N \rightarrow_d Z.
\]

and where \( H \) is a normally distributed \( K_g \)-vector such that

\[
\begin{bmatrix} \frac{1}{T} \otimes P_g' \sqrt{N} \text{vec} \left( \frac{\hat{\varepsilon}' - \hat{\sigma}^2 I_T}{N} \right) \rightarrow_d N \left( 0_{K_g}, 0 \left( \frac{1}{T} \otimes P_g' \right) U_r \left( \frac{1}{T} \otimes P_g \right) \right) \end{bmatrix} = H
\]

with \( D = \frac{P_g (\hat{\varepsilon}' - \hat{\sigma}^2 I_T) P_g}{N} - \frac{P_g \otimes P_g}{N^2} \).

**Proof:** See Appendix 3.C.

The result of Proposition 3.1 makes the Shanken estimator unusable in the presence of useless factors, as the estimated risk premia of the useless factors converges to random values. As we show in the proof of the Proposition 3.1 (see Appendix 3.C), such randomness comes from the term that contains the bias adjustment in the estimator in (3.11). Removing the adjustment gives us the CSR OLS estimator in (3.10). Therefore, a natural question concerns how the OLS estimator would behave in this case, when \( T \) is fixed and \( N \) is large, and the model uses spurious factors in the estimation. What we find is that the OLS estimator represents a very powerful tool to dissect useless factors in a large-\( N \) environment. Indeed, the risk premium OLS estimator of the useless factors is found to be consistent, asymptotically normally distributed, and converges to zero as \( N \) goes to infinity. We establish the limiting distribution of the CSR OLS estimator in (3.10) in the following theorem.
Theorem 3.1 Assume that the representation in (3.1) holds and that (3.3) is satisfied

(i) Under Assumptions 3.1-3.4,

\[ \hat{\gamma}_g - \begin{bmatrix} \gamma_0 \\ 0_K \end{bmatrix} = O_p \left( \frac{1}{\sqrt{N}} \right) . \]

(ii) Under Assumptions 3.1-3.5,

\[ \sqrt{N} \left( \hat{\gamma}_g - \begin{bmatrix} \gamma_0 \\ 0_K \end{bmatrix} \right) \rightarrow_d N(0_{K+1}, V) \]

where

\[ V = \begin{bmatrix} \frac{\hat{\sigma}^2}{T} & 0_K \\ 0_K & \frac{1}{T} C' \hat{U}_C C \end{bmatrix}, \quad \text{with} \quad C = \left( \frac{1}{T} \otimes \hat{G} \right), \]

and where \( U_c = \lim \frac{1}{N} \sum_{i=1}^N E \left[ \text{vec}(\epsilon_i' \epsilon_i' - \sigma_\epsilon^2 I_T) \text{vec}(\epsilon_i' \epsilon_i' - \sigma_\epsilon^2 I_T) \right] \).

Proof: See Appendix 3.C.

Theorem 3.1 shows that the two-pass OLS estimator for the useless factor \( \hat{\gamma}_1 \) converges to zero as \( N \) goes to infinity and \( T \) is fixed. The OLS estimator for the zero-beta rate is also consistently estimated and converges to the true value \( \gamma_0 \). Notice also that, when a factor is quasi-useless in the sense of Kleibergen (2009) (that is, its corresponding beta is of the form \( \beta_0 = c_i / \sqrt{T} \)) it will not become useless asymptotically, since \( T \) is fixed.

The asymptotic covariance matrix of the OLS estimator need to be consistently estimated in order to conduct statistical inference. The results are summarized in the next theorem.

Theorem 3.2 Under Assumptions 3.2-3.4 and using the identification condition \( \kappa_4 = 0 \), as \( N \to \infty \), we have

\[ \hat{V} \rightarrow_p V \] (3.13)

where

\[ \hat{V} = \begin{bmatrix} \frac{\hat{\sigma}^2}{T} & 0_K \\ 0_K & \frac{1}{T} \tilde{C}' \hat{U}_C C \end{bmatrix}, \quad \text{with} \quad C = \left( \frac{1}{T} \otimes \hat{G} \right). \] (3.14)

where \( \hat{\sigma}^2 \) is defined in (3.12), \( \hat{U}_c \) is a consistent estimator of \( U_c \) (see Raponi et al. (2019)), obtained by replacing \( \sigma_4 \) with

\[ \hat{\sigma}_4 = \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \hat{\epsilon}_i^4 \]

\[ \text{with } M^{(2)} = M \oplus M, \text{ with } M = I_T - D(D' D)^{-1} D', \quad D = [1_T, G] \text{ and where } \oplus \text{ denotes the Hadamard product operator.} \]

Proof: See Appendix 3.C.

The above result suggests that a correctly-sized Wald test for the null hypothesis \( H_0 : \gamma_1 = 0_{K_0} \) can be used and proper goodness-of-fit measures can be derived. In particular, we show that the \( t \)-test, \( F \)-test and \( R \)-squared maintain traditional distributions when \( N \to \infty \) and \( T \) is fixed.
3.2. TWO PASS-REGRESSION WITH USELESS FACTORS WHEN \( N \) IS LARGE AND \( T \) IS FIXED\(^{169} \)

Let \( \hat{\epsilon}_g = R - \hat{X}_g \hat{\beta}_g \) denote the vector of OLS residuals from the second pass in (3.9) and define

\[
s^2 = \frac{\hat{\epsilon}_g' \hat{\epsilon}_g}{N - K_g - 1}.
\]

Let

\[
t_k = \frac{\hat{\gamma}_k}{s \cdot \sqrt{c_{kk}}}, \quad 2 \leq k \leq K_g + 1
\]

be the \( t \) statistic for the \( k \)-th regression coefficient in (3.9), where \( c_{kk} \) denotes the \( (k,k) \)-th element of the matrix \((\hat{X}_g' \hat{X}_g)^{-1}\). Moreover, let

\[
R_{CSR}^2 = 1 - \frac{\hat{\epsilon}_g' \hat{\epsilon}_g}{R' \hat{M} \hat{R}}
\]

be \( R \)-squared of the model in (3.9), where we define \( \hat{M} = I_N - \frac{1}{N} \hat{X}_g \hat{X}_g' \). Finally, let

\[
F_{CSR} = \frac{R_{CSR}^2}{(1 - R_{CSR}^2)/(N - K_g - 1)}
\]

be the \( F \)-statistic to test whether all the \( K_g \) coefficients (except for the intercept) are zero in (3.9). Then, under the representation in (3.1), and assuming correct model specification, we can establish the following theorem.

**Theorem 3.3** Assume that the representation in (3.1) holds and that (3.3) is satisfied. Then, under Assumptions 3.1-3.5 we have

(i)

\[
t_{g,k} \rightarrow_d \left( 0, \frac{\kappa_0 + \sigma_0}{\sigma^2} \right)
\]

(ii)

\[
R_{CSR}^2 \rightarrow 0
\]

(iii)

\[
F_{CSR} \overset{d}{\rightarrow} \chi^2_K \left( \frac{\kappa_0 + \sigma_0}{\sigma^2} \right)/K
\]

**Proof:** See Appendix 3.C.

Notice that, if we also impose that \( \sigma_0 = \sigma^2 \), together with \( \kappa_0 = 0 \), the results in Theorem 3.3 reduce to traditional limiting distributions, that is \( t_k \rightarrow_d N(0,1) \) and \( F_{CSR} \rightarrow_d \chi^2_{K_g}/K_g \).

Theorem 3.3 shows that the presence of any useless factor in the estimated asset-pricing model can be properly detected using standard \( t \) and \( F \) tests. If one also admits a certain degree of weak cross-sectional dependence among asset returns, then the \( t \) and \( F \) statistics have non conventional distributions that depend on the nuisance parameters \( \kappa_0 \) and \( \sigma_0 \). These quantities can be consistently estimated using the same results and assumptions of Rapponi et al. (2019), making all the inference still feasible in practice. Notice also that the \( R \)-squared of the model is not inflated in any way by the presence of useless factors, and it converges to zero as expected. This suggests that this measure can be still used as a measure of goodness-of-fit, unlike the standard large-\( T \) setting, where the sample \( R \)-squared tends to one as \( T \) tends to infinity (Kan and Zhang (1999a)).
3.2.2 The base case under potential model misspecification

In this section we consider the case of potential model misspecification. In particular, we assume that the relation in (3.3) is not satisfied anymore and expected returns can now vary across assets (see Kan and Zhang (1999b)):

$$E[R_{it}] = c_i, \quad i = 1, \ldots, N.$$  \hfill (3.20)

The assumption in (3.20) implies that the pricing errors, $e_i$, are now such that

$$E[R_{it}] = \gamma_0 + c_i \equiv c_i.$$  \hfill (3.21)

This implies that, under (3.20),

$$\hat{R}_t = c + 0_{N \times K} \gamma_1^p + \xi$$  \hfill (3.22)

where $\gamma_1^p = \gamma_1 + \bar{y} - E[y]$ and $c = (c_1, \ldots, c_N)$. Let us consider again the simplest case in (3.1) and assume that, instead of (3.1), we estimate the model in (3.6) using OLS. Then, the following results can be established.

**Theorem 3.4** Assume that the representation in (3.1) holds and that (3.20) is satisfied. Then,

(i) Under Assumptions 3.1-3.4 and Assumption 3.6

$$\hat{\mu}_c \sim \left[ \begin{array}{c} \mu_c \\ 0_{K_s} \end{array} \right] = O_p \left( \frac{1}{\sqrt{N}} \right)$$  \hfill (3.23)

where $\mu_c = \text{plim} N^{-1} \sum_{i=1}^N c_i$.

(ii) Under Assumptions 3.1-3.6

$$\sqrt{N} \left( \hat{\mu} - \left[ \begin{array}{c} \mu_c \\ 0_{K_s} \end{array} \right] \right) \rightarrow_d N \left( 0_{K_s+1}, V + W \right)$$  \hfill (3.24)

where $W = \frac{\nu_c}{\nu} \bar{G} \hat{G}$, with $\nu_c = \text{plim} \frac{1}{N} \sum_{i=1}^N c_i^2 a_i^2$ and where $a_{ij}$ denotes the $(i,j)$-th element of the matrix $A_e = M_{1N} c c' M_{1N}$, with $M_{1N} = I_N - \frac{1}{N} 1_{N} 1_{N}'$.

**Proof:** See Appendix 3.C.

Theorem 3.4 shows that, when $T$ is fixed, the OLS estimator of the risk premia associated to the useless factors, $g_i$, still converges to zero at the rate of $O_p(1/\sqrt{N})$ even when the model is potentially misspecified. When $c = \gamma_0 1_N$, then Theorem 3.4 reduces to the case of correctly specified models.

The asymptotic covariance matrix in (3.24) can be consistently estimated using $\hat{V}$ as derived in Theorem 3.2 and where we use $\hat{\mu}_c = \sum_{t=1}^T \sum_{i=1}^N R_{it}/NT$ and $\hat{\nu}_c$ (see Raponi et al. (2019), Lemma 9) as consistent estimators of $\mu_c$ and $\nu_c$, respectively.

In order to conduct inference on the null hypothesis $H_0 : \gamma_1 = 0_{K_s}$ and adopt proper goodness-of-fit measures, we can now generalize Theorem 3.3 to the case of model misspecification. Let us define the t-statistic for the k-th regressor coefficient in the model (3.9) as in (3.17), with $s^2$ defined in (3.16). Let $R^2_{DSS}$ be the R-squared of the model in (3.9) as defined in (3.18) and define the $F$-statistic to test whether all the $K_s$ coefficients (except for the intercept) are jointly zero as in (3.19). Then, under the assumption in (3.20), we can establish the following theorem.
Theorem 3.5 Assume that the representation in (3.1) holds and that (3.20) is satisfied. Then, under Assumptions 3.1-3.6, we have

(i)

\[ t_{g,k} \rightarrow_d \left( 0, \frac{\nu_c + \frac{\kappa_g + T \sigma^2}{T + \nu_c}}{\nu_c + \frac{\sigma^2}{T}} \right) \]

(ii)

\[ R^2_{CSR} \rightarrow 0 \]

(iii)

\[ F_{CSR} \rightarrow \chi^2_K \left( \frac{\bar{\gamma}_0 + \sigma^4}{\sigma^4} \right) / K_g \]

The proof of this theorem is available upon request.

The results of Theorem 3.5 show how inference can be robustified under potential model misspecification. Notice that, these results are not qualitatively different from the correctly-specified case and, when \( \nu_c = 0 \), Theorem 3.5 reduces to Theorem 3.3. Moreover, since misspecification of asset-pricing models is now well-documented in practice, it would always worth computing standard errors which are robust to potential model misspecification.

3.3. Including useless factors in a two-pass regression with known useful factors

In the previous section we analyzed the simplest case where the true model in (3.1) did not contain any relevant factor to explain the cross-sectional variation of stock returns. Even though this situation is quite unrealistic in practice, we gave the intuition of why the CRS OLS estimator of risk premia can be adopted when \( N \) is large and \( T \) is fixed: the risk premia estimator associated to the useless factor converges to zero, and also the estimator of \( \gamma_0 \) converges to its true value. The estimator is also asymptotically normally distributed and its corresponding standard error can be easily derived. Therefore, we are able to test whether a factor is useless or not in this simple base case.

Given the extreme simplification adopted in model (3.1), one might reasonably ask what are the consequences of including useless factors \( g_t = [g_{1t}, \ldots, g_{Kf} \gamma]^\prime \), when the true asset-pricing relation specifies a multi-factor model with a set of \( K_f \) useful factors \( f_t = [f_{1t}, \ldots, f_{Kf} \gamma]^\prime \). Formally, let us assume that asset returns are generated by the following model:

\[ R_t = \alpha + B_f f_t + \epsilon_t, \quad t = 1, \ldots, T \]  \hspace{1cm} (3.25)

where \( B_f = [B_{f1}, \ldots, B_{fKf}] \) is an \((N \times K_f)\) matrix of regression coefficients. As before, without knowing that \( g_t \) is useless, the researcher estimates the betas of the \( N \) assets with respect to both the useful and useless factors in the first-pass time series regression. That is,

\[ R_t = \tilde{\alpha} + \tilde{B}_f f_t + \tilde{B}_g g_t + \tilde{\epsilon}_t, \quad t = 1, \ldots, T \]  \hspace{1cm} (3.26)
where we define \( \tilde{\alpha} = [\tilde{\alpha}_1, \ldots, \tilde{\alpha}_N]' \), \( \tilde{B}_f = [\tilde{B}_{f1}, \ldots, \tilde{B}_{fN}]' \), with \( \tilde{\beta}_{fi} = [\tilde{\beta}_{fi1}, \ldots, \tilde{\beta}_{fiK_f}]' \), \( \tilde{B}_g = [\tilde{B}_{g1}, \ldots, \tilde{B}_{gN}]' \), with \( \tilde{\beta}_{gi} = [\tilde{\beta}_{gi1}, \ldots, \tilde{\beta}_{giK_g}]' \), and \( \tilde{\epsilon}_i = [\tilde{\epsilon}_{i1}, \ldots, \tilde{\epsilon}_{iN}] \).

Notice that two possible situations could actually occur in this scenario. In the first case, together with \( g_i \), the researcher is including all the relevant \( K_f \) factors in the model, i.e., the model is correctly specified and does not suffer from any omitted variable problem. In a second case, instead, the researcher does not observe all the \( K_f \) usefull factors and includes only a subset \( K_{f1} \) of relevant factors, i.e., he is omitting \( K_{f2} \) usefull factors, such that \( K_f = K_{f1} + K_{f2} \). In this section we focus on the first case. The second situation is analyzed in the next section.

Define \( \tilde{X}_f = [\tilde{\beta}_{f1}, \ldots, \tilde{\beta}_{fN}]' \), \( \tilde{B}_g = [\tilde{\beta}_{g1}, \ldots, \tilde{\beta}_{gN}]' \), and let \( \tilde{X}_{fg} = [1_N, \tilde{B}_f, \tilde{B}_g] \) be an \( (N \times K_f + K_g + 1) \) matrix. Then, let us estimate the vector of risk premia in the second pass by running the following CSR

\[
\tilde{R} = \tilde{X}_{fg} \Gamma^P + \eta_{fg}, \tag{3.27}
\]

where \( \eta_{fg} = (\tilde{X}_{fg} - X_{fg}) \Gamma^P \) and \( X_{fg} = [1_N, B_f, 0_{N \times K_g}] \). The OLS estimator of \( \Gamma^P \) in (3.27) is therefore

\[
\hat{\Gamma}_{fg} = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_{i1} \\ \hat{\gamma}_{i2} \end{bmatrix} = (\tilde{X}_{fg}' \tilde{X}_{fg})^{-1} \tilde{X}_{fg}' \tilde{R}. \tag{3.28}
\]

We can now derive the asymptotic properties of the estimator in (3.28) as \( N \rightarrow \infty \) and \( T \) is fixed. Let \( F = [f_1, \ldots, f_T]' \) be a \( (T \times K_f) \) matrix of useful factors and let \( G = [g_1, \ldots, g_T]' \) be a \( (T \times K_g) \) matrix of useless factors. For simplicity, from now on, we will assume that \( F \) and \( G \) are orthogonal to each other, that is

\[
F'M_T G = 0_{K_f \times K_g} \tag{3.29}
\]

where \( M_T = I_T - \frac{1}{T} J_T \). We show later how the assumption in (3.29) can be relaxed, at the price of more complicated derivations and testing procedures. We also impose that the assumption in (3.3) of exact pricing holds, i.e. the model is correctly specified.

**Theorem 3.6** Assume that the representation in (3.25) holds and that (3.3) is satisfied. Then, the CSR OLS estimator in (3.28) under (3.29) satisfies:

(i) Under Assumptions 3.1-3.4

\[
\sqrt{N} \left( \hat{\Gamma}_{fg} - \begin{bmatrix} \gamma_0 + d_0 \\ 0_{K_g} \end{bmatrix} \right) \rightarrow_d \left( 0_{(K_f + K_g + 1)}, (\Sigma_{X_{fg}} + \Lambda_{fg})^{-1} (V_{fg} + W_{fg}) (\Sigma_{X_{fg}} + \Lambda_{fg})^{-1} \right) \tag{3.30}
\]

where \( d_0 = \sigma^2 \mu_{\beta_f} D^{-1} (F'M_T F)^{-1} \gamma_{i1}^T, d_1 = -\sigma^2 D^{-1} (F'M_T F)^{-1} \gamma_{i2}^T \) and \( D = \Sigma_{\beta_f} - \mu_{\beta_f} \mu_{\beta_f}' + \sigma^2 (F'M_T F)^{-1} \).

(ii) Under Assumptions 3.1-3.5

\[
\sqrt{N} \left( \hat{\Gamma}_{fg} - \begin{bmatrix} \gamma_0 + d_0 \\ \gamma_{i1}^T + d_1 \end{bmatrix} \right) \rightarrow_d \left( 0_{(K_f + K_g + 1)}, (\Sigma_{X_{fg}} + \Lambda_{fg})^{-1} (V_{fg} + W_{fg}) (\Sigma_{X_{fg}} + \Lambda_{fg})^{-1} \right) \tag{3.31}
\]
where
\[
\Sigma_{X_{fg}} = \begin{bmatrix}
\Sigma_{X_f} & 0_{(K_f-1) \times K_g} \\
0_{K_g \times (K_f+1)} & 0_{K_g \times K_g}
\end{bmatrix}, \quad \text{with} \quad \Sigma_{X_f} = \begin{bmatrix}
1 & \mu_{\beta_f} \\
\mu_{\beta_f} & \Sigma_{\beta_f}
\end{bmatrix},
\]
\[
\Lambda_{fg} = \begin{bmatrix}
0 & 0_{K_f} \\
0_{K_g} & \sigma^2 (F'P'F)^{-1}
\end{bmatrix},
\]
\[
V_{fg} = \sigma^2 \left( \frac{1}{T} + \langle \gamma_{1_f} + d_1 \rangle' (F'P'F)^{-1} (\gamma_{1_f} + d_1) \right) \Sigma_{X_{fg}} + \sigma^2 \Omega,
\]
\[
\Omega = \begin{bmatrix}
0 & 0_{K_f} \\
0_{K_g} & d_1' \Sigma_{\beta_f} d_1 (F'F)^{-1} - \Sigma_{\beta_f} d_1' Q_f P_f - P_f' Q_f d_1' \Sigma_{\beta_f} - 0_{K_f \times K_g} \\
0_{K_g} & 0_{K_g \times K_f} \\
0_{K_g} & d_1' \Sigma_{\beta_f} d_1 (G'G)^{-1}
\end{bmatrix},
\]
and
\[
W_{fg} = \begin{bmatrix}
0 & 0_{K_f} \\
0_{K_g} & (Q_f' \otimes P_f') U_c (Q_f' \otimes P_f) \\
0_{K_g} & (Q_g' \otimes P_g') U_c (Q_g' \otimes P_g)
\end{bmatrix},
\]
with the $T$-vector $Q_f = \frac{1}{T} - P_f (\gamma_{1_f} + d_1)$.

**Proof:** See Appendix 3.C.

Theorem 3.6 shows that, even when the true model follows a factor model with $K_f$ (known) useful factors, the risk premia of the $K_g$ useless factors still converge to zero, even though the asymptotic distribution becomes more complicated. The estimator of the zero beta rate ($\hat{\gamma}_0$) and the risk premia of the useful factors ($\hat{\gamma}_{1_f}$) are biased, with the two bias terms given by $d_0$ and $d_1$ in (3.30), respectively. Even though these quantities are not the primary focus of our analysis, a consistent estimation of both $\hat{\gamma}_0$ and $\hat{\gamma}_{1_f}$ would be required in order to estimate the two bias terms and, hence, derive correctly-sized standard errors. This can be obtained by using the bias-adjusted estimator proposed in Rapo et al. (2019). Therefore, as in the base case, tests with correct size and power can be properly derived. We summarize the results in the next theorem.

Let $\hat{e}_{fg} = \hat{R} - \hat{X}_{fg} \hat{Y}_{fg}$ be the $N$-vector of OLS residuals from the regression in (3.27) and define
\[
s_{fg}^2 = \frac{\hat{e}_{fg}' \hat{e}_{fg}}{N - K_f - K_g - 1}.
\]
(3.31)

Let
\[
t_k = \frac{\hat{\gamma}_k}{s_{fg} \sqrt{(\hat{X}_{fg}' \hat{X}_{fg})^{-1}_{k,k}}}, \quad k_g = 1, ..., K_g
\]
(3.32)

be the $t$-statistic for the $k_g$-th regression coefficient in (3.27), where $(\hat{X}_{fg}' \hat{X}_{fg})^{-1}_{k,k}$ denotes the $(k_g, k_g)$-th element of the matrix $(\hat{X}_{fg}' \hat{X}_{fg})^{-1}$, for $k_g = 1, ..., K_g$. Let
be the $R$-squared of the model in (3.27). Finally, let $\hat{\varepsilon}_f^*$ be the $N$-vector of OLS residuals from the regression

$$\bar{R} = \bar{X}_f \Gamma_f^P + \eta_f^P$$

where $\bar{X}_f = [1, B_f, \Gamma_f^P = [\gamma_0, \gamma_{(1)}^P, \ldots, \gamma_{(K_f)}^P]'$, $\eta_f^P = (\bar{r} - (\bar{X}_f - X_f) \Gamma_f^P)'$ with $X_f = [1, B_f]$. Let

$$F_{CSR_{fs}} = \frac{(\hat{\varepsilon}_f^* \hat{\varepsilon}_f^* - \hat{\varepsilon}_f^* \hat{\varepsilon}_fs) / K_g}{\hat{\varepsilon}_f^* \hat{\varepsilon}_fs / (N - (K_f + K_g + 1))}$$

be the $F$-statistic to test the null hypothesis that $\gamma_{(p)}^P = 0_{K_g}$ in (3.27). Then, under the representation in (3.25), and assuming correct model specification, we can establish the following theorem.

**Theorem 3.7** Assume that the representation in (3.25) holds and that (3.9) is satisfied. Assume also that (3.29) is satisfied. Then, under Assumptions 3.1-3.5, we have

(i) 

$$t_{g, k_g} \rightarrow d \left( 0, \frac{d_1^\ast \Sigma_{\beta^T} d_1 + \sigma^{-2} W_{[k_g, k_g]}}{\frac{\sigma^2}{T} + \gamma_{(1)}^P \sigma^2 (\bar{P}' \bar{P})^{-1} D^{-1} \Sigma_{\beta^T} \gamma_{(1)}^P} \right)$$

where $W_{[k_g, k_g]}$ denotes the $(k_g, k_g)$-th element of the matrix $(Q_f' \otimes \tilde{G}) U (Q_f \otimes P_g)$

(ii) 

$$R_{\text{ CRS}_{fs}}^2 \rightarrow 1 - \frac{\frac{\sigma^2}{T} + \gamma_{(1)}^P \sigma^2 (\bar{P}' \bar{P})^{-1} D^{-1} \Sigma_{\beta^T} \gamma_{(1)}^P}{\frac{\sigma^2}{T} + \gamma_{(1)}^P \Sigma_{\beta^T} \gamma_{(1)}^P}$$

(iii) 

$$F_{\text{ CSR}_{fs}} \rightarrow [Z_1', \ Z_2'] \frac{W_{f_{12}} / K_g}{\frac{\sigma^2}{T} - d_1^\ast \Sigma_{\beta^T} d_1} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

where $Z_1 \equiv N(0_{T^2}, U_r)$ and $Z_2 \equiv N(0_{T}, d_1^\ast \Sigma_{\beta^T} d_1 I_T)$ are two normally distributed vectors of dimension $T^2 \times 1$ and $T \times 1$, respectively, and where

$$W_{f_{12}} = \frac{1}{\sigma^2} \begin{bmatrix} Q_f Q_f' \otimes P_g \tilde{G} & -Q_f \otimes P_g \tilde{G}' \\ -Q_f' \otimes P_g \tilde{G} & P_g \tilde{G}' \end{bmatrix}$$
3.4. TWO-PASS REGRESSION WITH USELESS FACTORS AND OMITTED USEFUL FACTORS 175

Proof: See Appendix 3.C.

Before moving to the next section, we want to highlight that the assumption in (3.29) leads to a very useful simplification in the derivation of the results in Theorems 3.6 and 3.7. As explained before, this assumption could be further relaxed, even though the results become more complicated. In particular, we show in the next proposition that, when $F$ and $G$ are allowed to be correlated (as it is often the case in practice), the risk premia estimator $\hat{\gamma}_{12}$ of the useless factors will not converge to zero anymore. However, under the null hypothesis of no useless factors, a set of linear restrictions can be still identified, allowing us to derive a corresponding testing procedure. The results are established in the next proposition.

Proposition 3.2 Assume that the representation in (3.25) holds and that (3.3) is satisfied. Then, under Assumptions 3.1-3.5, we have

$$
\hat{I}_{f_2} = \begin{bmatrix}
\hat{\gamma}_0 \\
\hat{\gamma}_{11} \\
\hat{\gamma}_{12}
\end{bmatrix} \rightarrow_{p} \begin{bmatrix}
\gamma_0 - \mu'_{\beta_f}(I_{K_f} - E^{-1}\Sigma_{\beta_f})\gamma_{f_2} \\
E^{-1}\Sigma_{\beta_f}\gamma_{f_2} \\
-AE^{-1}\Sigma_{\beta_f}\gamma_{f_2}
\end{bmatrix} = \begin{bmatrix}
\gamma_0 - \mu'_{\beta_f}(I_{K_f} - E^{-1}\Sigma_{\beta_f})\gamma_{f_2} \\
\theta_f \\
\theta_g
\end{bmatrix}
$$

where $\theta_f = E^{-1}\Sigma_{\beta_f}\gamma_{f_2}^P$, $\theta_g = -AE^{-1}\Sigma_{\beta_f}\gamma_{f_2}^P$, with

$$
E = \Sigma_{\beta_f} + \sigma^2\left((\hat{F}'\hat{F})^{-1} + (\hat{F}'\hat{F})(\hat{G}'\hat{G})(\hat{G}'\hat{F})(\hat{F}'\hat{F})^{-1} - Q_{f_2}DQ_{f_2}'\right),
$$

$$
D = (\hat{G}'\hat{G}) - (\hat{G}'\hat{F})(\hat{F}'\hat{F})^{-1}(\hat{G}'\hat{G}),
$$

and $Q_{f_2} = \sigma^2(\hat{F}'\hat{F})^{-1}\hat{F}'\hat{G}D^{-1}$, and $A = -\hat{G}'\hat{F}(\hat{F}'\hat{F})^{-1}$.  

Proof: See Appendix 3.C.

Proposition 3.2 states that, when $F$ and $G$ are not orthogonal to each other, the risk premium estimator $\hat{\gamma}_{12}$ associated to the useless factors does not converge to zero and its true value depends on the degree of correlation between the two sets of factors. This result complicates statistical inference, and we heed not test whether $G$ is a set of truly useless factors or not using the same procedure derived before. However, a set of testable linear restrictions that holds under the null hypothesis of useless factors can still be derived. Indeed, notice that, for any observable matrix $A$ of dimension $(K_g \times K_f)$, and under the null hypothesis that $g_i$ is a useless factor, we have that $\theta_g = A\theta_f$. Therefore, the following null hypothesis can be tested

$$
H_0: \theta_g = A\theta_f
$$

and the distribution of the corresponding statistical test can be derived using the results of our Theorem 3.6. Notice that, when $F$ and $G$ are uncorrelated, then $A = 0_{K_g \times K_f}$ and the null hypothesis in (3.36) reduces to $H_0: \theta_g = 0$, as already derived in our Theorem 3.6.

3.4. Two-pass regression with useless factors and omitted useful factors

In this section we generalize the results of the previous section to the case of potential model misspecification. As already mention before, misspecification can take various forms, e.g., incorrect specification of the functional form and/or mismeasured or missing factors. In this section, we consider the case of missing relevant factors from the analysis, when estimating the model using some useless factors. Formally, let us assume that the true model has the following specification

$$
R_t = \alpha + B_{[1]}' f_{[1]} + B_{[2]}' f_{[2]} + \epsilon_t
$$

$$
= \alpha + B_{f} f_t + \epsilon_t
$$
where \( f_t = \left[ f_{t1}^{[1]}, f_{t2}^{[2]}\right]' \) is a \( K_f \)-vector of factors, with \( f_{t1}^{[1]} \) and \( f_{t2}^{[2]} \) being two vectors of dimension \( K_{f1} \) and \( K_{f2} \), respectively, and such that \( K_f = K_{f1} + K_{f2} \). \( B_f = \left[ B_{f1}^{[1]}, B_{f2}^{[2]}\right] \) is an \( (N \times K_f) \) matrix of coefficients, which can be partitioned in \( B_{f1}^{[1]} \) and \( B_{f2}^{[2]} \) matrices, of dimension \( (N \times K_{f1}) \) and \( (N \times K_{f2}) \), respectively.

Suppose that, instead of the true model in (3.37), we estimate the following model:

\[
R_t = \alpha + B_{g}^{[1]} f_{t1}^{[1]} + B_{g}^{[2]} f_{t2}^{[2]} + \epsilon_t
\]

(3.39)

where again \( g_t \) is a \( K_g \)-vector of useless factors and \( B_g \) is an \( (N \times K_g) \) matrix of coefficients. In other words, the researcher is incidentally omitting the subset of factors \( f_{t2}^{[2]} \) from the analysis and, at the same time, he introduces some useless factors \( g_t \) when estimating the model. Notice that, under the true model in (3.37), \( B_g = 0_{N \times K_g} \). Let \( \hat{B}_f^{[1]} = \left[ \hat{\beta}_{f11}, \ldots, \hat{\beta}_{f1K_{f1}} \right]' \) and \( \hat{B}_g = \left[ \hat{\beta}_{g11}, \ldots, \hat{\beta}_{gK_{g}} \right]' \) be the two matrices of estimated coefficients from the first pass, of dimension \( (N \times K_{f1}) \) and \( (N \times K_g) \), respectively. Let \( K_{f2g} = K_{f2} + K_g \) and define the following \( (N \times (K_{f1g} + 1)) \) matrix

\[
\bar{X}_{f1g} = \begin{bmatrix} 1_N, & \hat{B}_f^{[1]}, & \hat{B}_g \end{bmatrix}
\]

(3.40)

Assume for now that the orthogonality relation in (3.29) is satisfied, i.e., useless and useful factors are orthogonal to each other. Then,

\[
\hat{B}_g = B_g + \epsilon' P_g = \epsilon' P_g
\]

(3.41)

and

\[
\hat{B}_f^{[1]} = B_f^{[1]} + \epsilon' P_f^{[1]},
\]

(3.42)

where \( P_f^{[1]} = \hat{F}^{[1]} (\hat{F}^{[1]}' \hat{F}^{[1]})^{-1} \hat{F}^{[1]} = \hat{M}_T F^{[1]} \), with \( \hat{M}_T = I_T - \frac{1_T 1_T'}{T} \) and \( F^{[1]} \) is the \( (T \times K_{f1}) \) matrix of useful factors included in the fitted model in (3.39).

Let \( \hat{R}_t = \frac{1}{T} \sum_{t=1}^T R_{it} \) and \( \hat{R} = [\hat{R}_1, \ldots, \hat{R}_N]' \). In order to estimate \( \Gamma_f \), we then run the following CSR:

\[
\hat{R} = \hat{X}_{f1g} \Gamma_{f1g} + \eta_{f1g}
\]

(3.43)

where \( \Gamma_{f1g} = [\gamma_{f11}, \gamma_{f12}, \gamma_{f13}]' \) and \( \eta_{f1g} = (\tau + X_f \Gamma_f - \hat{X}_{f1g} \Gamma_{f1g})' \), with \( X_f = \begin{bmatrix} 1_N, B_f^{[1]}, B_f^{[2]} \end{bmatrix} \) and \( \Gamma_f = \begin{bmatrix} \gamma_{f11}, \gamma_{f12}, \gamma_{f13} \end{bmatrix}' \). Using the feasible representation in (3.43), then the OLS estimator has the following representation

\[
\hat{\Gamma}_{f1g} = \left( \hat{X}_{f1g}' \hat{X}_{f1g} \right)^{-1} \hat{X}_{f1g}' \hat{R}.
\]

(3.44)

Define \( \hat{\Sigma}_{f1} = \left( \hat{\Sigma}_{f1} - \mu_{f1}' \mu_{f1} \right)' \), \( \hat{\Sigma}_{f2} = \left( \hat{\Sigma}_{f2} - \mu_{f2}' \mu_{f2} \right)' \), \( \hat{\Sigma}_{f3} = \left( \hat{\Sigma}_{f3} - \mu_{f3}' \mu_{f3} \right)' \), where \( \mu_{f1} = \text{plim} \frac{1}{N} \sum_{i=1}^N \hat{F}_i^{[1]}, \mu_{f2} = \text{plim} \frac{1}{N} \sum_{i=1}^N \hat{F}_i^{[2]}, \mu_{f3} = \text{plim} \frac{1}{N} \sum_{i=1}^N \hat{F}_i^{[3]} \). Let \( D_0 = \sigma^2 \mu_{f1}' D_1^{-1} (\hat{F}_{f1}' \hat{F}_{f1})^{-1} \gamma_{f1}' + \mu_{f2}' \gamma_{f2}' + \mu_{f3}' \gamma_{f3}' \) and \( D_1 = D_1^{-1} \hat{\Sigma}_{f1} \gamma_{f1}' + \hat{\Sigma}_{f2} \gamma_{f2}' + \hat{\Sigma}_{f3} \gamma_{f3}' \). Then, under the representation in (3.37) and assuming (3.29), we can establish the following theorems.
Theorem 3.8 Assume that the representation in (3.37) holds. Then, the CSR OLS estimator in (3.44) under (3.29) satisfies:

(i) Under Assumptions 3.1-3.4

\[ \hat{\Gamma}_{f,g} - \begin{bmatrix} \gamma_0 + \tilde{d}_0 \\ \tilde{d}_{11} \gamma_1^{P[1]} + \tilde{d}_{12} \gamma_1^{P[2]} \\ 0_{K_s} \end{bmatrix} = O_P \left( \frac{1}{\sqrt{N}} \right) \]

(ii) Let \( \vartheta = \left( d_1 \Sigma_{\beta_1}^{[1]} \hat{d}_1 + \gamma_1^{P[1]} \Sigma_{\hat{d}_1}^{[1]} \hat{d}_1 - 2 \gamma_1^{P[2]} \Sigma_{\hat{d}_1}^{[2]} \hat{d}_1 \right) \). Then, under Assumptions 3.1-3.5

\[ \sqrt{N} \begin{bmatrix} \gamma_0 + \tilde{d}_0 \\ \tilde{d}_{11} \gamma_1^{P[1]} + \tilde{d}_{12} \gamma_1^{P[2]} \\ 0_{K_s} \end{bmatrix} \xrightarrow{d} \left( 0_{(K_s + K_s + 1)} \right), \left( \Sigma_{X_{f,s}} + \Lambda_{f,g}^{[1]} \right)^{-1} \left( V_{f,g} + W_{f,g} \right) \left( \Sigma_{X_{f,s}} + \Lambda_{f,g}^{[1]} \right)^{-1} \]

where

\[ \Sigma_{X_{f,s}}^{[1]} = \begin{bmatrix} 1 & \mu_{\beta_1}^{[1]} & 0_{K_s} \\ \mu_{\beta_1}^{[1]} & \Sigma_{\beta_1}^{[1]} & 0_{K_s \times K_s} \\ 0_{K_s} & 0_{K_s \times K_s} & 0_{K_s} \end{bmatrix}, \quad \Lambda_{f,g}^{[1]} = \begin{bmatrix} 0 & 0_{K_s}^{[1]} & 0_{K_s} \\ 0_{K_s}^{[1]} & \sigma^2 (\hat{F}_f^{[1]} \hat{F}_f^{[1]})^{-1} & 0_{K_s \times K_s} \\ 0_{K_s} & 0_{K_s \times K_s} & \sigma^2 (\Sigma_f^{[1]} \Sigma_f^{[1]})^{-1} \end{bmatrix}, \]

\[ V_{f,g} = \sigma^2 \left( \frac{1}{T} + (\tilde{d}_{11} \gamma_1^{P[1]} + \tilde{d}_{12} \gamma_1^{P[2]}) (\hat{F}_f \hat{F}_f)^{-1} (\tilde{d}_{11} \gamma_1^{P[1]} + \tilde{d}_{12} \gamma_1^{P[2]}) \right) \Sigma_{X_{f,s}} + \sigma^2 \Omega_{f,g}, \]

with

\[ \Omega_{f,g} = \begin{bmatrix} 0 & 0_{K_s}^{[1]} \\ 0_{K_s}^{[1]} & \vartheta (\hat{F}_f^{[1]} \hat{F}_f^{[1]})^{-1} \end{bmatrix} - \left( \tilde{d}_{11} \Sigma_{\hat{d}_1}^{[1]} \gamma_1^{P[1]} - \tilde{d}_{12} \Sigma_{\hat{d}_1}^{[2]} \gamma_1^{P[2]} \right) \left( \tilde{d}_{11} \Sigma_{\hat{d}_1}^{[1]} \gamma_1^{P[1]} - \tilde{d}_{12} \Sigma_{\hat{d}_1}^{[2]} \gamma_1^{P[2]} \right), \]

and where

\[ W_{f,g} = \begin{bmatrix} 0_{K_s}^{[1]} \\ 0_{K_s}^{[1]} \left( Q_f^{[1,2]} \otimes P_f^{[1]} \right) U_{s_f} (Q_f^{[1,2]} \otimes P_f^{[1]}) \\ 0_{K_s} \\ 0_{K_s} \left( Q_f^{[1,2]} \otimes P_f^{[2]} \right) U_{s_f} (Q_f^{[1,2]} \otimes P_f^{[2]}) \end{bmatrix}, \]

with

\[ Q_f^{[1,2]} = \left( \frac{1}{T} - P_f^{[1]} \tilde{d}_{11} \gamma_1^{P[1]} - P_f^{[2]} \tilde{d}_{12} \gamma_1^{P[2]} \right). \]

Proof: See Appendix 3.C.

Finally, let \( \hat{e}_{f,g} = \hat{R} - \hat{X}_{f,g} \hat{\Gamma}_{f,g} \) denote the N-vector of OLS residuals from the second pass in (3.43), and define

\[ s_{f,g}^2 = \frac{\hat{e}_{f,g} \hat{e}_{f,g}}{N - K_{f_s} - K_g - 1}. \]
Let

\[ t_{f_1, g, k_\alpha} = \frac{\hat{e}_{f_1, g, k_\alpha}}{s_{f_1, g, k_\alpha} / \sqrt{c_{g, k_\alpha}}} \]  

(3.46)

be the t-statistic for the \( k_\alpha \)-th regression coefficient in (3.43), with \( k_\alpha = 1, \ldots, K_\theta \) and where \( e_{f_1, g, k_\alpha} \) denotes the \((k_\alpha, k_\beta)\)-th element of the matrix \((\hat{X}_{f_1, g}^\prime \hat{X}_{f_1, g})^{-1}\). Let

\[ R_{_{CSR_{f_1, g}}}^2 = 1 - \frac{\hat{e}_{f_1, g}^\prime \hat{e}_{f_1, g}}{R_0 \mathcal{M}_N \bar{R}} \]  

(3.47)

be the \( R \)-squared of the model in (3.43), and let \( \hat{e}_{f_1, g} \) be the N-vector of OLS residuals from the restricted regression under the null hypothesis that there are no useless factors, i.e.,

\[ \bar{R} = \hat{X}_{f_1} \Gamma_{f_1}^P + \eta_{f_1}^P \]  

(3.48)

where \( \hat{X}_{f_1} = [1_N, \hat{B}^{[1]}_f] \), \( \Gamma_{f_1}^P = \begin{bmatrix} \gamma_0, \gamma_{1,f}^{P[1]}' \end{bmatrix}', \eta_{f_1}^P = \left( \bar{\tau} - (\hat{B}^{[1]}_f - \hat{B}^{[1]}_f) \gamma_{1,f}^{P[1]} + \hat{B}^{[2]}_f \gamma_{1,f}^{P[2]} \right) \). Let

\[ F_{_{CSR_{f_1, g}}} = \frac{(\hat{e}_{f_1, g}^\prime \hat{e}_{f_1, g} - \hat{e}_{f_1, g}^\prime \hat{e}_{f_1, g}) / K_\alpha}{\hat{e}_{f_1, g}^\prime \hat{e}_{f_1, g} / (N - K_{f_1} - K_\alpha - 1)} \]  

(3.49)

be the \( F \)-statistic to test the null hypothesis that all the useless factors are jointly equal to zero in the model (3.43), that is, \( H_0 : \gamma_{1,f}^P = 0_{K_\alpha} \). Then, the following asymptotic results hold.

**Theorem 3.9** Assume that the representation in (3.37) holds and that (3.29) is satisfied. Then, under Assumptions 3.1-3.5, we have

(i)

\[ t_{f_1, g, k_\alpha} \rightarrow_d \left( 0, \frac{\sigma^2 W_{[k_\alpha, k_\beta]}}{\bar{\tau} + \Gamma_{f_1} \Sigma_X \Gamma_{f_1}^\prime} \right) \]  

(3.50)

where \( W_{[k_\alpha, k_\beta]} \) denotes the \((k_\alpha, k_\beta)\)-th element of the matrix \((Q_f^{[1,2]}' \otimes \tilde{G}'_0) U_0 (Q_f^{[1,2]} \otimes P_0)\),

\[ \Gamma_{f_1}^P = \begin{bmatrix} \gamma_{1,f}^{P[1]} \gamma_{1,f}^{P[2]} \end{bmatrix}' \]  

and where

\[ \Sigma_X = \begin{bmatrix} \Sigma_{\tilde{\beta}_f} \quad \Sigma_{\tilde{\beta}_f} D^{-1} \Sigma_{\tilde{\beta}_f} \\ \sigma^2 \Sigma_{\tilde{\beta}_f} D^{-1} (\hat{\beta}_f^{[1]} \hat{\beta}_f^{[1]})^{-1} \Sigma_{\tilde{\beta}_f} \quad \sigma^2 \Sigma_{\tilde{\beta}_f} D^{-1} (\hat{\beta}_f^{[2]} \hat{\beta}_f^{[2]})^{-1} \Sigma_{\tilde{\beta}_f} \end{bmatrix} \].

(ii)

\[ R_{_{CSR_{f_1, g}}}^2 ightarrow_p 1 - \frac{\sigma^2}{\bar{\tau} + \Gamma_{f_1} \Sigma_{\tilde{\beta}_f} \Gamma_{f_1}^\prime} + \frac{\Gamma_{f_1} \Sigma_{\tilde{\beta}_f} \Gamma_{f_1}^\prime}{\bar{\tau} + \Gamma_{f_1} \Sigma_{\tilde{\beta}_f} \Gamma_{f_1}^\prime} \]  

where \( \Gamma_{f_1}^P = \begin{bmatrix} \gamma_{1,f}^{P[1]} \gamma_{1,f}^{P[2]} \end{bmatrix}' \) and \( \Sigma_{\tilde{\beta}_f} = \begin{bmatrix} \Sigma_{\tilde{\beta}_f} \Sigma_{\tilde{\beta}_f}^{[1,2]} \\ \Sigma_{\tilde{\beta}_f}^{[1]} \Sigma_{\tilde{\beta}_f}^{[2]} \end{bmatrix} \).
(iii) 

\[ F_{CSR_{i,t}} \overset{d}{=} \left[ Z'_1, \frac{W_{lf}/K_2}{\frac{1}{\sigma^2} + \Gamma_{lf}'\Sigma_{lf}^{-1}\Gamma_{lf}} Z'_2 \right] \]

where \( Z_1 \sim \mathcal{N}(0, \Sigma_{lf}) \) and \( Z_2 \sim \mathcal{N}(0, \delta \sigma^2 I_T) \) are two normally distributed vectors of dimension \((T^2 \times 1)\) and \((T \times 1)\), respectively and where

\[ W_{lf} = \frac{1}{\sigma^2} \begin{bmatrix} Q_{lf}Q_{lf}' \otimes P_{lf} \hat{G}' & -Q_{lf} \otimes P_{lf} \hat{G}' \\ -Q_{lf}' \otimes P_{lf} \hat{G}' & P_{lf} \hat{G}' \end{bmatrix}. \]

The proof of this theorem is available upon request.

Theorems 3.8 and 3.9 derive the asymptotic distributions of the CSR OLS estimator, the \( t \)-statistics, \( R \)-squared and \( F \)-statistics when the estimated model is not only including useless factors in the regression, but is also omitting some useful factors. Given that this situation could be quite realistic in practice, the two theorems provide the asymptotic results that would allow us to estimate and make inference on model parameters. However, non trivial issues arise in this case, which could make inferential procedures completely unfeasible in practice. Indeed, all the asymptotic results of Theorems 3.8 and 3.9 crucially depend on quantities (e.g., \( \mu_{F}^{[2]}, \Sigma_{F}^{[2]}, \) and \( \gamma_{F}^{[2]} \)) related to the set \( F^{[2]} \) of omitted factors, that need to be consistently estimated in order to implement our results in practice. Although one can assess the validity of these results using simulation exercises, in practice, of course, \( F^{[2]} \) is not observed by the researcher, rendering all the procedures empirically unusable.

Solutions to such problem can be found by deriving appropriate bounds for the limiting distributions in the theorems above (see Kan and Zhang (1999a)). However, they would be quite inaccurate under our large-\( N \) setting. Other possible alternatives, such as Principal Component techniques (see Giglio and Xiu (2017)), could be also adopted empirically, but they are not analyzed in this paper.

3.5. Simulation results

In this section, we undertake a Monte Carlo simulation experiment to assess the sample properties of the CSR OLS estimator in the presence of useless factors and the empirical rejection rates of the \( t \)-test and \( F \)-test derived in Sections 3.2 and 3.3. We focus here on the case of correctly specified models. In Section 3.5.1 we report the results for the base case derived in Section 3.2.1. Section 3.5.2 analyzes the more general case of models containing both useful and useless factors, assuming that there are no missing factors, as formalized in Section 3.3.

In all of our simulation experiments, we consider balanced panels with a time-series dimension of \( T = 36 \) and \( T = 72 \) monthly observations and calibrate our parameters of interest by employing a cross-section of 3,000 stocks from the Center for Research in Security Prices (CRSP) database, for the period from January 1966 to December 2013. We consider cross-sections of \( N = 100, 500, 1,000, \) and \( 3,000 \) stocks. All results are based on 10,000 Monte Carlo replications. Our econometric approach, designed for large \( N \) and fixed \( T \), should be able to handle this large number of assets over relative short time spans.

In all the cases, the return-generating process follows a linear specification, where the error term has a multivariate normal distribution with variance-covariance matrix \( \Sigma \), i.e., \( \epsilon_t \sim \mathcal{N}(0, \Sigma) \). For the calibration of the matrix \( \Sigma \), we use the same setting of Raponi et al. (2019). In particular, we consider three different specifications. We start from the case in which \( \Sigma \) is spherical, i.e., \( \Sigma = \sigma^2 I_T \). We then analyze the case where \( \Sigma \) is either diagonal or full. As all our theoretical results rely on the assumption that the model disturbances are weakly cross-sectionally correlated, we generate shocks under a weak factor structure, using the same data-generating process of Raponi et al. (2019), where

\[ \epsilon^{(1)} = \eta \left( \frac{\sqrt{\theta}}{N^{1/2}} \right) \epsilon' + \sqrt{1 - \theta} Z, \]
where $\eta$ and $c$ are $T$ and $N$-vectors of i.i.d. standard normal random variables, respectively, $Z$ is a $T \times N$ matrix of i.i.d. standard normal random variables, $0 \leq \theta \leq 1$ is a shrinkage parameter that controls the weight assigned to the diagonal and extra-diagonal elements of $\Sigma$, and $\delta$ is a parameter that controls the strength of the cross-sectional dependence of the shocks (the bigger $\delta$ is, the weaker the dependence). Our $(T \times N)$ matrix of shocks is then generated as

$$
\epsilon = \epsilon^{(1)} \cdot \begin{bmatrix} \frac{\sigma_1^2}{\hat{N} \hat{\Sigma} c_1^2 \left(1 - \theta \right)} & \cdots & \frac{\sigma_1^2}{\hat{N} \hat{\Sigma} c_N^2 \left(1 - \theta \right)} \\
\vdots & \ddots & \vdots \\
\frac{\sigma_N^2}{\hat{N} \hat{\Sigma} c_N^2 \left(1 - \theta \right)} & \cdots & \frac{\sigma_N^2}{\hat{N} \hat{\Sigma} c_N^2 \left(1 - \theta \right)} 
\end{bmatrix}^{-0.5},
$$

where $c_i$ is the $i$-th element of $c$. Given this specification for the shocks, for our theoretical results to hold, we require $\delta > 0$.

### 3.5.1 Base case

In this Section we analyze the finite-sample properties of the CSR OLS estimator, together with the behavior of the $R$-squared and the empirical rejection rates of the $t$-test and $F$-test, under the base case of Section 3.2.1.

The return-generating process under the null of a correctly specified asset-pricing model is given by

$$
R_t = \gamma_0 1_N + \epsilon_t,
$$

where $\epsilon_t \sim N(0, \Sigma)$. The fitted model is a one-factor model

$$
R_{it} = a_i + b_i' g_t + u_{it},
$$

where $g_t$ is the excess market return (from Kenneth French’s website) from January 2008 to December 2010 for $T=36$, and the excess market return from January 2008 to December 2013 for $T=72$.

Tables 3.1-3.4 report the percentage error (Bias) and the root mean squared error (RMSE), in percent, of the OLS estimator of $\gamma_0$ together with the $R$-squared ($R^2$) of the estimated model. For $\hat{\gamma}_1$ we report the absolute bias (i.e. the difference between the estimated coefficient and its true value), as the true value in this case is exactly equal to zero. Panels A and Panel B are for $T=36$ and $T=72$, respectively. Table 3.1 refers to the case of $\Sigma = \sigma^2 I_T$, while Table 3.2 considers the case of $\Sigma$ diagonal, that is obtained when we set $\theta = 0$ in the generating process (3.51). Finally, in Tables 3.3 and 3.4 we allow for weak cross-sectional dependence of the model disturbances, by setting $\theta = 0.5$, and considering $\delta = 0.5$ and $\delta = 0.25$ in (3.51), respectively.

The percentage bias of the OLS estimator is very small even for $N = 100$ (0.32% for $\hat{\gamma}_0$ and null for $\hat{\gamma}_1$) and becomes negligible as $N$ increases. As expected, also the RMSE becomes smaller as $N$ increases. In particular, $\hat{\gamma}_0$ exhibits a smaller RMSE than $\hat{\gamma}_1$, with such difference becoming smaller as $N$ increases.

The $R$-squared of the model converges to zero, confirming our theoretical results derived in Section 3.2.1. Panel B for $T = 72$ conveys a similar message. As expected from the theoretical analysis, the larger time-series dimension helps in reducing the bias and RMSE associated with the OLS estimator.

The OLS estimator continues to perform very well also in the presence of cross-sectional dependence, either when $\Sigma$ is diagonal and full, with a slightly higher bias and RMSE compared to the spherical case.

Tables 3.5-3.8 consider the empirical rejection rates of the $t$-test derived in Theorem 3.3. The null hypothesis is that the risk premium of the useless factor is equal to zero. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of time-series and cross-sectional observations. The model disturbances for the four tables are generated as in Tables 3.1 through Table 3.4. The $t$-statistics are compared with the critical values of a standard normal distribution. The performance of
the $t$-statistic is very good for all $T$ and $N$ and for different cases of cross-sectional dependence. Only when $\delta = 0.25$ (Table 3.8) we start to notice a slight over-rejection of the $t$-test.

Finally, in Tables 3.9-3.12 we report the size properties of the $F$-statistic derived in Theorem 3.3. The null hypothesis is that all the $K_g$ parameters (except for the intercept) are all equal to zero in the fitted model in (3.53). As before, we consider different levels of significance (10%, 5%, and 1%) and different values of time-series and cross-sectional observations. The $F$-statistics are compared with the critical values of a Chi-squared distribution, with $\sigma_4/\sigma^4$ and $K_g$ degrees of freedom. Overall, the results suggest that the proposed $F$-test is relatively well behaved even when moving toward a fairly strong factor structure in the residuals (i.e., up to the case of $\delta = 0.5$). When $\delta = 0.25$, as before, we start to observe a slight over-rejection, which is consistent with our theoretical results.

Table 3.1  
Bias and RMSE of the OLS Estimator in a one-factor model with a useless factor ($\Sigma$ scalar)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Panel A: $T = 36$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias($\gamma_0$)</td>
<td>0.32%</td>
<td>0.18%</td>
<td>0.12%</td>
<td>0.11%</td>
</tr>
<tr>
<td>RMSE($\gamma_0$)</td>
<td>0.184</td>
<td>0.083</td>
<td>0.058</td>
<td>0.035</td>
</tr>
<tr>
<td>Bias($\gamma_1$)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE($\gamma_1$)</td>
<td>0.429</td>
<td>0.191</td>
<td>0.134</td>
<td>0.082</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.006</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Panel B: $T = 72$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias($\gamma_0$)</td>
<td>0.05%</td>
<td>0.04%</td>
<td>0.03%</td>
<td>0.04%</td>
</tr>
<tr>
<td>RMSE($\gamma_0$)</td>
<td>0.146</td>
<td>0.066</td>
<td>0.046</td>
<td>0.028</td>
</tr>
<tr>
<td>Bias($\gamma_1$)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE($\gamma_1$)</td>
<td>0.379</td>
<td>0.166</td>
<td>0.119</td>
<td>0.072</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets. The return generating process is given by $R_t = \gamma_0 1_N + \epsilon_t$, where $\epsilon_t \sim N(0, \Sigma)$ and where we calibrate $\gamma_0$ as $\gamma_0 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{t=1}^{T} R_{it}$. The fitted Model is a One-Factor Model $R_{it} = a_i + b_i \gamma_0 + u_{it}$, where $\gamma_0$ is the excess market return (from Kenneth French’s website) from January 2008 to December 2010 for $T=36$, and the excess market return from January 2008 to December 2013 for $T=72$. The table also reports the $R$-squared ($R^2$) of the fitted model for different cross-sections of $N = 100, 500, 1000, 3000$ stocks.
Table 3.2  
Bias and RMSE of the OLS Estimator in a one-factor model with a useless factor ($\Sigma$ Diagonal).

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Panel A: $T = 36$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias($\gamma_0$)</td>
<td>-0.15%</td>
<td>0.15%</td>
<td>0.08%</td>
<td>0.02%</td>
</tr>
<tr>
<td>RMSE($\gamma_0$)</td>
<td>0.923</td>
<td>0.425</td>
<td>0.308</td>
<td>0.190</td>
</tr>
<tr>
<td>Bias($\gamma_1$)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE($\gamma_1$)</td>
<td>0.764</td>
<td>0.330</td>
<td>0.227</td>
<td>0.135</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.030</td>
<td>0.006</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Panel B: $T = 72$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias($\gamma_0$)</td>
<td>0.07%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.02%</td>
</tr>
<tr>
<td>RMSE($\gamma_0$)</td>
<td>0.400</td>
<td>0.160</td>
<td>0.127</td>
<td>0.075</td>
</tr>
<tr>
<td>Bias($\gamma_1$)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE($\gamma_1$)</td>
<td>1.070</td>
<td>0.521</td>
<td>0.332</td>
<td>0.208</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.069</td>
<td>0.018</td>
<td>0.008</td>
<td>0.003</td>
</tr>
</tbody>
</table>

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets. The return generating process is given by $R_t = \gamma_0 1_N + \epsilon_t$, where $\epsilon_t \sim N(0, \Sigma)$ and where we calibrate $\gamma_0$ as $\gamma_0 = \frac{1}{N} \sum_{t=1}^{N} \sum_{i=1}^{T} R_{it}$. The fitted Model is a One-Factor Model $R_{it} = a_i + b_i g_t + u_{it}$, where $g_t$ is the excess market return (from Kenneth French’s website) from January 2008 to December 2010 for $T=36$, and the excess market return from January 2008 to December 2013 for $T=72$. The table also reports the $R$-squared ($R^2$) of the fitted model for different cross-sections of $N = 100, 500, 1000, 3000$ stocks.
3.5. SIMULATION RESULTS

Table 3.3
Bias and RMSE of the OLS Estimator in a one-factor model with a useless factor ($\Sigma$ Full - $\delta = 0.5$).

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: $T = 36$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias((\gamma_0))</td>
<td>-0.16%</td>
<td>0.13%</td>
<td>0.06%</td>
<td>0.05%</td>
</tr>
<tr>
<td>RMSE((\gamma_0))</td>
<td>0.923</td>
<td>0.425</td>
<td>0.305</td>
<td>0.189</td>
</tr>
<tr>
<td>Bias((\gamma_1))</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE((\gamma_1))</td>
<td>1.253</td>
<td>0.474</td>
<td>0.349</td>
<td>0.196</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.031</td>
<td>0.006</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>Panel B: $T = 72$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias((\gamma_0))</td>
<td>-0.03%</td>
<td>0.02%</td>
<td>0.05%</td>
<td>0.03%</td>
</tr>
<tr>
<td>RMSE((\gamma_0))</td>
<td>0.353</td>
<td>0.178</td>
<td>0.118</td>
<td>0.078</td>
</tr>
<tr>
<td>Bias((\gamma_1))</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE((\gamma_1))</td>
<td>0.764</td>
<td>0.329</td>
<td>0.230</td>
<td>0.138</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.090</td>
<td>0.015</td>
<td>0.009</td>
<td>0.003</td>
</tr>
</tbody>
</table>

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets. The return generating process is given by $R_t = \gamma_0 1_N + \epsilon_t$, where $\epsilon_t \sim N(0, \Sigma)$ and where we calibrate $\gamma_0$ as $\gamma_0 = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N R_{it}$. The fitted Model is a One-Factor Model $R_{it} = \alpha_i + \beta g_t + \mu_{it}$, where $g_t$ is the excess market return (from Kenneth French's website) from January 2008 to December 2010 for $T=36$, and the excess market return from January 2008 to December 2013 for $T=72$. The table also reports the $R$-squared ($R^2$) of the fitted model for different cross-sections of $N = 100, 500, 1000, 3000$ stocks.
Table 3.4
Bias and RMSE of the OLS Estimator in a one-factor model with a useless factor (Σ Full - δ = 0.25).

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: $T = 36$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{Bias}(\hat{\gamma}_0)$</td>
<td>0.11%</td>
<td>0.22%</td>
<td>0.61%</td>
<td>0.14%</td>
</tr>
<tr>
<td>$\text{RMSE}(\hat{\gamma}_0)$</td>
<td>0.922</td>
<td>0.426</td>
<td>0.326</td>
<td>0.188</td>
</tr>
<tr>
<td>$\text{Bias}(\hat{\gamma}_1)$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\text{RMSE}(\hat{\gamma}_1)$</td>
<td>0.768</td>
<td>0.338</td>
<td>0.228</td>
<td>0.159</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.031</td>
<td>0.006</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>Panel B: $T = 72$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{Bias}(\hat{\gamma}_0)$</td>
<td>0.08%</td>
<td>0.06%</td>
<td>0.03%</td>
<td>0.01%</td>
</tr>
<tr>
<td>$\text{RMSE}(\hat{\gamma}_0)$</td>
<td>0.342</td>
<td>0.160</td>
<td>0.130</td>
<td>0.077</td>
</tr>
<tr>
<td>$\text{Bias}(\hat{\gamma}_1)$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\text{RMSE}(\hat{\gamma}_1)$</td>
<td>1.278</td>
<td>0.525</td>
<td>0.331</td>
<td>0.211</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.093</td>
<td>0.019</td>
<td>0.008</td>
<td>0.003</td>
</tr>
</tbody>
</table>

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets. The return generating process is given by $R_t = \gamma_0 1_N + \epsilon_t$, where $\epsilon_t \sim N(0, \Sigma)$ and where we calibrate $\gamma_0$ as $\hat{\gamma}_0 = \frac{1}{N} \sum_{t=1}^{N} \sum_{i=1}^{D} R_{it}$. The fitted Model is a One-Factor Model $R_{it} = \alpha_i + \beta_i g_t + u_{it}$, where $g_t$ is the excess market return (from Kenneth French’s website) from January 2008 to December 2010 for $T=36$, and the excess market return from January 2008 to December 2013 for $T=72$. The table also reports the $R$-squared ($R^2$) of the fitted model for different cross-sections of $N = 100, 500, 1000, 3000$ stocks.
Table 3.5
Empirical size of t-tests in a one-factor model with a useless factor (Σ Scalar)

<table>
<thead>
<tr>
<th>N</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: T = 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(\hat{\gamma}_0)</td>
<td>t(\hat{\gamma}_1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.105</td>
<td>0.053</td>
<td>0.010</td>
<td>0.105</td>
<td>0.055</td>
<td>0.013</td>
</tr>
<tr>
<td>500</td>
<td>0.107</td>
<td>0.052</td>
<td>0.011</td>
<td>0.108</td>
<td>0.054</td>
<td>0.011</td>
</tr>
<tr>
<td>1000</td>
<td>0.106</td>
<td>0.051</td>
<td>0.011</td>
<td>0.103</td>
<td>0.053</td>
<td>0.011</td>
</tr>
<tr>
<td>3000</td>
<td>0.102</td>
<td>0.050</td>
<td>0.010</td>
<td>0.102</td>
<td>0.052</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel B: T = 72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(\hat{\gamma}_0)</td>
<td>t(\hat{\gamma}_1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.103</td>
<td>0.053</td>
<td>0.009</td>
<td>0.098</td>
<td>0.055</td>
<td>0.011</td>
</tr>
<tr>
<td>500</td>
<td>0.103</td>
<td>0.051</td>
<td>0.010</td>
<td>0.098</td>
<td>0.046</td>
<td>0.009</td>
</tr>
<tr>
<td>1000</td>
<td>0.101</td>
<td>0.051</td>
<td>0.009</td>
<td>0.096</td>
<td>0.051</td>
<td>0.010</td>
</tr>
<tr>
<td>3000</td>
<td>0.101</td>
<td>0.051</td>
<td>0.010</td>
<td>0.099</td>
<td>0.050</td>
<td>0.009</td>
</tr>
</tbody>
</table>

The table presents the size properties of t-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1-2013:12. t(\cdot) denotes the t-statistic associated with the OLS estimator as derived in Theorem 3.3. The t-statistics are compared with the critical values from a standard normal distribution.
Table 3.6
Empirical size of t-tests in a one-factor model with a useless factor (Σ Diagonal)

<table>
<thead>
<tr>
<th>N</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: T = 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t(\gamma_0)$</td>
<td>$t(\gamma_1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.095</td>
<td>0.048</td>
<td>0.010</td>
<td>0.113</td>
<td>0.059</td>
<td>0.014</td>
</tr>
<tr>
<td>500</td>
<td>0.098</td>
<td>0.048</td>
<td>0.009</td>
<td>0.102</td>
<td>0.050</td>
<td>0.011</td>
</tr>
<tr>
<td>1000</td>
<td>0.099</td>
<td>0.050</td>
<td>0.011</td>
<td>0.103</td>
<td>0.052</td>
<td>0.011</td>
</tr>
<tr>
<td>3000</td>
<td>0.102</td>
<td>0.050</td>
<td>0.009</td>
<td>0.100</td>
<td>0.050</td>
<td>0.012</td>
</tr>
<tr>
<td>Panel B: T = 72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t(\gamma_0)$</td>
<td>$t(\gamma_1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.087</td>
<td>0.041</td>
<td>0.006</td>
<td>0.122</td>
<td>0.072</td>
<td>0.023</td>
</tr>
<tr>
<td>500</td>
<td>0.096</td>
<td>0.050</td>
<td>0.009</td>
<td>0.104</td>
<td>0.057</td>
<td>0.012</td>
</tr>
<tr>
<td>1000</td>
<td>0.097</td>
<td>0.047</td>
<td>0.010</td>
<td>0.108</td>
<td>0.053</td>
<td>0.013</td>
</tr>
<tr>
<td>3000</td>
<td>0.099</td>
<td>0.051</td>
<td>0.010</td>
<td>0.101</td>
<td>0.051</td>
<td>0.010</td>
</tr>
</tbody>
</table>

The table presents the size properties of t-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1-2013:12. $t(\cdot)$ denotes the t-statistic associated with the OLS estimator as derived in Theorem 3.3. The t-statistics are compared with the critical values from a standard normal distribution.
Table 3.7
Empirical size of t-tests in a one-factor model with a useless factor (Σ Full - δ = 0.5)

<table>
<thead>
<tr>
<th>N</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(t(\hat{\gamma}_0))</td>
<td></td>
<td></td>
<td>(t(\hat{\gamma}_1))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.096</td>
<td>0.048</td>
<td>0.010</td>
<td>0.113</td>
<td>0.059</td>
<td>0.014</td>
</tr>
<tr>
<td>500</td>
<td>0.097</td>
<td>0.048</td>
<td>0.010</td>
<td>0.101</td>
<td>0.051</td>
<td>0.011</td>
</tr>
<tr>
<td>1000</td>
<td>0.102</td>
<td>0.048</td>
<td>0.010</td>
<td>0.104</td>
<td>0.052</td>
<td>0.011</td>
</tr>
<tr>
<td>3000</td>
<td>0.099</td>
<td>0.050</td>
<td>0.010</td>
<td>0.103</td>
<td>0.051</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Panel A: \(T = 36\)

Panel B: \(T = 72\)

<table>
<thead>
<tr>
<th>N</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(t(\hat{\gamma}_0))</td>
<td></td>
<td></td>
<td>(t(\hat{\gamma}_1))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.087</td>
<td>0.044</td>
<td>0.008</td>
<td>0.129</td>
<td>0.073</td>
<td>0.022</td>
</tr>
<tr>
<td>500</td>
<td>0.102</td>
<td>0.050</td>
<td>0.009</td>
<td>0.105</td>
<td>0.052</td>
<td>0.011</td>
</tr>
<tr>
<td>1000</td>
<td>0.096</td>
<td>0.047</td>
<td>0.009</td>
<td>0.100</td>
<td>0.050</td>
<td>0.013</td>
</tr>
<tr>
<td>3000</td>
<td>0.100</td>
<td>0.049</td>
<td>0.010</td>
<td>0.097</td>
<td>0.049</td>
<td>0.011</td>
</tr>
</tbody>
</table>

The table presents the size properties of t-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008.1-2013.12. \(t(\cdot)\) denotes the t-statistic associated with the OLS estimator as derived in Theorem 3.3. The t-statistics are compared with the critical values from a standard normal distribution.
Table 3.8
Empirical size of t-tests in a one-factor model with a useless factor (\( \Sigma \) Full - \( \delta = 0.25 \))

<table>
<thead>
<tr>
<th>N</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t(\hat{\gamma}_0) )</td>
<td>( t(\hat{\gamma}_1) )</td>
<td>( t(\hat{\gamma}_0) )</td>
<td>( t(\hat{\gamma}_1) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.097</td>
<td>0.048</td>
<td>0.010</td>
<td>0.112</td>
<td>0.062</td>
<td>0.014</td>
</tr>
<tr>
<td>500</td>
<td>0.101</td>
<td>0.049</td>
<td>0.011</td>
<td>0.112</td>
<td>0.059</td>
<td>0.014</td>
</tr>
<tr>
<td>1000</td>
<td>0.102</td>
<td>0.052</td>
<td>0.11</td>
<td>0.120</td>
<td>0.067</td>
<td>0.017</td>
</tr>
<tr>
<td>3000</td>
<td>0.098</td>
<td>0.050</td>
<td>0.010</td>
<td>0.146</td>
<td>0.086</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Panel A: \( T = 36 \)

<table>
<thead>
<tr>
<th>N</th>
<th>0.088</th>
<th>0.044</th>
<th>0.007</th>
<th>0.128</th>
<th>0.073</th>
<th>0.025</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>( t(\hat{\gamma}_0) )</td>
<td>( t(\hat{\gamma}_1) )</td>
<td>( t(\hat{\gamma}_0) )</td>
<td>( t(\hat{\gamma}_1) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.097</td>
<td>0.049</td>
<td>0.009</td>
<td>0.111</td>
<td>0.059</td>
<td>0.013</td>
</tr>
<tr>
<td>1000</td>
<td>0.103</td>
<td>0.052</td>
<td>0.009</td>
<td>0.106</td>
<td>0.057</td>
<td>0.012</td>
</tr>
<tr>
<td>3000</td>
<td>0.102</td>
<td>0.046</td>
<td>0.009</td>
<td>0.121</td>
<td>0.065</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Panel B: \( T = 72 \)

The table presents the size properties of t-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (\( N \)) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1-2013:12. \( t(\cdot) \) denotes the t-statistic associated with the OLS estimator as derived in Theorem 3.3. The t-statistics are compared with the critical values from a standard normal distribution.
### Table 3.9
Empirical size of \(F\)-tests in a one-factor model with a useless factor (\(\Sigma\) scalar)

<table>
<thead>
<tr>
<th>(N)</th>
<th>Panel A: (T = 36)</th>
<th>Panel B: (T = 72)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.10  0.05  0.01</td>
<td>0.10  0.05  0.01</td>
</tr>
<tr>
<td>100</td>
<td>0.107 0.056 0.012</td>
<td>0.108 0.056 0.012</td>
</tr>
<tr>
<td>500</td>
<td>0.101 0.052 0.011</td>
<td>0.104 0.053 0.011</td>
</tr>
<tr>
<td>1000</td>
<td>0.101 0.051 0.011</td>
<td>0.101 0.052 0.010</td>
</tr>
<tr>
<td>3000</td>
<td>0.100 0.049 0.010</td>
<td>0.101 0.051 0.010</td>
</tr>
</tbody>
</table>

The table presents the size properties of \(F\)-tests of statistical significance. The null hypothesis is that all the \(K\) parameters, except for the intercept, are all equal to zero in the (fitted) model \(\hat{R} = \hat{X}_t\hat{\beta}^\prime + \eta_t\), and where the return generating process is given by \(R_t = \gamma_01_N + \epsilon_t\). The fitted Model is a One-Factor Model \(R_t = a_t + b_t'g_t + u_t\), where \(g_t\) is the excess market return (from Kenneth French's website) from January 2008 to December 2010 for \(T=36\), and the excess market return from January 2008 to December 2013 for \(T=72\). The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (\(N\)) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12. The \(F\)-statistics are compared with the critical values from a \(\chi^2_{K}(\frac{24}{K})\) distribution as derived in Theorem 3.3.

### Table 3.10
Empirical size of \(F\)-tests in a one-factor model with a useless factor (\(\Sigma\) Diagonal)

<table>
<thead>
<tr>
<th>(N)</th>
<th>Panel A: (T = 36)</th>
<th>Panel B: (T = 72)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.10  0.05  0.01</td>
<td>0.10  0.05  0.01</td>
</tr>
<tr>
<td>100</td>
<td>0.113 0.060 0.015</td>
<td>0.134 0.077 0.022</td>
</tr>
<tr>
<td>500</td>
<td>0.101 0.053 0.011</td>
<td>0.108 0.057 0.013</td>
</tr>
<tr>
<td>1000</td>
<td>0.102 0.052 0.011</td>
<td>0.106 0.053 0.011</td>
</tr>
<tr>
<td>3000</td>
<td>0.102 0.050 0.011</td>
<td>0.101 0.052 0.011</td>
</tr>
</tbody>
</table>

The table presents the size properties of \(F\)-tests of statistical significance. The null hypothesis is that all the \(K\) parameters, except for the intercept, are all equal to zero in the (fitted) model \(\hat{R} = \hat{X}_t\hat{\beta}^\prime + \eta_t\), and where the return generating process is given by \(R_t = \gamma_01_N + \epsilon_t\). The fitted Model is a One-Factor Model \(R_t = a_t + b_t'g_t + u_t\), where \(g_t\) is the excess market return (from Kenneth French's website) from January 2008 to December 2010 for \(T=36\), and the excess market return from January 2008 to December 2013 for \(T=72\). The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (\(N\)) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12. The \(F\)-statistics are compared with the critical values from a \(\chi^2_{K}(\frac{24}{K})\) distribution as derived in Theorem 3.3.
Table 3.11
Empirical size of F-tests in a one-factor model with a useless factor (Σ Full - δ = 0.5)

<table>
<thead>
<tr>
<th>N</th>
<th>Panel A: T = 36</th>
<th>Panel B: T = 72</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
</tr>
<tr>
<td>100</td>
<td>0.113 0.060 0.014</td>
<td>0.138 0.080 0.030</td>
</tr>
<tr>
<td>500</td>
<td>0.101 0.053 0.011</td>
<td>0.106 0.055 0.011</td>
</tr>
<tr>
<td>1000</td>
<td>0.103 0.053 0.011</td>
<td>0.101 0.052 0.014</td>
</tr>
<tr>
<td>3000</td>
<td>0.101 0.051 0.011</td>
<td>0.097 0.049 0.012</td>
</tr>
</tbody>
</table>

The table presents the size properties of F-tests of statistical significance. The null hypothesis is that all the K parameters, except for the intercept, are all equal to zero in the (fitted) model $\bar{R} = \bar{X}_t\beta + \eta^F_t$, and where the return generating process is given by $R_t = \gamma_01_N + \epsilon_t$. The fitted Model is a One-Factor Model $R_{it} = \alpha_i + b'_ig_t + u_{it}$, where $g_t$ is the excess market return (from Kenneth French’s website) from January 2008 to December 2010 for T=36, and the excess market return from January 2008 to December 2013 for T=72. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12. The F-statistics are compared with the critical values from a $\chi^2_K (\frac{df}{K})$ distribution as derived in Theorem 3.3.

Table 3.12
Empirical size of F-tests in a one-factor model with a useless factor (Σ Full - δ = 0.25)

<table>
<thead>
<tr>
<th>N</th>
<th>Panel A: T = 36</th>
<th>Panel B: T = 72</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
</tr>
<tr>
<td>100</td>
<td>0.117 0.062 0.015</td>
<td>0.136 0.082 0.028</td>
</tr>
<tr>
<td>500</td>
<td>0.114 0.061 0.014</td>
<td>0.108 0.058 0.014</td>
</tr>
<tr>
<td>1000</td>
<td>0.119 0.067 0.017</td>
<td>0.110 0.057 0.013</td>
</tr>
<tr>
<td>3000</td>
<td>0.148 0.086 0.025</td>
<td>0.123 0.065 0.016</td>
</tr>
</tbody>
</table>

The table presents the size properties of F-tests of statistical significance. The null hypothesis is that all the K parameters, except for the intercept, are all equal to zero in the (fitted) model $\bar{R} = \bar{X}_t\beta + \eta^F_t$, and where the return generating process is given by $R_t = \gamma_01_N + \epsilon_t$. The fitted Model is a One-Factor Model $R_{it} = \alpha_i + b'_ig_t + u_{it}$, where $g_t$ is the excess market return (from Kenneth French’s website) from January 2008 to December 2010 for T=36, and the excess market return from January 2008 to December 2013 for T=72. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12. The F-statistics are compared with the critical values from a $\chi^2_K (\frac{df}{K})$ distribution as derived in Theorem 3.3.
3.5.2 Correctly specified models with useful and useless factors

In this section we evaluate the finite sample properties of the OLS estimator under the assumption of correct model specification, when the true asset-pricing specification is assumed to follow a linear multi-factor model. In particular, under the null of correctly specified models, we generate asset returns from the following model:

\[ R_t = \gamma_0 1_N + B_f (\gamma_1 + f_t - E[f]) + \epsilon_t \]  

(3.54)

where \( \gamma_0 \) and \( \gamma_1 \) are calibrated using the OLS estimates from a one factor model, where \( f_t \) represents the excess market return. The error term, \( \epsilon_t \), is generated from a multivariate normal distribution, using the same setting of the previous section. Then, we estimate the model using a two-factor model

\[ R_t = \alpha + B_f f_t + B_g g_t + \epsilon_t \]  

(3.55)

where \( g_t \) is the useless factor, orthogonal to \( f_t \).

Tables 3.13-3.15 report the percentage bias (Bias) and the root mean squared error (RMSE) of the risk premia estimator for \( \gamma_0 \), the useful factor \( \gamma_1 \), and the useless factor \( \gamma_2 \). We also report the bias of the R-squared \( (R^2) \) that uses the asymptotic results derived in Theorem 3.7(ii). Overall, we find that the OLS estimator exhibits a slightly higher bias and RMSE compared to the results of Section 3.5.1, even though it continues to perform quite well for \( N \geq 500 \). The absolute bias of the risk premium associated to the useless factors \( (\gamma_2) \) is always null, for all the dimensions of \( N \) and \( T \), and regardless of the degree of cross-sectional dependence in the model disturbances. The bias of the R-squared is approximately 4\% when \( N = 100 \) and it sensibly reduces to less than 1\% when \( N \geq 500 \).

In Tables 3.16-3.18 we finally report the empirical rejection rates of the \( t \)-test of statistical significance, which is based on the asymptotic distribution of our Theorem 3.6(ii). The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for the 10\%, 5\%, and 1\% level of significance and are compared with the critical values of a standard normal distribution. We find that our \( t \)-test performs very well for all \( N \) and \( T \), not only when testing the significance of the risk premium associated to the useful factor, but also (and more importantly) when testing the risk premium corresponding to the useless factor. This result confirms that the OLS estimator represents a very powerful tool if we one wishes to detect spurious factors in linear asset-pricing models.
Table 3.13
Bias and RMSE of the OLS Estimator in a correctly specified model with useful and useless factors (∑ scalar).

<table>
<thead>
<tr>
<th>Statistics</th>
<th>N = 100</th>
<th>N = 500</th>
<th>N = 1000</th>
<th>N = 3000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Panel A: T = 36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_0$)</td>
<td>0.78%</td>
<td>0.06%</td>
<td>-0.15%</td>
<td>0.10%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_0$)</td>
<td>0.291</td>
<td>0.132</td>
<td>0.071</td>
<td>0.047</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_{1_f}$)</td>
<td>0.43%</td>
<td>0.07%</td>
<td>0.08%</td>
<td>-0.03%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_{1_f}$)</td>
<td>0.211</td>
<td>0.102</td>
<td>0.053</td>
<td>0.039</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_{1_p}$)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_{1_p}$)</td>
<td>1.769</td>
<td>0.766</td>
<td>0.543</td>
<td>0.326</td>
</tr>
<tr>
<td>Bias($R^2$)</td>
<td>4.66%</td>
<td>1.62%</td>
<td>0.38%</td>
<td>0.22%</td>
</tr>
<tr>
<td></td>
<td>Panel B: T = 72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_0$)</td>
<td>0.10%</td>
<td>0.05%</td>
<td>0.06%</td>
<td>0.07%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_0$)</td>
<td>0.582</td>
<td>0.195</td>
<td>0.079</td>
<td>0.055</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_{1_f}$)</td>
<td>0.27%</td>
<td>0.08%</td>
<td>0.11%</td>
<td>0.03%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_{1_f}$)</td>
<td>0.278</td>
<td>0.125</td>
<td>0.052</td>
<td>0.033</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_{1_p}$)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_{1_p}$)</td>
<td>1.215</td>
<td>0.540</td>
<td>0.376</td>
<td>0.227</td>
</tr>
<tr>
<td>Bias($R^2$)</td>
<td>4.00%</td>
<td>1.57%</td>
<td>0.65%</td>
<td>0.39%</td>
</tr>
</tbody>
</table>

The table reports the percentage bias (Bias) and root mean squared error (RMSE), over 10,000 simulated data sets. The return generating process is given by $R_t = \gamma_0 1_N + B_f(\gamma_1 + f_t - E[f]|\epsilon_t)$, where $\epsilon_t \sim N(0, \sigma^2 I_T)$ and where we calibrate $\gamma_0$ and $\gamma_1$ as the OLS estimates from the one factor model (CAPM). The fitted Model is a Two-Factor Model $R_t = \alpha + B_f f_t + B_g y_t + \epsilon_t$, where $y_t$ is an orthogonal factor to $f_t$. The table also reports the bias and the RMSE of the $R$-squared ($R^2$) of the fitted model for different cross-sections of $N = 100$, 500, 1000, 3000 stocks. Notice that, before estimating both the true and the fitted model, all the factors are orthogonalized to each other such that $F'F = 0_{K_f \times K_f}$. 
### Table 3.14
Bias and RMSE of the OLS Estimator in a correctly specified model with useful and useless factors (Σ Diagonal).

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias($\hat{\gamma}_0$)</td>
<td>0.09%</td>
<td>0.06%</td>
<td>0.02%</td>
<td>0.04%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_0$)</td>
<td>0.052</td>
<td>0.034</td>
<td>0.029</td>
<td>0.021</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_{1,1}$)</td>
<td>0.11%</td>
<td>0.05%</td>
<td>0.03%</td>
<td>0.02%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_{1,1}$)</td>
<td>0.038</td>
<td>0.029</td>
<td>0.025</td>
<td>0.017</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_{1,2}$)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_{1,2}$)</td>
<td>1.820</td>
<td>1.203</td>
<td>0.958</td>
<td>0.543</td>
</tr>
<tr>
<td>Bias($R^2$)</td>
<td>2.30%</td>
<td>0.90%</td>
<td>0.23%</td>
<td>0.18%</td>
</tr>
</tbody>
</table>

Panel B: $T = 72$

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias($\hat{\gamma}_0$)</td>
<td>0.09%</td>
<td>0.03%</td>
<td>0.02%</td>
<td>0.02%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_0$)</td>
<td>0.036</td>
<td>0.032</td>
<td>0.029</td>
<td>0.023</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_{1,1}$)</td>
<td>0.13%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_{1,1}$)</td>
<td>0.032</td>
<td>0.023</td>
<td>0.019</td>
<td>0.017</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_{1,2}$)</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_{1,2}$)</td>
<td>1.807</td>
<td>0.922</td>
<td>0.653</td>
<td>0.392</td>
</tr>
<tr>
<td>Bias($R^2$)</td>
<td>3.13%</td>
<td>1.07%</td>
<td>0.19%</td>
<td>0.17%</td>
</tr>
</tbody>
</table>

The table reports the percentage bias (Bias) and root mean squared error (RMSE), over 10,000 simulated data sets. The return generating process is given by $R_t = \gamma_0 1_N + B_1(\gamma_{1,1} + f_t - E[f_t])\epsilon_t$, where $\epsilon_t \sim N(0, \Sigma)$ and where we calibrate $\gamma_0$ and $\gamma_{1,1}$ as the OLS estimates from the one factor model (CAPM). The fitted Model is a Two-Factor Model $R_t = \alpha + B_1 f_t + B_2 g_t + \epsilon_t$, where $g_t$ is an orthogonal factor to $f_t$. The table also reports the bias and the RMSE of the $R$-squared ($R^2$) of the fitted model for different cross-sections of $N = 100, 500, 1000, 3000$ stocks. Notice that, before estimating both the true and the fitted model, all the factors are orthogonalized to each other such that $F'G = 0_{K_f \times K_g}$.
Table 3.15
Bias and RMSE of the OLS Estimator in a correctly specified model with useful and useless factors ($\Sigma$ Full, $\delta = 0.5$).

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$N = 100$</th>
<th>$N = 500$</th>
<th>$N = 1000$</th>
<th>$N = 3000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: $T = 36$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_0$)</td>
<td>0.42%</td>
<td>0.38%</td>
<td>0.09%</td>
<td>0.06%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_0$)</td>
<td>0.086</td>
<td>0.042</td>
<td>0.035</td>
<td>0.021</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_{1_f}$)</td>
<td>0.06%</td>
<td>0.03%</td>
<td>0.04%</td>
<td>0.01%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_{1_f}$)</td>
<td>0.066</td>
<td>0.023</td>
<td>0.020</td>
<td>0.017</td>
</tr>
<tr>
<td>$\hat{\gamma}_s$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_s$)</td>
<td>1.211</td>
<td>0.916</td>
<td>0.903</td>
<td>0.543</td>
</tr>
<tr>
<td>$R^2$</td>
<td>2.90%</td>
<td>1.42%</td>
<td>0.49%</td>
<td>0.38%</td>
</tr>
<tr>
<td>Panel B: $T = 72$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_0$)</td>
<td>0.02%</td>
<td>0.04%</td>
<td>0.01%</td>
<td>0.02%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_0$)</td>
<td>0.052</td>
<td>0.044</td>
<td>0.030</td>
<td>0.023</td>
</tr>
<tr>
<td>Bias($\hat{\gamma}_{1_f}$)</td>
<td>0.04%</td>
<td>0.07%</td>
<td>0.02%</td>
<td>0.02%</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_{1_f}$)</td>
<td>0.056</td>
<td>0.036</td>
<td>0.028</td>
<td>0.021</td>
</tr>
<tr>
<td>$\hat{\gamma}_s$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSE($\hat{\gamma}_s$)</td>
<td>1.872</td>
<td>0.864</td>
<td>0.653</td>
<td>0.393</td>
</tr>
<tr>
<td>$R^2$</td>
<td>2.19%</td>
<td>1.29%</td>
<td>0.46%</td>
<td>0.29%</td>
</tr>
</tbody>
</table>

The table reports the percentage bias (Bias) and root mean squared error (RMSE), over 10,000 simulated data sets. The return generating process is given by $R_t = \gamma_0 1_N + B_f(\gamma_{1_f} + f_t - E[f])\epsilon_t$, where $\epsilon_t \sim N(0, \Sigma)$ and where we calibrate $\gamma_0$ and $\gamma_{1_f}$ as the OLS estimates from the one factor model (CAPM). The fitted Model is a Two-Factor Model $R_t = \alpha + B_f f_t + B_s g_t + \epsilon_t$, where $g_t$ is an orthogonal factor to $f_t$. The table also reports the bias and the RMSE of the $R$-squared ($R^2$) of the fitted model for different cross-sections of $N = 100, 500, 1000, 3000$ stocks. Notice that, before estimating both the true and the fitted model, all the factors are orthogonalized to each other such that $F'G = 0_{K_f \times K_s}$.
Table 3.16
Empirical Size of t-tests in a correctly specified model with useful and useless factors (Σ Scalar)

<table>
<thead>
<tr>
<th></th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: T = 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(\tilde{\gamma}_0)</td>
<td>t(\tilde{\gamma}_{11})</td>
<td>t(\tilde{\gamma}_{12})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.102</td>
<td>0.049</td>
<td>0.011</td>
<td>0.099</td>
<td>0.053</td>
<td>0.009</td>
<td>0.100</td>
<td>0.053</td>
<td>0.012</td>
</tr>
<tr>
<td>500</td>
<td>0.102</td>
<td>0.052</td>
<td>0.009</td>
<td>0.098</td>
<td>0.048</td>
<td>0.008</td>
<td>0.099</td>
<td>0.051</td>
<td>0.012</td>
</tr>
<tr>
<td>1000</td>
<td>0.100</td>
<td>0.051</td>
<td>0.011</td>
<td>0.098</td>
<td>0.048</td>
<td>0.009</td>
<td>0.100</td>
<td>0.051</td>
<td>0.010</td>
</tr>
<tr>
<td>3000</td>
<td>0.099</td>
<td>0.050</td>
<td>0.010</td>
<td>0.101</td>
<td>0.052</td>
<td>0.010</td>
<td>0.099</td>
<td>0.049</td>
<td>0.010</td>
</tr>
<tr>
<td>Panel B: T = 72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(\tilde{\gamma}_0)</td>
<td>t(\tilde{\gamma}_{11})</td>
<td>t(\tilde{\gamma}_{12})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.097</td>
<td>0.049</td>
<td>0.009</td>
<td>0.101</td>
<td>0.048</td>
<td>0.010</td>
<td>0.110</td>
<td>0.054</td>
<td>0.012</td>
</tr>
<tr>
<td>500</td>
<td>0.103</td>
<td>0.051</td>
<td>0.013</td>
<td>0.096</td>
<td>0.048</td>
<td>0.009</td>
<td>0.098</td>
<td>0.048</td>
<td>0.010</td>
</tr>
<tr>
<td>1000</td>
<td>0.099</td>
<td>0.052</td>
<td>0.010</td>
<td>0.095</td>
<td>0.049</td>
<td>0.010</td>
<td>0.098</td>
<td>0.048</td>
<td>0.011</td>
</tr>
<tr>
<td>3000</td>
<td>0.101</td>
<td>0.052</td>
<td>0.010</td>
<td>0.102</td>
<td>0.052</td>
<td>0.010</td>
<td>0.100</td>
<td>0.049</td>
<td>0.009</td>
</tr>
</tbody>
</table>

The table presents the size properties of t-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008-1-2013-12. t(\cdot) denotes the t-statistic associated with the OLS estimator based on the asymptotic distribution derived in Theorem 3.6. The t-statistics are compared with the critical values from a standard normal distribution.
Table 3.17
Empirical Size of $t$-tests in a correctly specified model with useful and useless factors ($\Sigma$ Diagonal)

<table>
<thead>
<tr>
<th>N</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t(\hat{\gamma}_0)$</td>
<td>$t(\hat{\gamma}_{\delta_1})$</td>
<td>$t(\hat{\gamma}_{\delta_2})$</td>
<td>$t(\hat{\gamma}_0)$</td>
<td>$t(\hat{\gamma}_{\delta_1})$</td>
<td>$t(\hat{\gamma}_{\delta_2})$</td>
<td>$t(\hat{\gamma}_0)$</td>
<td>$t(\hat{\gamma}_{\delta_1})$</td>
<td>$t(\hat{\gamma}_{\delta_2})$</td>
</tr>
<tr>
<td>100</td>
<td>0.101</td>
<td>0.051</td>
<td>0.011</td>
<td>0.109</td>
<td>0.068</td>
<td>0.016</td>
<td>0.112</td>
<td>0.056</td>
<td>0.016</td>
</tr>
<tr>
<td>500</td>
<td>0.101</td>
<td>0.051</td>
<td>0.011</td>
<td>0.084</td>
<td>0.045</td>
<td>0.009</td>
<td>0.106</td>
<td>0.049</td>
<td>0.009</td>
</tr>
<tr>
<td>1000</td>
<td>0.102</td>
<td>0.052</td>
<td>0.010</td>
<td>0.099</td>
<td>0.051</td>
<td>0.010</td>
<td>0.103</td>
<td>0.054</td>
<td>0.011</td>
</tr>
<tr>
<td>3000</td>
<td>0.099</td>
<td>0.049</td>
<td>0.010</td>
<td>0.099</td>
<td>0.051</td>
<td>0.010</td>
<td>0.099</td>
<td>0.049</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Panel A: $T = 36$

Panel B: $T = 72$

The table presents the size properties of $t$-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks ($N$) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12. $t(\cdot)$ denotes the $t$-statistic associated with the OLS estimator based on the asymptotic distribution derived in Theorem 3.6. The $t$-statistics are compared with the critical values from a standard normal distribution.
Table 3.18
Empirical Size of t-tests in a correctly specified model with useful and useless factors (Σ Full - δ = 0.5)

<table>
<thead>
<tr>
<th>N</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t(γ₀)</td>
<td></td>
<td></td>
<td>t(γ₁₁)</td>
<td></td>
<td></td>
<td>t(γ₁₂)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.127</td>
<td>0.069</td>
<td>0.016</td>
<td>0.126</td>
<td>0.072</td>
<td>0.019</td>
<td>0.102</td>
<td>0.053</td>
<td>0.010</td>
</tr>
<tr>
<td>500</td>
<td>0.107</td>
<td>0.055</td>
<td>0.014</td>
<td>0.110</td>
<td>0.054</td>
<td>0.013</td>
<td>0.102</td>
<td>0.052</td>
<td>0.009</td>
</tr>
<tr>
<td>1000</td>
<td>0.104</td>
<td>0.052</td>
<td>0.012</td>
<td>0.103</td>
<td>0.049</td>
<td>0.008</td>
<td>0.100</td>
<td>0.050</td>
<td>0.010</td>
</tr>
<tr>
<td>3000</td>
<td>0.099</td>
<td>0.048</td>
<td>0.010</td>
<td>0.099</td>
<td>0.051</td>
<td>0.010</td>
<td>0.100</td>
<td>0.050</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>t(γ₀)</td>
<td></td>
<td></td>
<td>t(γ₁₁)</td>
<td></td>
<td></td>
<td>t(γ₁₂)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.076</td>
<td>0.035</td>
<td>0.007</td>
<td>0.065</td>
<td>0.036</td>
<td>0.006</td>
<td>0.104</td>
<td>0.056</td>
<td>0.013</td>
</tr>
<tr>
<td>500</td>
<td>0.088</td>
<td>0.040</td>
<td>0.009</td>
<td>0.088</td>
<td>0.044</td>
<td>0.009</td>
<td>0.102</td>
<td>0.051</td>
<td>0.010</td>
</tr>
<tr>
<td>1000</td>
<td>0.095</td>
<td>0.047</td>
<td>0.009</td>
<td>0.096</td>
<td>0.046</td>
<td>0.008</td>
<td>0.102</td>
<td>0.051</td>
<td>0.010</td>
</tr>
<tr>
<td>3000</td>
<td>0.099</td>
<td>0.048</td>
<td>0.010</td>
<td>0.099</td>
<td>0.049</td>
<td>0.010</td>
<td>0.099</td>
<td>0.048</td>
<td>0.010</td>
</tr>
</tbody>
</table>

The table presents the size properties of t-tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1-2013:12. \( t(\cdot) \) denotes the t-statistic associated with the OLS estimator based on the asymptotic distribution derived in Theorem 3.6. The t-statistics are compared with the critical values from a standard normal distribution.
3.6. Conclusion

In this paper we investigate the large cross-sectional properties of the standard two-pass methodology, when useless factors are included in the beta-pricing specification. In the conventional large-$T$ framework, the presence of a factor which is independent of all asset returns implies that risk premia parameters are no longer identified and, therefore, standard inference techniques become unreliable. Identification failure is further aggravated when the number of assets becomes large and/or the model is potentially misspecified.

We provide a simple statistical methodology to bypass this issue and show how it can be used to test for and detect useless factors. We assume a large cross-sectional setting, where the number of time series observations, $T$, is fixed and possibly very small, as it is often the case in empirical applications. The methodology is based on the simple two-pass CSR OLS estimator and allows us to derive correctly-sized $t$-ratios and $F$-tests for the null hypothesis that the risk premium of the useless factor(s) is equal to zero, together with a valid $R$-squared measure.

The results are also generalized to the case of potential model misspecification. We derive the asymptotic distribution of the CSR OLS estimator when the estimated model is not only including useless factors in the regression, but it is also omitting some useful factors. However, non trivial issues arise in this case, which could make inferential procedures unfeasible in practice. Indeed, all the asymptotic results depend on quantities related to the set of omitted factors, that need to be consistently estimated in order to conduct inference in practice.

Solution to such problems can be found and are left for future research. For example, appropriate bounds for the limiting distribution of the OLS estimator could be derived, in order to eliminate the problem of nuisance parameters. We can also implement a multi-step procedure, where we first extract all the relevant factors using PCA (see Giglio and Xiu (2017)) and then derive inference for the useless factors based on the PCA distribution. This would allow us to estimating, testing and detecting spurious factors even under the presence of omitted factors. Our methodology needs also to be implemented empirically, in order to identify a set of robust factors and those that suffer from spuriousness problems. Finally, results for the large-$N$ and large-$T$ setting could be derived, in order to provide statistical strategies to be used in high-dimensional frameworks.
3.A. Assumptions

Assumption 3.1 As $N \to \infty$, 
\[
\frac{1}{N} \sum_{i=1}^{N} \beta_i \to \mu_\beta \quad \text{and} \quad \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' \to \Sigma_\beta, \quad (3.4.1)
\]
such that the matrix 
\[
\Sigma_\beta = \begin{bmatrix} 1 & \mu_{\beta} \\ \mu_{\beta} & \Sigma_{\beta} \end{bmatrix}
\quad (3.4.2)
\]
is positive-definite

Assumption 3.2 The vector $\epsilon_t$ is independently and identically distributed (i.i.d.) over time with 
\[
E[\epsilon_t | F] = 0_N 
\quad (3.4.3)
\]
and a positive-definite matrix, 
\[
\text{Var}[\epsilon_t | F] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_N^2 \end{pmatrix} = \Sigma, \quad (3.4.4)
\]
where $0_N$ is a $N$-vector of zeros, and $\sigma_{ij}$ denotes the $(i,j)$-th element of $\Sigma$, for every $i,j = 1, \ldots, N$ with $\sigma_{ii}^2 = \sigma_{ii}$.

Assumption 3.3 $E[f_t]$ does not vary over time. Moreover, $F_t' F_t$ is a positive-definite matrix for every $T \geq K$.

Assumption 3.4 As $N \to \infty$, 

(i) 
\[
\frac{1}{N} \sum_{i=1}^{N} (\sigma_i^2 - \sigma^2) = o \left( \frac{1}{\sqrt{N}} \right), \quad (3.4.5)
\]
for some $0 < \sigma^2 < \infty$.

(ii) 
\[
\sum_{i,j=1}^{N} |\sigma_{ij}| \mathbb{I}_{(i \neq j)} = o(N), \quad (3.4.6)
\]
where $\mathbb{I}_{(i \neq j)}$ denotes the indicator function.

(iii) 
\[
\frac{1}{N} \sum_{i=1}^{N} \mu_{4i} \to \mu_4, \quad (3.4.7)
\]
for some $0 < \mu_4 < \infty$ where $\mu_{4i} = E[\epsilon_i^4]$.

(iv) 
\[
\frac{1}{N} \sum_{i=1}^{N} \sigma_i^4 \to \sigma_4, \quad (3.4.8)
\]
for some $0 < \sigma_4 < \infty$. 

(vi) \[
\sup_i \mu_{4i} \leq C < \infty,
\] (3.A.9)

for a generic constant \(C\).

(vii) \[
E[\kappa^4_{11}] = 0.
\] (3.A.10)

(viii) \[
\frac{1}{N} \sum_{i=1}^{N} \kappa_{4,iii} \rightarrow \kappa_4,
\] (3.A.11)

for some \(0 \leq |\kappa_4| < \infty\), where \(\kappa_{4,iii} = \kappa_4(\varepsilon_{it}, \varepsilon_{it}, \varepsilon_{it}, \varepsilon_{it})\) denotes the fourth-order cumulant of the residuals \(\{\varepsilon_{it}, \varepsilon_{it}, \varepsilon_{it}, \varepsilon_{it}\}\).

(viii) For every \(3 \leq h \leq 8\), all the mixed cumulants of order \(h\) satisfy

\[
\sup_{N} \sum_{i_2, \ldots, i_h=1}^{N} |\kappa_{h,i_1i_2,\ldots,i_h}| = o(N),
\] (3.A.12)

for at least one \(i_j\) \((2 \leq j \leq h)\) different from \(i_1\).

**Assumption 3.5** As \(N \rightarrow \infty\), we have

(i) \[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \varepsilon_i \xrightarrow{d} \mathcal{N}(0_T, \sigma^2 I_T).
\] (3.A.13)

(ii) \[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\varepsilon_i \varepsilon_i' - \sigma^2 I_T) \xrightarrow{d} \mathcal{N}(0 \tau^2, U_c).
\] (3.A.14)

(iii) For a generic \(T\)-vector \(C_T\), \[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( C_T' \otimes \left( \frac{1}{\beta_i} \right) \right) \varepsilon_i \xrightarrow{d} \mathcal{N}(0_{K+1}, V_c),
\] (3.A.15)

where \(V_c = \sigma^2 \Sigma_X\) and \(c = C_T' C_T\). In particular, \(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (C_T' \otimes \beta_i) \varepsilon_i \xrightarrow{d} \mathcal{N}(0_K, V_c^\dagger)\), where \(V_c^\dagger = \sigma^2 \Sigma_\beta\).

**Assumption 3.6** Let \(c = (c_1, \ldots, c_N)'\) and assume that \(E[R_t] = c\). Then

(i) \[
\frac{1}{N} \sum_{i=1}^{N} (\varepsilon_i - \mu_c) = o \left( \frac{1}{\sqrt{N}} \right)
\] (3.A.16)

for some \(\mu_c < \infty\).

(ii) \[
\frac{1}{N} \sum_{i=1}^{N} \varepsilon_i c_i = O_p \left( \frac{1}{\sqrt{N}} \right)
\] (3.A.17)
3.B. Lemmas

Lemma 3.1 Under Assumptions 3.2-3.4,
\[
\hat{\sigma}^2 - \sigma^2 = O_p \left( \frac{1}{\sqrt{N}} \right). \tag{3.B.1}
\]

Proof. Rewrite \( \hat{\sigma}^2 - \sigma^2 \) as
\[
\hat{\sigma}^2 - \sigma^2 = \left( \hat{\sigma}^2 - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right) + \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 - \sigma^2 \right)
= \left( \hat{\sigma}^2 - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right) + o \left( \frac{1}{\sqrt{N}} \right) \tag{3.B.2}
\]
by Assumption 3.5(i). Moreover,
\[
\hat{\sigma}^2 - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 = \frac{\text{tr} \left( M \epsilon' \right)}{N(T-K-1)} = \frac{\text{tr} \left( M \right)}{T-K-1} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2
= \frac{\text{tr} \left( P \left( \sum_{i=1}^{N} \sigma_i^2 I_T - \epsilon \epsilon' \right) \right)}{N(T-K-1)} + \frac{\text{tr} \left( \epsilon \epsilon' - T \sum_{i=1}^{N} \sigma_i^2 \right)}{N(T-K-1)}. \tag{3.B.3}
\]
As for the second term on the right-hand side of Eq. (3.B.3), we have
\[
\frac{\text{tr} \left( \epsilon \epsilon' - T \sum_{i=1}^{N} \sigma_i^2 \right)}{N(T-K-1)} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (\epsilon_{it}^2 - \sigma_i^2)}{N(T-K-1)} = O_p \left( \frac{1}{\sqrt{N(T-K-1)}} \right) \tag{3.B.4}
\]
As for the first term on the right-hand side of Eq. (3.B.3), we have
\[
\frac{\text{tr} \left( P \left( \sum_{i=1}^{N} \sigma_i^2 I_T - \epsilon \epsilon' \right) \right)}{N(T-K-1)} = \frac{\sum_{t=1}^{T} d_t \left( D' D \right)^{-1} D' \left( \sum_{i=1}^{N} \sigma_i^2 u_{i,t} - \sum_{i=1}^{N} \epsilon_i \epsilon_i \right)}{N(T-K-1)}
= \frac{\sum_{t=1}^{T} p_t \left( \sum_{i=1}^{N} \sigma_i^2 u_{i,t} - \sum_{i=1}^{N} \epsilon_i \epsilon_i \right)}{N(T-K-1)}, \tag{3.B.5}
\]
where \( u_{i,t} \) is a \( T \)-vector with one in the \( t \)-th position and zeros elsewhere, \( d_t \) is the \( t \)-th row of \( D = [1_T, \ F] \), and \( p_t = d_t \left( D' D \right)^{-1} D' \). Since Eq. (3.B.5) has a zero mean, we only need to consider its variance to determine the rate of convergence. We have
\[
\text{Var} \left( \frac{\sum_{t=1}^{T} p_t \left( \sum_{i=1}^{N} \sigma_i^2 u_{i,t} - \sum_{i=1}^{N} \epsilon_i \epsilon_i \right)}{N(T-K-1)} \right)
= \frac{1}{N^2(T-K-1)^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} p_t \left( \sigma_i^2 u_{i,t} - \epsilon_i \epsilon_i \right) \left( \sigma_j^2 u_{j,s} - \epsilon_j \epsilon_j \right)^t \frac{p_s^t}{p_s}
= \frac{1}{N^2(T-K-1)^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} p_t E \left[ \left( \sigma_i^2 u_{i,t} - \epsilon_i \epsilon_i \right) \left( \sigma_j^2 u_{j,s} - \epsilon_j \epsilon_j \right)^t \right] p_s \tag{3.B.6}
\]
Moreover, we have

\[
E \left[ (\sigma_{it}^2 u_{it} - \epsilon_i \epsilon_i) (\sigma_{it}^2 u_{it} - \epsilon_j \epsilon_j) \right] = E \left[ \sigma_{it}^2 u_{it}^2 + \epsilon_i \epsilon_j \epsilon_i \epsilon_j - \sigma_{it}^2 \epsilon_i \epsilon_i \epsilon_i \epsilon_i \right]
\]

\[
\begin{cases}
\mu_{ii} u_{it} u_{ts} + \sigma_{it}^2 (I_{t} - 2u_{it} u_{ts}) & \text{if } i = j, \ t = s \\
\kappa_{iijj} u_{it} u_{ts} + \sigma_{ij}^2 (I_{t} + u_{it} u_{ts}) & \text{if } i \neq j, \ t = s \\
\sigma_{it}^2 u_{it} u_{ts} & \text{if } i = j, \ t \neq s \\
\sigma_{ij}^2 u_{it} u_{ts} & \text{if } i \neq j, \ t \neq s.
\end{cases}
\] (3.8.7)

It follows that

\[
\begin{align*}
\text{Var} \left( \frac{\sum_{t=1}^{T} p_{t} (\sum_{i=1}^{N} \sigma_{it}^2 u_{it} - \sum_{i=1}^{N} \epsilon_i \epsilon_i)}{N(T-K-1)} \right) \\
&= \frac{1}{N^2(T-K-1)^2} \sum_{t=1}^{T} \sum_{i=1}^{N} p_{t} (\mu_{ii} u_{it} u_{ts} + \sigma_{it}^2 (I_{t} - 2u_{it} u_{ts})) p_{t} \\
&\quad + \frac{1}{N^2(T-K-1)^2} \sum_{t=1}^{T} \sum_{i \neq j} \sum_{t \neq s} p_{t} (\kappa_{iiij} u_{it} u_{ts} + \sigma_{ij}^2 (I_{t} + u_{it} u_{ts})) p_{t} \\
&\quad + \frac{1}{N^2(T-K-1)^2} \sum_{i=1}^{N} \sum_{t \neq s} p_{t} (\sigma_{it}^2 u_{it} u_{ts} p_{t} \sigma_{it}^2 u_{it} u_{ts} p_{t}) \\
&\quad + \frac{1}{N^2(T-K-1)^2} \sum_{i \neq j} \sum_{t \neq s} p_{t} \sigma_{ij}^2 u_{it} u_{ts} p_{t} \sigma_{ij}^2 u_{it} u_{ts} p_{t} \\
&= O \left( \frac{1}{N} \right) \quad (3.8.8)
\end{align*}
\]

by Assumptions 3.4(ii), 3.4(iii), 3.4(iv), and 3.4(viii), which implies that the first term on the right-hand side of Eq. (3.8.3) is \( O_p \left( \frac{1}{\sqrt{N}} \right) \). Putting the pieces together concludes the proof. \( \blacksquare \)

**Lemma 3.2** Let

\[
\Lambda = \begin{bmatrix} 0 & 0_{K}^T \\ 0_{K} & \sigma^2 (\hat{F}' \hat{F})^{-1} \end{bmatrix}.
\] (3.9.9)

(i) Under Assumptions 3.1-3.4,

\[
\hat{X} \hat{X} = O_p(N).
\] (3.10)

In addition, under Assumption 3.5,

(ii) \( \hat{\Sigma}_{XX} \to_p \Sigma_{XX} + \Lambda \),

and

(iii) \( \frac{\hat{X} - X}'(\hat{X} - X) \to_p \Lambda \).

(3.11) (3.12)
Proof.

(i) Consider

\[ \hat{X}' \hat{X} = \begin{bmatrix} N & 1_N \hat{B} \\ \hat{B}'1_N & \hat{B}'\hat{B} \end{bmatrix}. \] (3.B.13)

Then,

\[ \hat{B}'1_N = \sum_{i=1}^{N} \hat{\beta}_i = \sum_{i=1}^{N} \beta_i + P' \sum_{i=1}^{N} \epsilon_i. \] (3.B.14)

Under Assumptions 3.3-3.4,

\[ \text{Var} \left( \sum_{t=1}^{T} \sum_{i=1}^{N} \epsilon_{it}(f_t - \bar{f}) \right) = \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} (f_t - \bar{f})(f_s - \bar{f})'E[\epsilon_{it}\epsilon_{js}] \]

\[ \leq \sum_{t=1}^{T} \sum_{i,j=1}^{N} (f_t - \bar{f})(f_t - \bar{f})'|\sigma_{ij}| \]

\[ = O \left( N\sigma^2 \sum_{t=1}^{T} (f_t - \bar{f})(f_t - \bar{f})' \right) = O(NT). \] (3.B.15)

Using Assumption 3.1, we have

\[ \hat{B}'1_N = O_p \left( N + \left( \frac{N}{T} \right)^{\frac{1}{2}} \right) = O_p(N). \] (3.B.16)

Next, consider

\[ \hat{B}'\hat{B} = \sum_{i=1}^{N} \hat{\beta}_i \hat{\beta}_i' \]

\[ = \sum_{i=1}^{N} (\beta_i + P'\epsilon_i) (\beta_i' + \epsilon_i'P) \]

\[ = \sum_{i=1}^{N} \beta_i \beta_i' + P' \left( \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) P + P' \left( \sum_{i=1}^{N} \epsilon_i \beta_i' \right) + \left( \sum_{i=1}^{N} \beta_i \epsilon_i' \right) P. \] (3.B.17)

By Assumption 3.1,

\[ \sum_{i=1}^{N} \beta_i \beta_i' = O(N). \] (3.B.18)

Using similar arguments as for Eq. (3.B.15),

\[ P' \left( \sum_{i=1}^{N} \epsilon_i \beta_i' \right) = O_p \left( \left( \frac{N}{T} \right)^{\frac{1}{2}} \right) \] (3.B.19)

and

\[ \left( \sum_{i=1}^{N} \beta_i \epsilon_i' \right) P = O_p \left( \left( \frac{N}{T} \right)^{\frac{1}{2}} \right). \] (3.B.20)
For $P' \left( \sum_{i=1}^{N} \epsilon_{i} \epsilon_{i}' \right) P$, consider its central part and take the norm of its expectation. Using Assumptions 3.3-3.4,

$$
\left\| E \left[ \tilde{F}' \left( \sum_{i=1}^{N} \epsilon_{i} \epsilon_{i}' \right) \tilde{F} \right] \right\| = \left\| E \left[ \sum_{t,s=1}^{T} \sum_{i=1}^{N} (f_{t} - \tilde{f})(f_{s} - \tilde{f})' \epsilon_{it} \epsilon_{is} \right] \right\| \\
\leq \sum_{t,s=1}^{T} \sum_{i=1}^{N} \left\| (f_{t} - \tilde{f})(f_{s} - \tilde{f})' \right\| \left\| E \left[ \epsilon_{it} \epsilon_{is} \right] \right\| \\
= \sum_{t=1}^{T} \sum_{i=1}^{N} \left\| (f_{t} - \tilde{f})(f_{t} - \tilde{f})' \right\| \sigma_{t}^{2} \\
= O \left( N \sigma^{2} \sum_{t=1}^{T} \left\| (f_{t} - \tilde{f})(f_{t} - \tilde{f})' \right\| \right) = O(NT). \tag{3.21}
$$

Then, we have

$$
P' \left( \sum_{i=1}^{N} \epsilon_{i} \epsilon_{i}' \right) P = O_{p} \left( \frac{N}{T} \right) \tag{3.22}
$$

and

$$
\tilde{B}' \tilde{B} = O_{p} \left( N + \left( \frac{N}{T} \right)^{\frac{1}{2}} + \frac{N}{T} \right) = O_{p}(N). \tag{3.23}
$$

This concludes the proof of part (i).

(ii) Using part (i) and under Assumptions 3.2-3.5, we have

$$
N^{-1} \tilde{B}' 1_{N} = \frac{1}{N} \sum_{i=1}^{N} \beta_{i} + O_{p} \left( \frac{1}{\sqrt{N}} \right) \tag{3.24}
$$

and

$$
N^{-1} \tilde{B}' \tilde{B} = \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' + P' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_{i} \epsilon_{i}' \right) P + P' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_{i} \beta_{i}' \right) + \left( \frac{1}{N} \sum_{i=1}^{N} \beta_{i} \epsilon_{i}' \right) P \\
= \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' + P' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_{i} \epsilon_{i}' - \frac{1}{N} \sum_{i=1}^{N} \epsilon_{i}^{2} I_{T} \right) + \frac{1}{N} \sum_{i=1}^{N} \sigma_{i}^{2} I_{T} - \sigma^{2} I_{T} + \sigma^{2} I_{T} \right) P \\
+ P' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_{i} \beta_{i}' \right) + \left( \frac{1}{N} \sum_{i=1}^{N} \beta_{i} \epsilon_{i}' \right) P \\
= \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' + P' \left( \frac{1}{N} \sum_{i=1}^{N} \left( \epsilon_{i} \epsilon_{i}' - \epsilon_{i}^{2} I_{T} \right) \right) P + \frac{1}{N} \sum_{i=1}^{N} \left( \sigma_{i}^{2} - \sigma^{2} \right) P' P + \sigma^{2} P' P \\
+ P' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_{i} \beta_{i}' \right) + \left( \frac{1}{N} \sum_{i=1}^{N} \beta_{i} \epsilon_{i}' \right) P \\
= \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' + \sigma^{2} P' P + O_{p} \left( \frac{1}{\sqrt{N}} \right) + o \left( \frac{1}{\sqrt{N}} \right) + O_{p} \left( \frac{1}{\sqrt{N}} \right) + O_{p} \left( \frac{1}{\sqrt{N}} \right). \tag{3.25}
$$
Assumption 3.1 concludes the proof of part (ii).

(iii) Note that

\[
\frac{(\hat{X} - X)'(\hat{X} - X)}{N} = \frac{1}{N} \left[ \begin{array}{c} 0_N' \\ (B - B)' \end{array} \right] \left[ \begin{array}{c} 0_N \\ (B - B) \end{array} \right] \\
= \left[ \begin{array}{cc} 0 & 0_K' \\ 0_K & \mathcal{P}' \mathcal{P} \end{array} \right].
\]

(3.B.26)

As in part (ii) we can write

\[
\frac{ee'}{N} = \frac{1}{N} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma_i^2 I_T) + \left( \frac{1}{N} \sum_{i=1}^{N} (\sigma_i^2 - \sigma^2) \right) I_T + \sigma^2 I_T.
\]

(3.B.27)

Assumptions 3.4(i) and 3.5(ii) conclude the proof since

\[
\mathcal{P}' \frac{ee'}{N} \mathcal{P} = \sigma^2 \mathcal{P}' \mathcal{P} + O_p \left( \frac{1}{\sqrt{N}} \right) + o \left( \frac{1}{\sqrt{N}} \right).
\]

(3.B.28)

**Lemma 3.3**

*Under Assumptions 3.1-3.4,*

\[
X' \bar{\epsilon} = O_p \left( \sqrt{N} \right).
\]

(3.B.29)

**Proof.** We have

\[
X' \bar{\epsilon} = \frac{1}{T} \sum_{t=1}^{T} \left[ 1_N' B' \right] \epsilon_t
\]

(3.B.30)

and

\[
\text{Var} \left( \frac{1}{T} \sum_{t=1}^{T} 1_N' \epsilon_t \right) = \frac{1}{T^2} \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} E[\epsilon_i \epsilon_{js}] \\
\leq \frac{1}{T^2} \sum_{t=1}^{T} \sum_{i,j=1}^{N} |\sigma_{ij}| \\
= O \left( \frac{NT}{T^2} \sigma^2 \right) = O \left( N \right).
\]

(3.B.31)

Moreover, using Assumptions 3.1 and 3.4(ii),

\[
\text{Var} \left( \frac{1}{T} \sum_{t=1}^{T} B' \epsilon_t \right) = \frac{1}{T^2} \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} E[\epsilon_i \epsilon_{js}] \beta_i \beta_j' \\
\leq \frac{1}{T^2} \sum_{t=1}^{T} \sum_{i,j=1}^{N} |\beta_i \beta_j'| |\sigma_{ij}| \\
= O \left( \frac{NT}{T^2} \sigma^2 \right) = O \left( N \right).
\]

(3.B.32)

Putting the pieces together, \( X' \bar{\epsilon} = O_p \left( \sqrt{N} \right) \).
Lemma 3.4

Under Assumptions 3.2-3.4,

\[(\hat{X} - X)'X\Gamma^p = O_p\left(\sqrt{N}\right).\]  \hfill (3.B.33)

Proof. We have

\[(\hat{X} - X)'X\Gamma^p = \begin{bmatrix} 0' \\ P't \end{bmatrix} X\Gamma^p.\]  \hfill (3.B.34)

Using similar arguments to Eq. (3.B.15) concludes the proof. ■

Lemma 3.5

Under Assumptions 3.2-3.4,

\[(\hat{X} - X)'\hat{\epsilon} = O_p\left(\sqrt{N}\right).\]  \hfill (3.B.35)

Proof.

\[(\hat{X} - X)'\hat{\epsilon} = \begin{bmatrix} 0' \\ P't \end{bmatrix} \begin{bmatrix} 0 \\ \epsilon' \end{bmatrix} = \begin{bmatrix} 0' \\ P't \end{bmatrix} \begin{bmatrix} 0 \\ \epsilon' \begin{bmatrix} \frac{1}{T} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} P' \left[ \left( \epsilon \epsilon' - \sum_{i=1}^{N} \sigma_i^2 I_T \right) + \left( \sum_{i=1}^{N} \sigma_i^2 \right) I_T \right] \begin{bmatrix} \frac{1}{T} \end{bmatrix} \end{bmatrix} = O_p(\sqrt{N})\]  \hfill (3.B.36)

by Assumption 3.4. ■

Lemma 3.6 Under Assumption 3.4 and the identification assumption \(\kappa_4 = 0\), we have

\[\hat{\sigma}_4 \rightarrow_p \sigma_4.\]  \hfill (3.B.37)

Proof. We need to show that (i) \(E(\hat{\sigma}_4) \rightarrow \sigma_4\) and (ii) \(\text{Var}(\hat{\sigma}_4) = O\left(\frac{1}{N}\right)\).

(i) By Assumptions 3.4(iv), 3.4(vi), and 3.4(vii), we have

\[E \left[ \frac{1}{N} \sum_{i=1}^{T} \sum_{t=1}^{N} \hat{\epsilon}_{it}^2 \right] = \frac{1}{N} \sum_{i=1}^{T} \sum_{t=1}^{N} E[\hat{\epsilon}_{it}^2] \]

\[= \frac{1}{N} \sum_{i=1}^{T} \sum_{t=1}^{N} \sum_{s_1, s_2, s_3, s_4 = 1} m_{t_{s_1}} m_{t_{s_2}} m_{t_{s_3}} m_{t_{s_4}} E[\epsilon_{i s_1} \epsilon_{i s_2} \epsilon_{i s_3} \epsilon_{i s_4}] \]

\[= \frac{1}{N} \sum_{i=1}^{T} \sum_{t=1}^{N} \kappa_{4, i t t} \sum_{s=1}^{T} \frac{N}{N} \sum_{i=1}^{T} \sum_{s=1}^{N} \sigma_{i s}^2 \left( \sum_{s=1}^{T} \sum_{i=1}^{N} m_{i s}^2 \right)^2 \]

\[\rightarrow \kappa_4 \sum_{t=1}^{T} \sum_{s=1}^{T} m_{t s}^2 + 3\sigma_4 \sum_{t=1}^{T} \left( \sum_{s=1}^{T} m_{t s}^2 \right)^2,\]  \hfill (3.B.38)

where \(\hat{\epsilon}_{it} = i'_{t, T} M\epsilon_i\) and \(M = [m_{ts}]\) for \(t, s = 1, \ldots, T\). Note that

\[\sum_{s=1}^{T} m_{t s}^2 = ||m_t||^2\]

\[= i'_{t} M i_t\]

\[= \epsilon_t' \left( I_T - D(D'D)^{-1} D' \right) \epsilon_t\]

\[= 1 - \text{tr} \left( D(D'D)^{-1} D' i_t \epsilon' \right)\]

\[= 1 - \text{tr} \left( P_t i_t \epsilon' \right)\]

\[= 1 - p_{tt},\]  \hfill (3.B.39)
where \( p_{tt} \) is the \((t, t)\)-element of \( P \). Then, we have
\[
\sum_{t=1}^{T} \left( \sum_{s=1}^{T} m_{ts}^2 \right)^2 = \sum_{t=1}^{T} m_{tt}^2 = \text{tr} \left( M^{(2)} \right).
\]
(3.B.40)

By setting \( \kappa_4 = 0 \), it follows that
\[
E \left[ \hat{\sigma}_4 \right] \to \sigma_4.
\]
(3.B.41)

This concludes the proof of part (i).

(ii) As for the variance of \( \hat{\sigma}_4 \), we have
\[
\text{Var} \left( \frac{1}{N} \sum_{t=1}^{N} \sum_{t=1}^{T} \xi_{tt} \right) = \frac{1}{N^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \text{Cov} \left( \xi_{it}, \xi_{jt} \right)
\]
\[
= \frac{1}{N^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \sum_{u_1, u_2, v_1, v_2, u_3, u_4, v_3, v_4 = 1}^{T} m_{tu_1} m_{tu_2} m_{tu_3} m_{tu_4} m_{su_1} m_{su_2} m_{su_3} m_{su_4}
\]
\[
\times \text{Cov} \left( \xi_{tu_1}, \xi_{tu_2}, \xi_{tu_3}, \xi_{tu_4}, \xi_{sv_1}, \xi_{sv_2}, \xi_{sv_3}, \xi_{sv_4} \right)
\]
\[
= \frac{1}{N^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \sum_{u_1, u_2, v_1, v_2, u_3, u_4, v_3, v_4 = 1}^{T} m_{tu_1} m_{tu_2} m_{tu_3} m_{tu_4} m_{su_1} m_{su_2} m_{su_3} m_{su_4}
\]
\[
\times \left( \kappa_4 \left( \xi_{tu_1}, \xi_{tu_2}, \xi_{tu_3}, \xi_{tu_4}, \xi_{sv_1}, \xi_{sv_2}, \xi_{sv_3}, \xi_{sv_4} \right)
\right.
\]
\[
+ \sum \kappa_6 \left( \xi_{tu_1}, \xi_{tu_2}, \xi_{tu_3}, \xi_{tu_4}, \xi_{sv_1}, \xi_{sv_2} \right) \text{Cov} \left( \xi_{sv_1}, \xi_{sv_2} \right)
\]
\[
+ \sum \kappa_4 \left( \xi_{tu_1}, \xi_{tu_2}, \xi_{tsv_1}, \xi_{tsv_2} \right) \text{Cov} \left( \xi_{tsv_1}, \xi_{tsv_2} \right)
\]
\[
+ \sum \text{Cov} \left( \xi_{tsv_1}, \xi_{tsv_2} \right) \text{Cov} \left( \xi_{tsv_3}, \xi_{tsv_4} \right) \text{Cov} \left( \xi_{tsv_5}, \xi_{tsv_6} \right)
\]
\[
\left. \right) \right)
\]
(3.B.42)

where \( \kappa_4 (\cdot) \), \( \kappa_6 (\cdot) \), and \( \kappa_8 (\cdot) \) denote the fourth-, sixth-, and eighth-order mixed cumulants, respectively.

By \( \sum(\nu_1, \nu_2, \ldots, \nu_k) \) we denote the sum over all possible partitions of a group of \( K \) random variables into \( k \) subgroups of size \( \nu_1, \nu_2, \ldots, \nu_k \), respectively. As an example, consider \( \sum \left( 6, 2 \right) \). \( \sum (6, 2) \) defines the sum over all possible partitions of the group of eight random variables \( \{ \xi_{t11}, \xi_{t12}, \xi_{t13}, \xi_{t14}, \xi_{j11}, \xi_{j12}, \xi_{j13}, \xi_{j14} \} \) into two subgroups of size six and two, respectively. Moreover, since \( E \left[ \xi_{it} \right] = E \left[ \xi_{it}^2 \right] = 0 \), we do not need to consider further partitions in the relation above.\(^1\) Then, under Assumptions 3.4(i), 3.4(ii), 3.4(v), and 3.4(viii), it follows that
\[
\text{Var} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \xi_{it} \right) = O \left( \frac{1}{N} \right)
\]
(3.B.43)

\(^1\)According to the theory on cumulants (Brillinger (2001)), evaluation of \( \text{Cov} \left( \xi_{t11}, \xi_{t12}, \xi_{t13}, \xi_{t14}, \xi_{j11}, \xi_{j12}, \xi_{j13}, \xi_{j14} \right) \) requires considering the indecomposable partitions of the two sets, \( \{ \xi_{t11}, \xi_{t12}, \xi_{t13}, \xi_{t14} \} \) and \( \{ \xi_{j11}, \xi_{j12}, \xi_{j13}, \xi_{j14} \} \), meaning that there must be at least one subset that includes an element of both sets.
and \( \text{Var}(\hat{\sigma}_s) = O\left(\frac{1}{N}\right) \). This concludes the proof of part (ii). 

**Lemma 3.7** Under Assumptions 3.1-3.4 and assuming (3.29), then

\[
\frac{\hat{B}_f \hat{B}_s}{N} = O_p\left(\frac{1}{\sqrt{N}}\right)
\]

**Proof.** Using (3.C.94) and (3.C.95), under Assumption 3.4(i) and and imposing (3.29), we have

\[
\frac{\hat{B}_f \hat{B}_s}{N} = \frac{(B'_f + P'_f \epsilon)M_N(\epsilon' P_g)}{N}
\]

\[
= \frac{(B'_f + P'_f \epsilon)(I_N - \frac{1}{N}1_N'1_N)}{N}
\]

\[
= \frac{B'_f \epsilon' P_g}{N} + \frac{P'_f \epsilon' \epsilon P_g}{N} - \frac{B'_f 1_N' \epsilon' P_g}{N} - \frac{P'_f \epsilon' 1_N' \epsilon' P_g}{N}
\]

\[
= \frac{B'_f \epsilon' P_g}{N} + \frac{P'_f \left(\epsilon \epsilon' / N - \overline{\epsilon}^2 I_T\right) P_g}{N} - \frac{B'_f 1_N' \epsilon' P_g}{N} - \frac{P'_f \epsilon' 1_N' \epsilon' P_g}{N}
\]

\[
= O_p\left(\frac{1}{\sqrt{N}}\right) \]

**Lemma 3.8** Under Assumptions 3.1-3.4

\[
\frac{1}{N} \hat{\epsilon} \hat{\sigma}_s = O_p\left(\frac{1}{\sqrt{N}}\right)
\]

**Proof.** Using the fact that \( \hat{\epsilon}_s = P'_s \epsilon M_N \), and rewriting \( \hat{\epsilon} = \epsilon' \frac{1}{T} \), we have

\[
\frac{1}{N} \hat{\epsilon} \hat{\sigma}_s = \frac{1}{N} P'_s \epsilon \left(I_N - \frac{1}{N}1_N'1_N\right) \epsilon' \frac{1}{T}
\]

\[
= \frac{1}{N} P'_s \epsilon \frac{1}{T} - \frac{P'_s \epsilon 1_N1_N' \epsilon}{N} \frac{1}{T}
\]

\[
= P'_s \frac{1}{N} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \overline{\epsilon}^2) \frac{1}{T} + O_p\left(\frac{1}{N}\right) + o_p\left(\frac{1}{\sqrt{N}}\right)
\]

\[
= O_p\left(\frac{1}{\sqrt{N}}\right) \]

**Lemma 3.9** Under Assumptions 3.1-3.4

\[
\frac{1}{N} \hat{\epsilon} \hat{\sigma} = O_p\left(\frac{1}{\sqrt{N}}\right)
\]

**Proof.** Using the fact that \( \hat{\epsilon}_f = (B'_f + P'_f \epsilon)M_N \), and rewriting \( \hat{\epsilon} = \epsilon' \frac{1}{T} \), we have
\[
\frac{1}{N} \hat{B}_f' \epsilon = \frac{B_f' \epsilon' 1_T}{N} \frac{1}{T} - \frac{B_f' \epsilon' 1_T}{N} \frac{1}{T} N + \frac{P_f' \epsilon' 1_T}{N} \frac{1}{T} - \frac{P_f' \epsilon' 1_T}{N} \frac{1}{T} N - \frac{1}{N} \sum_{i=1}^{N} \epsilon_i' \frac{1_T}{T} - O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{\sqrt{N}} \right)
\]

\[
= \frac{B_f' \epsilon' 1_T}{N} \frac{1}{T} + \mu_f \epsilon' \frac{1_T}{T} N + P_f^'} \frac{1}{N} \frac{1}{T} \sum_{i=1}^{N} (\epsilon_i' - \sigma^2 1_T) \frac{1_T}{T} - O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{\sqrt{N}} \right)
\]

\[
= O_p \left( \frac{1}{\sqrt{N}} \right)
\]

Lemma 3.10 Under Assumptions 3.1-3.4

\[
\frac{1}{N} \hat{B}_f' \epsilon' P_f = \sigma^2 (\hat{F}^* \hat{F})^{-1} + O_p \left( \frac{1}{\sqrt{N}} \right)
\]

Proof. Using the fact that \(\hat{B}_f' = (B_f' + \sigma^2 \hat{F}) M_N\) we have

\[
\frac{1}{N} \hat{B}_f' \epsilon' P_f = \frac{B_f' \epsilon' P_f}{N} \frac{1}{T} + \frac{P_f' \epsilon' \epsilon' P_f}{N} \frac{1}{T} - \frac{B_f' \epsilon' \epsilon' P_f}{N} \frac{1}{T} - \frac{P_f' \epsilon' \epsilon' P_f}{N} \frac{1}{T} - \frac{1}{N} \sum_{i=1}^{N} \epsilon_i' \frac{1_T}{T} - O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{\sqrt{N}} \right)
\]

\[
= \frac{B_f' \epsilon' P_f}{N} + P_f' \left( \frac{\epsilon_i' - \sigma^2 1_T}{T} \right) P_f + \sigma^2 (\hat{F}^* \hat{F})^{-1} - \frac{B_f' \epsilon' P_f}{N} \frac{1}{T} - O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{\sqrt{N}} \right)
\]

\[= \sigma^2 (\hat{F}^* \hat{F})^{-1} + O_p \left( \frac{1}{\sqrt{N}} \right)
\]

Lemma 3.11 Under Assumptions 3.1-3.4 and imposing (3.29), then

\[
\hat{\Delta}_{f \delta} = O_p \left( \frac{1}{N} \right)
\]

Proof. Starting from the definition in (3.113)-(3.113) and noticing that

\[
\eta_{f \delta} = \varepsilon - (\hat{X}_{f \delta} - X_{f \delta} \Gamma_{f \delta}) \gamma_{f \delta}^p = \varepsilon - (\hat{B}_{f} - B_f) \gamma_{f \delta}^p - \hat{B}_{g} \gamma_{f \delta}^p,
\]

then, we have

\[
\hat{\Delta}_{f \delta} = - (\hat{B}_f' \hat{B}_f')^{-1} \hat{B}_f' \eta_{f \delta}^p
\]

\[+ (\hat{B}_f' \hat{B}_g')^{-1} (\hat{B}_f' \hat{B}_g') \hat{\Delta}_{f \delta} (\hat{B}_f' \hat{B}_f')^{-1} \hat{B}_f' \eta_{f \delta}^p
\]

\[+ (\hat{B}_f' \hat{B}_g')^{-1} (\hat{B}_f' \hat{B}_g') \hat{\Delta}_{f \delta} (\hat{B}_f' \hat{B}_f')^{-1} \hat{B}_f' \eta_{f \delta}^p
\]

Then, collecting (3.44) with (3.47), and using the definition of \(\hat{\Delta}_{f \delta} = (\hat{B}_g' \hat{B}_g) - (\hat{B}_g' \hat{B}_g) (\hat{B}_f' \hat{B}_f)^{-1} (\hat{B}_f' \hat{B}_g)\)
\[ \hat{\Delta}_{fg} = (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1}(\hat{B}'_g \hat{B}_g - \hat{D}_{fg}) \gamma_{fg}' \]

\[ + (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1}(\hat{B}'_g \hat{B}_g)(\hat{B}'_f \hat{B}_f)^{-1}\hat{B}_f' \eta_{fg}' \]

\[ - (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1} \hat{B}_g' (\tau - (\hat{B}_f - B_f) \gamma_{fg}') \]

\[ = (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1}(\hat{B}'_g \hat{B}_g)(\hat{B}'_f \hat{B}_f)^{-1}\hat{B}_f' (\tau - (\hat{B}_f - B_f) \gamma_{fg}') \]

\[ + (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1}(\hat{B}'_g \hat{B}_g)(\hat{B}'_f \hat{B}_f)^{-1}\hat{B}_f' \epsilon' P_f \gamma_{fg}' \]

\[ - (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1} \hat{B}_g' (\tau - (\hat{B}_f - B_f) \gamma_{fg}') \]

\[ + (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1} \hat{B}_g' \epsilon' P_f \gamma_{fg}' \]

\[ = \alpha_p \left( \frac{1}{N \sqrt{N}} \right) \]  

\[ - (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1} \hat{B}_g' \tau \]  

\[ - (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1}(\hat{B}'_g \hat{B}_g)(\hat{B}'_f \hat{B}_f)^{-1} \hat{B}_f' \tau \epsilon' P_f \gamma_{fg}' \]  

\[ + (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1} \hat{B}_g' \epsilon' P_f \gamma_{fg}' \]  

Thus, collecting (3.B.54) and (3.B.56), then

\[ \hat{\Delta}_{fg} = -(\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1}(\hat{B}'_g \hat{B}_g)(\hat{B}'_f \hat{B}_f)^{-1} N \sigma^2 (\hat{F}' \hat{F}^{-1} \gamma_{fg}' + O_p \left( \frac{1}{N \sqrt{N}} \right) \]  

\[ - (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1} \hat{B}_g' \tau + P_f \gamma_{fg}' \]  

\[ + O_p \left( \frac{1}{N \sqrt{N}} \right) \]  

Therefore, collecting (3.B.54) and (3.B.56), then

\[ \hat{\Delta}_{fg} = -(\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1}(\hat{B}'_g \hat{B}_g)(\hat{B}'_f \hat{B}_f)^{-1} N \sigma^2 (\hat{F}' \hat{F}^{-1} \gamma_{fg}' + O_p \left( \frac{1}{N \sqrt{N}} \right) \]  

\[ - (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1} \hat{B}_g' \tau + P_f \gamma_{fg}' \]  

\[ + O_p \left( \frac{1}{N \sqrt{N}} \right) \]  

Now, notice that

\[ \hat{D}_{fg} = \hat{B}_g' \hat{B}_g - \hat{B}_g' \hat{B}_f (\hat{B}'_f \hat{B}_f)^{-1} \hat{B}_f' \hat{B}_g \]

\[ = O_p(N) + O_p(1) = O_p(N) \]

which implies that \( \hat{D}_{fg}^{-1} = O_p \left( \frac{1}{N} \right) \). Lemmas 3.7, 3.8, and Assumption 3.5 conclude the proof since

\[ (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_g \hat{B}_g) \hat{D}_{fg}^{-1}(\hat{B}'_g \hat{B}_g)(\hat{B}'_f \hat{B}_f)^{-1} N \sigma^2 (\hat{F}' \hat{F}^{-1} \gamma_{fg}' = O_p \left( \frac{1}{N} \right) , \]
and

\[(\hat{B}_f^e \hat{B}_f)^{-1}(\hat{B}_f^e \hat{B}_f)\hat{D}_f \hat{B}_f^e' \left( \frac{1_T}{T} - P_f \gamma_r^p \right) = O_p \left( \frac{1}{N} \right) \]  \[\blacksquare\]

**Lemma 3.12** Under Assumptions 3.1-3.5 and imposing (3.29),

\[\hat{d}_{tf}^r \to_p -\sigma^2 \left( \Sigma_{\beta_f} - \mu_{\beta_f} \mu_{\beta_f}' + \sigma^2 (F'M_T F)^{-1} \right)^{-1} (F'M_T F)^{-1} \gamma_r^p \]

\[\equiv -\sigma^2 D^{-1} (F'M_T F)^{-1} \gamma_r^p \]

\[\equiv d_1\]

**Proof.** Starting from the definition in (3.111) and defining \(\eta_f^p = (\bar{x} - (\bar{X}_f - X_f)\Gamma_f^p)\), we have

\[\hat{d}_{tf}^r = (\hat{B}_f^e \hat{B}_f)^{-1} \hat{B}_f^e' (\bar{x} - (\bar{X}_f - X_f)\Gamma_f^p) \]

\[\equiv (\hat{B}_f^e \hat{B}_f)^{-1} \hat{B}_f^e' \bar{x} - (\hat{B}_f^e \hat{B}_f)^{-1} \hat{B}_f^e' P_f \gamma_r^p \quad \text{(3.B.60)}\]

\[\equiv (\hat{B}_f^e \hat{B}_f)^{-1} \hat{B}_f^e' \bar{x} - (\hat{B}_f^e \hat{B}_f)^{-1} \hat{B}_f^e' P_f \gamma_r^p \quad \text{(3.B.61)}\]

Consider the first term on the right-hand side of (3.B.61). We have

\[\langle \hat{B}_f^e \hat{B}_f \rangle^{-1} \hat{B}_f^e \bar{x} = (\hat{B}_f^e \hat{B}_f)^{-1} \hat{B}_f^e \bar{x} + (\hat{B}_f^e \hat{B}_f)^{-1} (\hat{B}_f - B_f)' \bar{x},\]

where

\[\frac{\hat{B}_f^e \bar{x}}{N} = \frac{1}{N} B_f \left( I_N - \frac{1}{N} Y_N' N \right) \bar{x}\]

\[= \frac{1}{N} B_f \bar{x}' \frac{1_T}{T} \frac{1}{N} - \frac{1}{N} B_f \bar{x}' \frac{1_T}{T} \frac{1_Y}{N} \bar{x}' \frac{1_T}{T}\]

\[= \frac{1}{N} \sum_{i=1}^{N} \beta_i \xi_i' \frac{1_T}{T} - \frac{1}{N} \sum_{i=1}^{N} \beta_i \xi_i' \frac{1_T}{T} + o_p \left( \frac{1}{\sqrt{N}} \right)\]

\[= O_p \left( \frac{1}{\sqrt{N}} \right) + o_p \left( \frac{1}{N \sqrt{N}} \right) + o_p \left( \frac{1}{\sqrt{N}} \right)\]

\[= O_p \left( \frac{1}{\sqrt{N}} \right)\]

and where
\[
\frac{\hat{\beta}_f - \beta_f}{N} = \frac{1}{N} P_f' \epsilon' \left( I_N - \frac{1}{N} \frac{1}{N} \right) \epsilon' \frac{1}{T} = \frac{1}{N} P_f' \epsilon' \frac{1}{T} - \frac{1}{N} \frac{1}{N} \epsilon' \frac{1}{T} \\
= P_f' \frac{1}{N} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T \frac{1}{T}) - \frac{1}{N^2} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T \frac{1}{T}) + o_p \left( \frac{1}{\sqrt{N}} \right) \\
= O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{N \sqrt{N}} \right) + o_p \left( \frac{1}{\sqrt{N}} \right) \\
= O_p \left( \frac{1}{\sqrt{N}} \right)
\]

by Assumptions 3.4 and 3.5. Therefore, using the fact that \( \frac{\hat{\beta}_f \hat{\beta}_f}{N} = O_p(1) \), we have

\[
(\hat{\beta}_f \hat{\beta}_f)^{-1} \hat{\beta}_f' \epsilon = O_p \left( \frac{1}{\sqrt{N}} \right).
\]

Consider now the second term on the right-hand-side of (3.B.61) and rewrite it as:

\[
(\hat{\beta}_f \hat{\beta}_f)^{-1} \hat{\beta}_f' \epsilon' P_f \gamma^P = (\hat{\beta}_f \hat{\beta}_f)^{-1} \hat{\beta}_f' \epsilon' P_f \gamma^P + (\hat{\beta}_f \hat{\beta}_f)^{-1} (\hat{\beta}_f - \beta_f)' \epsilon' P_f \gamma^P,
\]

where, using the same steps as before

\[
(\hat{\beta}_f \hat{\beta}_f)^{-1} \hat{\beta}_f' \epsilon' P_f \gamma^P = O_p \left( \frac{1}{\sqrt{N}} \right)
\]

while

\[
\frac{1}{N} (\hat{\beta}_f - \beta_f)' \epsilon' P_f \gamma^P = \frac{1}{N} P_f' \epsilon' \left( I_N - \frac{1}{N} \frac{1}{N} \right) \epsilon' P_f \gamma^P \\
= P_f' \frac{1}{N} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T) P_f \gamma^P + \sigma^2 (\hat{\beta}_f \hat{\beta}_f)^{-1} \gamma^P \\
- P_f' \frac{1}{N^2} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T) P_f \gamma^P + o_p \left( \frac{1}{N} \right) \\
= \sigma^2 (\hat{\beta}_f \hat{\beta}_f)^{-1} \gamma^P + O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{N} \right).
\]

Using the result in (3.C.125) concludes the proof. \(\blacksquare\)
3.C. PROOFS OF THEOREMS

3.C. Proofs of Theorems

Proof of Proposition 3.1. Starting from the definition in (3.11) we have

\[
\frac{\hat{X}_g' \hat{X}_g}{N} - \bar{\Lambda} = \begin{bmatrix}
1 & \frac{1_g^b \bar{\theta}_g}{N} \\
\bar{\beta}_g' \bar{\beta}_g \ & \bar{\beta}_g' \bar{\beta}_g
\end{bmatrix} - \begin{bmatrix}
0 & 0_g' \\
0_{K_g} & \bar{\delta}^2 (\bar{G}' \bar{G})^{-1}
\end{bmatrix} = \frac{1}{N} \begin{bmatrix}
1 & \frac{1_g' \bar{P}_g}{N} \\
\bar{P}_g' \bar{P}_g \ & \bar{P}_g' \bar{P}_g
\end{bmatrix}.
\]

Hence, defining \( D = \frac{P_g' (\bar{G}' - \frac{1}{N} \Sigma \bar{G}) P_g}{N} - \frac{P_g' \bar{P}_g}{N} \), and using the rule for the inverse of a partitioned matrix, we can write

\[
\left( \frac{\hat{X}_g' \hat{X}_g}{N} - \bar{\Lambda} \right)^{-1} = \begin{bmatrix}
1 + \frac{1_g' \bar{P}_g}{N} D^{-1} & \frac{1_g' \bar{P}_g}{N} D^{-1} \\
- D^{-1} \frac{P_g' \bar{P}_g}{N} & D^{-1}
\end{bmatrix}.
\]

Notice that, by Assumptions 3.5(i) and 3.5(ii), \( D = O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{N} \right) \), and hence \( D^{-1} = O_p \left( \sqrt{N} \right) \). It implies that

\[
\left( \frac{\hat{X}_g' \hat{X}_g}{N} - \bar{\Lambda} \right)^{-1} = \begin{bmatrix}
1 + \frac{1_g' \bar{P}_g}{N} D^{-1} \frac{P_g' \bar{P}_g}{N} D^{-1} & \frac{1_g' \bar{P}_g}{N} D^{-1} \\
- D^{-1} \frac{P_g' \bar{P}_g}{N} & D^{-1}
\end{bmatrix} = \begin{bmatrix}
1 + O_p \left( \frac{1}{\sqrt{N}} \right) & O_p \left( 1 \right) \\
O_p \left( 1 \right) & O_p \left( \sqrt{N} \right)
\end{bmatrix}.
\]

Consider now the term \( \frac{\hat{X}_g' \hat{R}}{N} \). We have

\[
\frac{\hat{X}_g' \hat{R}}{N} = \frac{\hat{X}_g' \bar{X} \Gamma P + \bar{r}}{N} = \begin{bmatrix}
1 & \bar{0}_K \\
\bar{\beta}_g' \bar{\beta}_g \ & \bar{0}_{K \times K}
\end{bmatrix} \Gamma P + \begin{bmatrix}
\frac{1_{g'}' \bar{P}_g}{N} \\
\bar{P}_g' \bar{P}_g \ & \bar{P}_g' \bar{P}_g
\end{bmatrix} = \gamma_0 \begin{bmatrix}
1 \\
\bar{P}_g' \bar{P}_g \ & \bar{P}_g' \bar{P}_g
\end{bmatrix} = \gamma_0 \begin{bmatrix}
1 + O_p \left( \frac{1}{\sqrt{N}} \right) \\
O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{N} \right)
\end{bmatrix}.
\]

Then,
\[ \hat{\Gamma}^* = \left( \frac{\hat{X}'_g \hat{X}_g}{N} - \hat{\Lambda} \right)^{-1} \frac{\hat{X}'_g \hat{R}}{N} \]
\[ = \gamma^0 \begin{bmatrix} \left[ 1 + \frac{1}{N} \frac{\epsilon \epsilon'}{N} P_g \mathcal{D}^{-1} P_g \mathcal{D}^{-1} \right] \begin{bmatrix} 1 \\ P_g \mathcal{D}^{-1} \end{bmatrix} + \left[ 1 + \frac{1}{N} \frac{\epsilon \epsilon'}{N} \mathcal{D}^{-1} P_g \mathcal{D}^{-1} \right] \begin{bmatrix} \frac{\epsilon \epsilon'}{N} \frac{1}{\sqrt{N}} \\ P_g \mathcal{D}^{-1} \frac{\epsilon \epsilon'}{N} \frac{1}{\sqrt{N}} \end{bmatrix} \\ \mathcal{D}^{-1} \frac{\epsilon \epsilon'}{N} \frac{1}{\sqrt{N}} \end{bmatrix} \end{bmatrix} \]
\[ = \gamma^0 \begin{bmatrix} \left[ 1 + \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right) \right] \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right) + \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right) \\ \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right) + \mathcal{D}^{-1} P_g \left( \frac{\epsilon \epsilon'}{N} - \sigma^2 I_T \right) \frac{1}{\sqrt{N}} \end{bmatrix} \]
\[ = \gamma^0 \begin{bmatrix} \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right) \\ \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right) \end{bmatrix} \]
\[ \mathcal{D}^{-1} P_g \left( \frac{\epsilon \epsilon'}{N} - \sigma^2 I_T \right) \frac{1}{\sqrt{N}} \]
\[ = \gamma^0 \begin{bmatrix} \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right) \\ \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right) \end{bmatrix} \]
\[ \mathcal{D}^{-1} P_g \left( \frac{\epsilon \epsilon'}{N} - \sigma^2 I_T \right) \frac{1}{\sqrt{N}} \]

where \( Z \) is a matrix of Normal random variables such that

\[ \sqrt{N} D = \sqrt{N} \frac{P_g \left( \epsilon \epsilon' - \sigma^2 I_T \right) P_g}{N} + \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right) \]

and where \( H \) is a normally distributed \( K_g \)-vector such that

\[ \left( \frac{\epsilon \epsilon'}{T} \right) \sqrt{N} \mathcal{O}_p \left( \frac{\epsilon \epsilon'}{N} - \sigma^2 I_T \right) \to_d \left( 0_{K_g}, \left( \frac{\epsilon \epsilon'}{T} \otimes P_g \right) U_\epsilon \left( \frac{\epsilon \epsilon'}{T} \otimes P_g \right) \right) \equiv H \]

**Proof of Theorem 3.1.** Starting from (3.10) and using (3.9), we have

\[ \hat{\Gamma}_g = \left( \frac{\hat{X}'_g \hat{X}_g}{N} \right)^{-1} \frac{\hat{X}'_g \hat{R}}{N} \]
\[ = \left( \frac{\hat{X}'_g \hat{X}_g}{N} \right)^{-1} \frac{\hat{X}'_g \left( \hat{\Gamma} + \hat{\epsilon} - (\hat{X}_g - X)\hat{\Gamma} \right)}{N} \]
\[ = \hat{\Gamma} + \left( \frac{\hat{X}'_g \hat{X}_g}{N} \right)^{-1} \left( \frac{\hat{X}_g \hat{\epsilon}}{N} - \frac{\hat{X}_g (\hat{X}_g - X)\hat{\Gamma}}{N} \right). \]

By Lemma 3.1 and using the definition in (3.8), then
\[
\frac{\hat{X}_2'\hat{X}_2}{N} = \left[ \frac{1}{P'_g N} \begin{bmatrix} 1 & \frac{1}{\sqrt{N}} \bar{\epsilon} \bar{\epsilon}' & 0 \end{bmatrix} P_g \right] \rightarrow P_0 \left[ \begin{bmatrix} 0_{K_g} \\ \sigma^2 (G'G_1) \end{bmatrix} \right]^{-1}.
\]

(3.C.4)

Moreover, rewriting \( \epsilon = \epsilon' \frac{1}{\sqrt{N}} \) and assuming that \( B = 0_{N \times K} \), then

\[
\frac{\hat{X}'_2 \bar{\epsilon}}{N} = \frac{(\hat{X}_2 - X)' \bar{\epsilon}}{N} + \frac{X' \bar{\epsilon}}{N}
\]

\[
= \frac{1}{N} \begin{bmatrix} 0 \\ P'_g \frac{\epsilon \epsilon'}{\sqrt{N}} \end{bmatrix} + \frac{1}{N} \begin{bmatrix} 1_N' \\ B' \end{bmatrix} \epsilon \frac{1_N}{\sqrt{N}}
\]

\[
= \begin{bmatrix} 0 \\ \frac{P'_g \epsilon \epsilon'}{\sqrt{N}} \end{bmatrix} + \begin{bmatrix} \frac{1_N'}{N} \\ 0_K \end{bmatrix}
\]

\[
= \begin{bmatrix} \frac{1_N'}{N} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma_i^2 I_{K}) \frac{1}{\sqrt{N}} \end{bmatrix} + o_p \left( \frac{1}{\sqrt{N}} \right)
\]

\[
= O_p \left( \frac{1}{\sqrt{N}} \right) + o_p \left( \frac{1}{\sqrt{N}} \right)
\]

(3.C.5)

by Lemmas 3.3 and 3.5 and using Assumptions 3.4(i) 3.5(ii). Finally, consider \( \frac{\hat{X}'_2}{N} (\hat{X}_2 - X) \Gamma^P \) and rewrite it as

\[
\frac{\hat{X}'_2 (\hat{X}_2 - X) \Gamma^P}{N} = \frac{(\hat{X}_2 - X)'(\hat{X}_2 - X) \Gamma^P}{N} + \frac{X'(\hat{X}_2 - X) \Gamma^P}{N}
\]

\[
= \begin{bmatrix} 0 \\ 0_{K_g} \end{bmatrix} \frac{1}{P'_g \frac{\epsilon \epsilon'}{\sqrt{N}}} \Gamma^P + \begin{bmatrix} 0 \\ 0_{K_g} \end{bmatrix} \frac{1_N'}{N} \frac{1}{P'_g \frac{\epsilon \epsilon'}{\sqrt{N}}} \Gamma^P
\]

\[
= \begin{bmatrix} 0 \\ 0_{K_g} \end{bmatrix} \frac{1}{\sigma^2 (G'G_1)} \Gamma^P + o_p \left( \frac{1}{\sqrt{N}} \right) + o_p \left( \frac{1}{\sqrt{N}} \right)
\]

(3.C.6)

by Assumption 3.4(i). Then, using the results in (3.C.4) together with (3.C.6), we have

\[
\frac{\hat{X}'_2 (\hat{X}_2 - X) \Gamma^P}{N} = \frac{1}{0_{K_g} \sigma^2 (G'G_1)} \begin{bmatrix} 0 \\ 0_{K_g} \end{bmatrix} \left( \begin{bmatrix} 0 \\ 0_{K_g} \end{bmatrix} \sigma^2 (G'G_1)^{-1} \right) \Gamma^P + o_p \left( \frac{1}{\sqrt{N}} \right) + o_p \left( \frac{1}{\sqrt{N}} \right)
\]

(3.C.7)

and, by the result in (3.C.5),

\[
\left( \frac{\hat{X}'_2 \hat{X}}{N} \right)^{-1} \frac{\hat{X}_2' \bar{\epsilon}}{N} = O_p \left( \frac{1}{\sqrt{N}} \right)
\]

(3.C.8)

Hence, using (3.C.7) and (3.C.8), Equation (3.C.3) can be written as
\[
\hat{\gamma}_g = \Gamma^P - \begin{bmatrix} 0 & 0'_{K_s} \\ 0_{K_s} & I_{K_s} \end{bmatrix} \Gamma^P + O_p\left(\frac{1}{\sqrt{N}}\right) + o_p\left(\frac{1}{\sqrt{N}}\right)
\]
\[
= \begin{bmatrix} \gamma_0 \\ 0_{K_s} \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N}}\right).
\] (3.C.9)

which concludes the proof of part (i).

(ii) Starting from (3.C.3), and using the results in (3.C.5) and (3.C.6) in part (i), we have

\[
\hat{\gamma}_g = \Gamma^P + \left(\frac{\hat{X}_g'\hat{X}_g}{N}\right)^{-1} \left(\left[ \begin{array}{c} \frac{1}{2} \frac{Q}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T) \frac{1}{\sqrt{T}} \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] \Gamma^P - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s \times K_s} \end{array} \right] \Gamma^P \right) + \frac{1}{\sqrt{N}} \left[ \begin{array}{c} 0 \\ \gamma_1^P \end{array} \right]
\]

Hence, defining the T-vector \( Q = \frac{1}{\sqrt{T}} - P_g \gamma_1^P \), then

\[
\sqrt{N} \left( \hat{\gamma}_g - \begin{bmatrix} \gamma_0 \\ 0_{K_s} \end{bmatrix} \right) = 
\]
\[
= \sqrt{N} \left( \frac{\hat{X}_g'\hat{X}_g}{N} \right)^{-1} \left( \left[ \begin{array}{c} \frac{1}{2} \frac{Q}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T) \frac{1}{\sqrt{T}} \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] \Gamma^P - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s \times K_s} \end{array} \right] \Gamma^P \right) + \sqrt{N} \left[ \begin{array}{c} 0 \\ \gamma_1^P \end{array} \right]
\]
\[
= \sqrt{N} \left( \frac{\hat{X}_g'\hat{X}_g}{N} \right)^{-1} \left( \left[ \begin{array}{c} \frac{1}{2} \frac{Q}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T) \frac{1}{\sqrt{T}} \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] \Gamma^P - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s \times K_s} \end{array} \right] \Gamma^P \right)
\]
\[
+ \sqrt{N} \left( \frac{\hat{X}_g'\hat{X}_g}{N} \right)^{-1} \left( \left[ \begin{array}{c} \frac{1}{2} \frac{Q}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T) \frac{1}{\sqrt{T}} \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] \Gamma^P - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s \times K_s} \end{array} \right] \Gamma^P \right)
\]
\[
= \left( \frac{\hat{X}_g'\hat{X}_g}{N} \right)^{-1} \left( \left[ \begin{array}{c} \frac{1}{2} \frac{Q}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T) \frac{1}{\sqrt{T}} \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] \Gamma^P - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s \times K_s} \end{array} \right] \Gamma^P \right)
\]
\[
\left( \frac{\hat{X}_g'\hat{X}_g}{N} \right)^{-1} \left[ \begin{array}{c} \frac{1}{2} \frac{Q}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T) \frac{1}{\sqrt{T}} \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] \Gamma^P - \left[ \begin{array}{c} 0 \\ 0'_{K_s} \\ 0_{K_s \times K_s} \end{array} \right] \Gamma^P \right)
\] (3.C.10)

We need now to derive the variance of (3.C.10). Using the result in (3.C.4), we have

\[
\left( \frac{\hat{X}_g'\hat{X}_g}{N} \right)^{-1} \rightarrow_p \begin{bmatrix} 1 & 0'_{K_s} \\ 0_{K_s} & \frac{\sigma^2}{\sigma^2} \end{bmatrix}.
\] (3.C.11)

Moreover, under Assumption 3.4(vi),

\[
\text{Var} \left( \left[ \begin{array}{c} \frac{1}{2} \frac{Q}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T) \frac{1}{\sqrt{T}} \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] \right) = \frac{1}{N} \left[ \begin{array}{c} \text{Var} \left( \frac{1}{2} \frac{Q}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2 I_T) \frac{1}{\sqrt{T}} \right) \\ 0'_{K_s} \\ 0_{K_s} \end{array} \right] \text{Var} \left( \left[ \begin{array}{c} 0'_{K_s} \\ 0_{K_s} \end{array} \right] \right).
\] (3.C.12)

Notice that
by Assumption 3.4(i) and 3.4(ii). Using also Assumption 3.5(ii)

\[
\frac{1}{N} \text{Var} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i^T Y_i}{T} \right) = \frac{1}{N} \text{Var} \left( \sum_{i=1}^{N} \frac{Y_i^T Y_i}{T} \right)
\]

\[
= \frac{1}{N} \text{Var} \left( \sum_{i=j}^{N} \frac{Y_i^T Y_i}{T} + \sum_{i \neq j}^{N} \frac{Y_i^T Y_i}{T} \right)
\]

\[
= \frac{1}{N} \left( \frac{Y_j^T Y_j}{T} \sum_{i=j}^{N} \sigma_i^2 + \frac{1}{T} \sum_{i \neq j}^{N} \sigma_{ij} \right)
\]

\[
\to \frac{\sigma^2}{T}
\]  \hspace{1cm} (3.13)

Hence, substituting (3.13) and (3.14) in (3.12), we get that

\[
\text{Var} \left( \sqrt{N} \left( \hat{\Gamma}_G - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right) \right) = \left( \frac{\hat{X}_G^T \hat{X}_G}{N} \right)^{-1} \text{Var} \left( \begin{bmatrix} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (Y_i^T Y_i) \frac{Y_i^T Y_i}{T} \end{bmatrix} \right) \left( \frac{\hat{X}_G^T \hat{X}_G}{N} \right)^{-1}
\]

\[
\to_p \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \frac{\sigma^2}{T} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (Y_i^T Y_i) \right) \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (Y_i^T Y_i) \right) \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (Y_i^T Y_i) \right) \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (Y_i^T Y_i) \right)
\]

Defining \( C = \left( \frac{1}{\sqrt{T}} \otimes \hat{G} \right) \) concludes the proof of part (ii). \( \blacksquare \)

**Proof of Theorem 3.2.** By Lemma 3.1, \( \hat{\sigma}^2 \) is a consistent estimator of \( \sigma^2 \). A consistent estimator of \( V \) requires a consistent estimate of the matrix \( U_e \), which can be obtained using Lemma 3.6. This concludes the proof of Theorem 3.2. \( \blacksquare \)

**Proof of Theorem 3.3.** (i) To derive the asymptotic distribution of \( t_k \), it is convenient to rewrite it as

\[
t_{g,k} = \sqrt{\frac{\hat{N}}{N \cdot \frac{\sigma^2}{c_k}}} = \sqrt{\frac{\hat{N} \gamma_{1k}}{\frac{\sigma^2}{c_k}}} \frac{1}{\sqrt{\frac{\hat{X}_G^T \hat{X}_G}{N} \left( \begin{bmatrix} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (Y_i^T Y_i) \end{bmatrix} \right)^{-1}}}
\]  \hspace{1cm} (3.15)
Let us start by analyzing $\hat{s}^2$. Using the definition in (3.16) and noticing that $\hat{e}_g = \hat{R} - \hat{X}_g \hat{\Gamma}_g$, we have that

$$\begin{align*}
\frac{\hat{e}_g^T \hat{e}_g}{N} &= \frac{\langle \hat{R} - \hat{\Gamma}_g \hat{X}_g \rangle (\hat{R} - \hat{X}_g \hat{\Gamma}_g)}{N} \\
&= \frac{\hat{R}^T \hat{R}}{N} - 2 \frac{\hat{R}^T \hat{X}_g \hat{\Gamma}_g}{N} + \frac{\hat{\Gamma}_g^T \hat{X}_g \hat{\Gamma}_g}{N} \\
&= \frac{\hat{R}^T \hat{R}}{N} - 2 \frac{\hat{R}^T \hat{X}_g \hat{X}_g \hat{\Gamma}_g}{N} + \frac{\hat{\Gamma}_g^T \hat{X}_g \hat{X}_g \hat{\Gamma}_g}{N} + 2 \frac{\hat{e}_g^T \hat{X}_g \hat{\Gamma}_g}{N} + \frac{\hat{\Gamma}_g^T \hat{X}_g \hat{\Gamma}_g}{N} \\
&= \frac{\hat{R}^T \hat{R}}{N} - \frac{\hat{\Gamma}_g^T \hat{X}_g \hat{X}_g \hat{\Gamma}_g}{N} + \frac{\hat{\Gamma}_g^T \hat{X}_g \hat{\Gamma}_g}{N},
\end{align*}$$

(3.16)

since $\hat{X}_g \hat{\Gamma}_g = 0_{K_g}$. Consider the first term on the right-hand side of (3.16). Using the fact that $\hat{R} = X \Gamma' + \tilde{\xi}$, we have

$$\begin{align*}
\frac{\hat{R}^T \hat{R}}{N} &= \frac{(\Gamma' X' + \tilde{\xi})(X \Gamma' + \tilde{\xi})}{N} \\
&= \frac{\tilde{\xi} \tilde{\xi}}{N} + \frac{\Gamma' X' X \Gamma'}{N} + 2 \frac{\Gamma' X' \tilde{\xi}}{N} \\
&= \frac{\tilde{\xi} \tilde{\xi}}{N} + \frac{\Gamma' X' X \Gamma'}{N} + O_p \left( \frac{1}{\sqrt{N}} \right) \\
\rightarrow_p & \quad \frac{\sigma^2}{T} + \left[ \gamma_0, \gamma_1' \right] \begin{bmatrix} 1 & 0_{0 \times K_g} \\ 0_{K_g} & 0_{K_g \times K_g} \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1' \end{bmatrix} \\
&= \frac{\sigma^2}{T} + \gamma_0^2.
\end{align*}$$

(3.17)

Consider now the second term on the right-hand side of (3.16). Lemma 3.2 and Theorem 3.1 implies that

$$\begin{align*}
\frac{\hat{\Gamma}_g^T \hat{X}_g \hat{X}_g \hat{\Gamma}_g}{N} &\rightarrow_p \left[ \gamma_0, 0_{0 \times K_g} \right] \begin{bmatrix} 1 & 0_{0 \times K_g} \\ 0_{K_g} & \sigma^2 (\hat{G}' \hat{G})^{-1} \end{bmatrix} \begin{bmatrix} \gamma_0 \\ - \gamma_0^2 \end{bmatrix} \\
&= \gamma_0^2.
\end{align*}$$

(3.18)

Hence, using (3.17) and (3.18), together with (3.16), we have

$$\begin{align*}
\frac{\hat{e}_g^T \hat{e}_g}{N} &\rightarrow_p \frac{\sigma^2}{T} + \gamma_0^2 - \gamma_0^2 \\
&= \frac{\sigma^2}{T}.
\end{align*}$$

(3.19)

Then, Lemma 3.2 and the result in (3.19) imply that

$$s^2 \left( \frac{\hat{X}_g \hat{X}_g}{N} \right)^{-1} \rightarrow_p \begin{bmatrix} \frac{\sigma^2}{T} & 0_{0 \times K_g} \\ 0_{K_g} & \frac{-\gamma_0}{T} \end{bmatrix}.$$  

(3.20)

Consider now the numerator of (3.15). Remember that, from Theorem 3.1
\[ \sqrt{N}(\hat{\gamma}_{p,1}) \rightarrow_d \left( 0_K, \frac{1}{\sigma^2} C' U C \right) \]

where \( C = \left( \frac{1}{T^2} \otimes \tilde{G} \right) = \frac{1}{T} [\tilde{G}', \tilde{G}', ..., \tilde{G}']' \) is a \( T^2 \times K \) matrix. Using the form of \( U \) as in Raponi et al. (2019) (see Appendix C), we have

\[ C' U C = \frac{1}{T^2} \tilde{G}' \left( \sum_{t,s=1}^{T} U_{t,s} \right) \tilde{G} \]
\[ = \frac{1}{T^2} \tilde{G}' \left( \sum_{t=1}^{T} U_{t,t} + \sum_{t \neq s}^{T} U_{t,s} \right) \tilde{G} \]
\[ = \frac{1}{T^2} \tilde{G}' \left( (k_4 + (T + 1)\sigma_4) I_T + \sigma_4 (I_T 1_T' - I_T) \right) \tilde{G} \]
\[ = \frac{1}{T^2} \tilde{G}' \left( (k_4 + T\sigma_4) I_T + \sigma_4 1_T 1_T' \right) \tilde{G} \]
\[ = \frac{1}{T^2} \tilde{G}' \left( \frac{k_4 + T\sigma_4}{T^2} G \right) \tilde{G} \]
\[ = \frac{(k_4 + T\sigma_4)}{T^2} \tilde{G}' \tilde{G} \]  

(3.21)

since, by construction, \( \tilde{G}' 1_T = 0_K \). It follows that

\[ \sqrt{N}(\hat{\gamma}_{1k}) \rightarrow_d \left( 0_K, \left( \frac{k_4 + T\sigma_4}{\sigma^4 T^2} \right) \tilde{G}' \tilde{G} \right) , \]  

(3.22)

and hence, combining (3.21) with (3.20),

\[ t_k = \frac{\hat{\gamma}_{1k}}{s \sqrt{\text{Var}}} \rightarrow_d \left( 0, \left( \frac{k_4 + T\sigma_4}{\sigma^4} \right) \right) \]

which concludes the proof of part (i).

(ii) By definition

\[ R_{GRS}^2 = 1 - \frac{\hat{c}'_g \hat{c}_g / N}{R'MR / N} . \]  

(3.23)

Moreover, notice that

\[ \frac{R' 1_N}{N} = \frac{(\hat{c}'_g + \hat{\gamma}_g X'_g) 1_N}{N} \]
\[ = \frac{\hat{c}'_g 1_N}{N} + \frac{\hat{\gamma}_g X'_g 1_N}{N} \]
\[ \rightarrow_p \gamma_0 \]  

(3.24)

since \( \hat{c}'_g 1_N = 0 \). Then, using (3.24) together with (3.17), we get
\[
\frac{\bar{R}'\bar{M}\bar{R}}{N} = \frac{\bar{R}'\bar{R}}{N} - \frac{\bar{R}'1_N 1_N'\bar{R}}{N^2} \\
\rightarrow_p \frac{\sigma^2}{T} + \gamma_0^2 - \gamma_0^2 \\
= \frac{\sigma^2}{T}
\] (3.25)

It implies that

\[
R_{CRS}^2 = 1 - \frac{\hat{e}_g'\hat{e}_g/N}{\bar{R}'\bar{M}\bar{R}/N} \\
\rightarrow_p 1 - \frac{\sigma^2/T}{\sigma^2/T} \\
= 0.
\] (3.26)

This concludes the proof of part (ii).

(iii) Consider (3.19) and rewrite it as

\[
F_{CSR} = \frac{R_{CSR}/K_g}{(1 - R_{CSR}^2)/N - (K_g + 1)} \\
= \frac{NR_{CSR}^2}{K_g(1 - R_{CSR}^2)} \left( \frac{N - K_g - 1}{N} \right).
\] (3.27)

Using (3.23), (3.16) and (3.25), we have

\[
R_{CSR_g}^2 = 1 - \frac{\hat{e}_g'\hat{e}_g}{\bar{R}'\bar{M}\bar{R}} \\
= 1 - \frac{\bar{R}'R - \hat{\Gamma}_g'\hat{\Xi}_g'\hat{\Xi}_g\hat{\Gamma}_g}{\bar{R}'\bar{M}\bar{R}/N} \\
= \frac{\bar{R}'\bar{R}/N - \bar{R}'1_N 1_N'\bar{R}/N^2 - \bar{R}'\bar{R}/N + \hat{\Gamma}_g'\hat{\Xi}_g'\hat{\Xi}_g\hat{\Gamma}_g/N}{\bar{R}'\bar{M}\bar{R}/N} \\
= \frac{\hat{\Gamma}_g'\hat{\Xi}_g'\hat{\Xi}_g\hat{\Gamma}_g/N - \hat{\Gamma}_g'\hat{\Xi}_g' \bar{1}_N' \bar{1}_N \hat{\Xi}_g \hat{\Gamma}_g/N^2}{\bar{R}'\bar{M}\bar{R}/N} \\
= \frac{\hat{\Gamma}_g'\hat{\Xi}_g'\hat{\Xi}_g\hat{\Gamma}_g/N - \hat{\Gamma}_g'\hat{\Xi}_g' \bar{1}_N' \bar{1}_N \hat{\Xi}_g \hat{\Gamma}_g/N^2}{\bar{R}'\bar{M}\bar{R}/N} \\
= \frac{\hat{\Gamma}_g \left( \hat{\Xi}_g'\hat{\Xi}_g/N - \hat{\Xi}_g' \bar{1}_N' \hat{\Xi}_g \hat{\Gamma}_g/N \right)}{\bar{R}'\bar{M}\bar{R}/N}.
\] (3.28)

Moreover, notice that, using Lemma 3.3,
\[
\frac{\tilde{X}_g 1_N}{N} = \left[ \frac{1}{N} \right] 1_N \\
= \left[ \frac{1}{P'_{g} \epsilon P_g} \right] = \left[ \frac{1}{O_p \left( \frac{1}{\sqrt{N}} \right)} \right]
\]  
(3.29)

which implies that

\[
\frac{\tilde{X}_g 1_N}{N} \frac{1_N}{N} \tilde{X}_g = \left[ \frac{1}{P'_{g} \epsilon P_g} \frac{1}{\sqrt{N}} \epsilon \right] \frac{1}{\sqrt{N}} \epsilon \frac{\epsilon}{N} = \left[ \frac{1}{P'_{g} \epsilon P_g} \frac{1}{\sqrt{N}} \epsilon \right] \frac{1}{\sqrt{N}} \epsilon \frac{\epsilon}{N}.
\]  
(3.30)

Therefore, using (3.24) and (3.30) in (3.28), we have

\[
NR^2_{CSR} = \frac{\sqrt{N} \gamma}{\sqrt{\gamma} \gamma} \left( \frac{\tilde{X}_g 1_N}{N} \epsilon \frac{\epsilon}{N} \right) \frac{1}{\sqrt{N}} \epsilon \frac{\epsilon}{N} = \frac{\sqrt{N} \gamma}{\sqrt{\gamma} \gamma} \frac{1}{\sqrt{N}} \epsilon \frac{\epsilon}{N} + o_p(1)
\]  
(3.31)

Now, by (3.22) and using the fact that \( \frac{P'_{g} \epsilon P_g}{N} \rightarrow_p \sigma^2 (\tilde{G'} \tilde{G})^{-1} \), then

\[
\sqrt{N} \gamma \left( \frac{P'_{g} \epsilon P_g}{N} \right) \frac{1}{\sqrt{N}} \epsilon \frac{\epsilon}{N} \rightarrow \mathcal{Z} \left( \frac{k_4 + T \sigma_4}{\sigma^4 T^2} \right) \tilde{G}' \tilde{G} \sigma^2 (\tilde{G}' \tilde{G})^{-1} \mathcal{Z} = \frac{k_4 + T \sigma_4}{\sigma^4 T^2} \mathcal{Z}' \mathcal{Z}
\]  
(3.32)

where \( \mathcal{Z} \) denotes a \( \mathcal{N}(0, I_K) \) random variable. Finally, using (3.32) and the result in (3.25) in (3.31) we have

\[
NR^2_{CSR} \rightarrow \frac{\frac{k_4 + T \sigma_4}{\sigma^4 T^2} \mathcal{Z}' \mathcal{Z}}{\sigma^2 / T} = \frac{k_4 / T + \sigma_4 \mathcal{Z}' \mathcal{Z}}{\sigma^4}.
\]  
(3.33)

Then, substituting the results in (3.33) and (3.26) in (3.27) implies that

\[
F_{CSR} \rightarrow \chi^2_K \left( \frac{k_4 / T + \sigma_4 \mathcal{Z}' \mathcal{Z}}{\sigma^4} \right) / K
\]
which concludes the proof of part (iii). ■

Proof of Theorem 3.4. (i) For the proof of this part we follow the main steps of Theorem 3.1. Using (3.22), we have that

\[
\hat{\Gamma}_g = \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \frac{\hat{X}_g^t \hat{R}}{N} 
\]

\[
= \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \frac{\hat{X}_g^t}{N} \left( \frac{0}{\gamma_1^p} \right) + \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \left( \frac{\hat{X}_g^t}{N} \hat{X}_g - \frac{\hat{X}_g^t}{N} \hat{X}_g \right)^p \left( \frac{\hat{X}_g^t}{N} \hat{R} - \frac{\hat{X}_g^t}{N} \hat{X}_g \right) \right). \tag{3.C.34}
\]

since \( \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \frac{\hat{X}_g^t}{N} \hat{B}_g \gamma_1^p = \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \frac{\hat{X}_g^t}{N} \gamma_1^p = \left[ \gamma_1^p \right] \). Moreover, by Assumption 3.6,

\[
\frac{\hat{X}_g^t}{N} c = \left[ \frac{\mu_c}{0_{K_x}} \right] + O_p \left( \frac{1}{\sqrt{N}} \right),
\]

implying that

\[
\left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \frac{\hat{X}_g^t}{N} c = \left[ \frac{\mu_c}{0_{K_x}} \right].
\]

Using the same results of Theorem 3.1, we also have that

\[
\left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \left( \frac{\hat{X}_g^t}{N} \hat{X}_g - \frac{\hat{X}_g^t}{N} \hat{X}_g \right)^p = - \left[ \begin{array}{c} 0 \\ 0_{K_x} \\ I_{K_x} \end{array} \right] \Gamma^p + O_p \left( \frac{1}{\sqrt{N}} \right) 
\]

\[
= \left[ \begin{array}{c} 0 \\ \gamma_1^p \end{array} \right] + O_p \left( \frac{1}{\sqrt{N}} \right).
\]

Therefore (3.C.34) reduces to

\[
\hat{\Gamma}_g = \left[ \begin{array}{c} \mu_c^p \\ \gamma_1^p \end{array} \right] - \left[ \begin{array}{c} 0 \\ \gamma_1^p \end{array} \right] + O_p \left( \frac{1}{\sqrt{N}} \right) 
\]

\[
= \left[ \begin{array}{c} \mu_c \\ 0_{K_x} \end{array} \right] + O_p \left( \frac{1}{\sqrt{N}} \right)
\]

which concludes the proof of part (i).

(ii) Starting from (3.C.34), we have
\[ \hat{\Gamma}_g = \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \frac{\hat{X}_g}{N} \left[ 0 \right] + \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \left( \frac{\hat{X}_g^t (\hat{X}_g - \bar{X})}{N} \right) \Gamma^P \]  
(3.C.35)

\[ = \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \frac{\hat{X}_g}{N} \left[ 0 \right] + \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \left( \frac{\hat{X}_g^t (\hat{X}_g - \bar{X})}{N} \right) \Gamma^P \]  
(3.C.36)

\[ + \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \left( \frac{\hat{X}_g^t (\hat{X}_g - \bar{X})}{N} \right) \Gamma^P - \left( \frac{\hat{X}_g^t (\hat{X}_g - \bar{X})}{N} \right) \Gamma^P \]  
(3.C.37)

Consider the term (3.C.36). Define \( D = \frac{\hat{B}_g \hat{B}_e^{N}}{-N} - \frac{\hat{B}_g \hat{B}_e^{-1}}{-D^{-1} \hat{B}_e^{N}} \), \( \hat{\mu}_c = \frac{\hat{V}_c}{N} \) and notice that
\[ \frac{1}{N} \hat{B}_e = O_p \left( \frac{1}{\sqrt{N}} \right), \quad \frac{\hat{B}_e^{N}}{N} = O_p \left( \frac{1}{\sqrt{N}} \right), \quad \text{and} \quad D^{-1} = O_p(1). \]  
(3.C.38)

Then,
\[ \left( \frac{\hat{X}_g^t \hat{X}_g}{N} \right)^{-1} \frac{\hat{X}_g}{N} \left[ 0 \right] = \left[ \frac{1}{N} \right] \frac{\hat{\mu}_c + O_p(N^{-1})}{D^{-1} \hat{B}_e^{N}} + \left[ 0 \right] \]  
(3.C.39)

Therefore, using Assumption 3.6(i), and using the results in (3.C.10), we have
\[ \sqrt{N} \left( \hat{\Gamma}_g - \left[ \frac{\hat{\mu}_c - \mu_c + \hat{\mu}_c}{0} \right] \right) = \sqrt{N} \left( \hat{\Gamma}_g - \left[ \frac{\hat{\mu}_c - \mu_c + \hat{\mu}_c}{0} \right] \right) = \]  
(3.C.39)

where we define \( M_{1, N} = I_N - \frac{1}{N} \hat{X}_g^t \hat{X}_g \). We now need to derive the variance of (3.C.39). For the first term in (3.C.39), we can use the same result of Theorem 3.1. Therefore,
The variance of the second term in (3.C.39) satisfies

\[
\text{Var}\left(D^{-1}P_g \epsilon M_{1,N} c \sqrt{N}\right) = D^{-1}P_g E \left(\frac{\epsilon M_{1,N} c c' M_{1,N} \epsilon}{N}\right) P_g D^{-1}
\]

Denoting by \( A_c = M_{1,N} c c' M_{1,N} \), with generic \((i,j)\)-th element \(a_{ij}\), we have

\[
E(\epsilon A_c) = \sum_{i,j=1}^{N} E(\epsilon_i a_{ij} \epsilon_j')
\]

\[
= \sum_{i=1}^{N} \epsilon_i \epsilon_i' a_i^2 + \sum_{i \neq j=1}^{N} \epsilon_i \epsilon_j' a_{ij}
\]

\[
= \sum_{i=1}^{N} \epsilon_i \epsilon_i' a_i^2 + o(N)
\]

Let \( \nu_c = \frac{1}{N} \sum_{i=1}^{N} \epsilon_i^2 a_i^2 \), and notice that \( D \rightarrow_p \sigma^2 \left( \hat{G}' \hat{G} \right)^{-1} \), then

\[
\text{Var}\left(D^{-1}P_g \epsilon M_{1,N} c \sqrt{N}\right) = \frac{\nu_c}{\sigma^4} \hat{G}' \hat{G} \equiv W
\]

Finally,

\[
\text{Cov}\left(D^{-1}P_g \epsilon M_{1,N} c \sqrt{N}, P_g' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \frac{1_T}{T}\right) = 0_{K_s}
\]

by Assumption 3.4(vi), and

\[
\text{Cov}\left(D^{-1}P_g \epsilon M_{1,N} c \sqrt{N}, \frac{1}{\sqrt{N}} \frac{1_T}{T} \frac{1_T}{T}\right) = 0_{K_s}
\]

since \( P_g' \frac{1_T}{T} = 0_{K_s} \) by construction. Hence

\[
\text{Var}\left(\sqrt{N} \left( \Gamma_g - \begin{bmatrix} \mu_c \\ 0_{K_s} \end{bmatrix} \right) \right) = V + W
\]

(3.C.41)

**Proof of Theorem 3.6.** (i) Notice that

\[
\hat{B}_f = R' \mathcal{M}_T F (F' \mathcal{M}_T F)^{-1}
\]

\[
= B_f + \epsilon' \mathcal{M}_T F (F' \mathcal{M}_T F)^{-1}
\]

\[
= B_f + \epsilon' P_f
\]

(3.C.42)
\[
\hat{B}_g = R'M_T (G'M_T G)^{-1} \\
= B_g + \epsilon'M_T G (G'M_T G)^{-1} \\
= B_g + \epsilon' P_g \\
= 0_{N \times K_a} + \epsilon' P_g
\] (3.43)

Then, starting from the definition in (3.28), we have

\[
\hat{X}_{f,g} = \begin{bmatrix}
1 & \frac{1}{N} \hat{B}_f & \frac{1}{N} \hat{B}_g \\
\hat{B}^T_{f} \hat{B}_f & \frac{1}{N} \hat{B}^T_{f} \hat{B}_g & \frac{1}{N} \hat{B}^T_{g} \hat{B}_g \\
\hat{B}^T_{g} \hat{B}_f & \frac{1}{N} \hat{B}^T_{g} \hat{B}_g & \frac{1}{N} \hat{B}^T_{g} \hat{B}_g \\
\end{bmatrix}
\] (3.44)

(3.42) and (3.43) imply that

\[
\frac{1}{N} \hat{B}_f = \frac{1}{N} B_f + \frac{1}{N} \epsilon' P_f \\
= \frac{1}{N} \sum_{i=1}^{N} \beta'_i + \frac{1}{N} \sum_{i=1}^{N} \epsilon'_i P_f \\
= \frac{1}{N} \sum_{i=1}^{N} \beta'_i + O_p \left( \frac{1}{\sqrt{N}} \right) \\
\rightarrow_p \mu_{\beta_f}
\]

and

\[
\frac{1}{N} \hat{B}_g = \frac{1}{N} B_g + \frac{1}{N} \epsilon' P_g \\
= 0_{K_g} + \frac{1}{N} \sum_{i=1}^{N} \epsilon'_i P_g \\
= 0_{K_g} + O_p \left( \frac{1}{\sqrt{N}} \right) \\
\rightarrow_p 0_{K_g'}
\]

Moreover, under (3.29) and using Assumption 3.4
\[
\frac{\hat{B}_f \hat{B}_g}{N} = \frac{B_f' B_g}{N} + \frac{(\hat{B}_f - B_f)' \hat{B}_g}{N} \\
= \frac{B_f' \epsilon' P_g}{N} + \frac{P_f' \epsilon \hat{B}_g}{N} \\
= \frac{B_f' \epsilon' P_g}{N} + \frac{P_f' \epsilon \hat{B}_g}{N} \\
= \frac{\sum_{i=1}^{N} \beta_{f_i} \epsilon_{i}' P_g}{N} + \frac{P_f' \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_{i}' P_g}{N} \\
= O_p \left( \frac{1}{\sqrt{N}} \right) + \frac{P_f' \frac{1}{N} \sum_{i=1}^{N} \left( \epsilon_{i}' \epsilon_{i} - \sigma_i^2 \right) P_g}{N} \\
+ \frac{1}{N} \sum_{i=1}^{N} \left( \sigma_i^2 - \sigma^2 \right) P_f' P_g + \sigma^2 P_f' P_g \\
= \sigma^2 (\hat{F}' \hat{F})^{-1} \hat{F}' \hat{G} (\hat{G}' \hat{G})^{-1} + O_p \left( \frac{1}{\sqrt{N}} \right) + \frac{1}{\sqrt{N}} \\
\rightarrow_p \sigma^2 (\hat{F}' \hat{F})^{-1} \hat{F}' \hat{G} (\hat{G}' \hat{G})^{-1} \\
\equiv 0_{K_f \times K_g}, \quad (3.45)
\]

\[
\frac{\hat{B}_f \hat{B}_g}{N} = \frac{P_f' \epsilon' P_g}{N} \\
= \frac{P_f' \frac{1}{N} \sum_{i=1}^{N} \left( \epsilon_{i}' \epsilon_{i} - \sigma_i^2 \right) P_g}{N} + \frac{1}{N} \sum_{i=1}^{N} \left( \sigma_i^2 - \sigma^2 \right) P_f' P_g + \sigma^2 P_f' P_g \\
= \frac{1}{\sqrt{N}} \sigma^2 (\hat{G}' \hat{G})^{-1} - O_p \left( \frac{1}{\sqrt{N}} \right) + \sigma^2 P_f' P_g \\
\rightarrow_p \sigma^2 (\hat{G}' \hat{G})^{-1}. \quad (3.46)
\]

Finally, using the same arguments and defining \( \Sigma_{\delta_f} = \text{plim}_{\frac{1}{N} \sum_{i=1}^{N} \beta_{f_i} \beta_{f_i}'}, \) we have

\[
\frac{\hat{B}_f \hat{B}_f}{N} \rightarrow_p \Sigma_{\delta_f} + \sigma^2 (\hat{F}' \hat{F})^{-1}. \quad (3.47)
\]

Then, using (3.125) - (3.125) in (3.44), we have

\[
\frac{\hat{X}_f' \hat{X}_g}{N} \rightarrow_p \left[ \begin{array}{ccc}
1 & \mu_{\beta_f}' & 0_{K_f}' \\
\mu_{\beta_f} & \Sigma_{\beta_f} + \sigma^2 (\hat{F}' \hat{F})^{-1} & 0_{K_f \times K_g}' \\
0_{K_f} & 0_{K_f \times K_g} & \sigma^2 (\hat{G}' \hat{G})^{-1}
\end{array} \right] \equiv \Sigma_{X_f s} + \Lambda_{f g}, \quad (3.48)
\]

where we define

\[
\Sigma_{X_f s} = \left[ \begin{array}{cc}
\Sigma_{X_f} & 0_{(K_f + 1) \times K_g} \\
0_{K_f \times (K_f + 1)} & 0_{K_g \times K_g}
\end{array} \right], \quad \Sigma_{X_f} = \left[ \begin{array}{c}
1 \\
\mu_{\beta_f}' \\
\Sigma_{\beta_f}
\end{array} \right]
\]
and where

\[
\Lambda_{fs} = \begin{bmatrix}
0 & 0'_{K_f} & 0'_{K_s} \\
0_{K_f} & \sigma^2 (P^T F)^{-1} & 0_{K_f \times K_s} \\
0_{K_s} & 0_{K_s \times K_f} & \sigma^2 (G' M_T G)^{-1}
\end{bmatrix}.
\]

Consider now \( \dot{\hat{X}}_{f_2}^T R \). Using (3.27), we have

\[
\frac{\dot{X}_{f_2}^T \dot{R}}{\dot{N}} = \left( \frac{\dot{X}_{f_2}^T \dot{X}_{f_2}}{\dot{N}} \right) \Gamma^p + \frac{\dot{X}_{f_2}^T}{\dot{N}} \left( \dot{\xi} - (\dot{\hat{X}}_{f_2} - X) \Gamma^p \right)
\]

where

\[
\frac{\dot{X}_{f_2}^T \dot{\xi}}{\dot{N}} = \frac{X_{f_2}^T}{\dot{N}} + (\hat{X}_{f_2} - X_{f_2}) \frac{\dot{\xi}}{\dot{N}}
\]

\[
= \begin{bmatrix}
1_N' \\
0_{K_s \times N}
\end{bmatrix} \frac{\dot{\xi}}{\dot{N}} + \begin{bmatrix}
0'_{N} \\
P'^T_{f_2} \frac{\dot{\xi}}{\dot{N}}
\end{bmatrix} \frac{\dot{\xi}}{\dot{N}} = \begin{bmatrix}
\frac{1_N' \dot{\xi}}{\dot{N}} \\
\frac{P'^T_{f_2} \dot{\xi}}{\dot{N}}
\end{bmatrix} + \begin{bmatrix}
0'_{N} \\
P'^T_{f_2} \frac{\dot{\xi}}{\dot{N}}
\end{bmatrix}
\]

\[
= O_p \left( \frac{1}{\sqrt{N}} \right) + P'_{f_2} \frac{1}{N} \sum_{i=1}^{N} (\varepsilon_i \dot{\varepsilon}_i - \sigma_i^2) \frac{\dot{\xi}}{\dot{N}} + P'_{f_2} \frac{1}{N} \sum_{i=1}^{N} (\sigma_i^2 - \sigma^2) \frac{\dot{\xi}}{\dot{N}} + \sigma^2 \frac{\dot{\xi}}{\dot{N}}
\]

\[
= O_p \left( \frac{1}{\sqrt{N}} \right) + \begin{bmatrix}
O_p \left( \frac{1}{\sqrt{N}} \right) \\
O_p \left( \frac{1}{\sqrt{N}} \right)
\end{bmatrix} + O_{K_f}
\]

and where

\[
\frac{\dot{X}_{f_2}^T (X - \dot{X}_{f_2})}{\dot{N}} \Gamma^p = \frac{X_{f_2}^T (X - \dot{X}_{f_2})}{\dot{N}} \Gamma^p + \frac{X^T - \dot{X}_{f_2})}{\dot{N}} \frac{\dot{X}_{f_2}^T - X}{\dot{N}} \Gamma^p
\]

\[
= \frac{1}{N} \begin{bmatrix}
1_N' \\
0_{K_s \times N}
\end{bmatrix} \begin{bmatrix}
0_N, -\varepsilon P_f, -\varepsilon P_g
\end{bmatrix} \Gamma^p - \frac{1}{N} \begin{bmatrix}
0'_{N} \\
P'^T_{f_2} \frac{\dot{\xi}}{\dot{N}}
\end{bmatrix} \begin{bmatrix}
0_{K_f} \\
0_{K_s \times K_f}
\end{bmatrix} \Gamma^p
\]

\[
= \begin{bmatrix}
0 & -\frac{1}{N} \varepsilon P_f & -\frac{1}{N} \varepsilon P_g \\
0_{K_f} & -\frac{1}{N} \varepsilon P_f & -\frac{1}{N} \varepsilon P_g \\
0_{K_s} & 0_{K_s \times K_f}
\end{bmatrix} \Gamma^p - \begin{bmatrix}
0'_{K_f} \\
0_{K_f} \frac{P'^T_{f_2} \dot{\xi}}{\dot{N}}
\end{bmatrix} \begin{bmatrix}
0_{K_s} \\
0_{K_s \times K_f}
\end{bmatrix} \Gamma^p
\]

\[
\Rightarrow \frac{1}{\gamma_{(1+K_f+K_s)}} \begin{bmatrix}
0'_{K_f} \\
0_{K_f} \sigma^2 (P^T F)^{-1} \\
0_{K_s \times K_f}
\end{bmatrix} \gamma_{(1+K_f+K_s)} = \begin{bmatrix}
0'_{K_f} \\
0_{K_f} \sigma^2 (P^T F)^{-1} \\
0_{K_s \times K_f}
\end{bmatrix} \gamma_{(1+K_f+K_s)}.
\]

Hence, using (3.126), together with (3.48), (3.126) and (3.126) we have
\[
\hat{\Gamma}_P^{F} \rightarrow_p \Gamma^P - \begin{bmatrix}
1 & \mu'_{\beta_f} & 0'_{K_f} \\
\mu_{\beta_f} & \Sigma_{\beta_f} + \sigma^2(F'M_TF)^{-1} & 0'_{K_f} \\
0_{K_f} & 0_{K_f \times K_f} & \Sigma^2(G'M_GG)^{-1}
\end{bmatrix}^{-1} \begin{bmatrix}
0 & 0'_{K_f} & 0'_{K_f} \\
0_{K_f} & \sigma^2(F'M_TF)^{-1} & 0'_{K_f} \\
0_{K_f \times K_f} & 0_{K_f \times K_f} & \sigma^2(G'M_GG)^{-1}
\end{bmatrix} \Gamma^P
\]

Notice that

\[
\begin{bmatrix}
1 & \mu'_{\beta_f} & 0'_{K_f} \\
\mu_{\beta_f} & \Sigma_{\beta_f} + \sigma^2(F'M_TF)^{-1} & 0'_{K_f} \\
0_{K_f} & 0_{K_f \times K_f} & \Sigma^2(G'M_GG)^{-1}
\end{bmatrix}^{-1} = \begin{bmatrix}
1 & \mu'_{\beta_f} \\
\mu_{\beta_f} & \Sigma_{\beta_f} + \sigma^2(F'M_TF)^{-1} \\
0_{K_f} & 0_{K_f \times K_f}
\end{bmatrix}^{-1} \begin{bmatrix}
0'_{K_f} \\
0_{K_f} & \sigma^2(F'M_TF)^{-1} \\
0_{K_f \times K_f}
\end{bmatrix} = \begin{bmatrix}
A_{11} & A_{12} & 0'_{K_f} \\
A_{21} & A_{22} & 0_{K_f \times K_f}
\end{bmatrix}
\]

where

\[
A_{11} = 1 + \mu'_{\beta_f} D^{-1} \mu_{\beta_f}
\]
\[
A_{21} = -D^{-1} \mu_{\beta_f}
\]
\[
A_{22} = D^{-1}
\]
\[
D = \Sigma_{\beta_f} - \mu_{\beta_f} \mu'_{\beta_f} + \sigma^2(F'M_TF)^{-1}
\]

It implies that

\[
\hat{\Gamma}_P^{F} \rightarrow_p \Gamma^P - \begin{bmatrix}
A_{11} & A_{12} & 0'_{K_f} \\
A_{21} & A_{22} & 0_{K_f \times K_f}
\end{bmatrix} \begin{bmatrix}
0 & 0'_{K_f} & 0'_{K_f} \\
0_{K_f} & \sigma^2(F'M_TF)^{-1} & 0'_{K_f} \\
0_{K_f \times K_f} & 0_{K_f \times K_f} & \sigma^2(G'M_GG)^{-1}
\end{bmatrix} \Gamma^P
\]

\[
= \Gamma^P - \begin{bmatrix}
0 & \sigma^2 A_{12} (F'M_TF)^{-1} & 0'_{K_f} \\
0_{K_f} & \sigma^2 A_{22} (F'M_TF)^{-1} & 0_{K_f \times K_f}
\end{bmatrix} \begin{bmatrix}
0'_{K_f} \\
0_{K_f} & \sigma^2 (F'M_TF)^{-1} & 0'_{K_f} \\
0_{K_f \times K_f} & 0_{K_f \times K_f} & \sigma^2 (G'M_GG)^{-1}
\end{bmatrix} \Gamma^P
\]

\[
= \begin{bmatrix}
\gamma_0 & 0'_{K_f} \\
\gamma_{1f} & \sigma^2 A_{12} (F'M_TF)^{-1} \gamma_{1f} \\
\gamma_{1s} & \sigma^2 A_{22} (F'M_TF)^{-1} \gamma_{1s}
\end{bmatrix} \begin{bmatrix}
\gamma_0 & \sigma^2 A_{12} (F'M_TF)^{-1} \gamma_{1f} \\
\sigma^2 A_{12} (F'M_TF)^{-1} \gamma_{1f} & \gamma_{1f} \\
\sigma^2 A_{22} (F'M_TF)^{-1} \gamma_{1s} & \gamma_{1s}
\end{bmatrix} \Gamma^P
\]

\[
= \begin{bmatrix}
\gamma_0 + \sigma^2 \mu_{\beta_f} D^{-1} (F'M_TF)^{-1} \gamma_{1f} \\
(I_{K_f} - \sigma^2 D^{-1} (F'M_TF)^{-1}) \gamma_{1f} \\
0_{K_f}
\end{bmatrix}
\]

Defining \( d_0 = \sigma^2 \mu_{\beta_f} D^{-1} (F'M_TF)^{-1} \gamma_{1f} \) and \( d_1 = -\sigma^2 D^{-1} (F'M_TF)^{-1} \gamma_{1f} \) concludes the proof of part (i).

(ii) Starting from the definition in (3.28) and using the result in (3.126) we have
\[ \hat{\Gamma}_{fs} = \Gamma^p + \left( \frac{\hat{X}_{fs} \hat{X}_{fs}}{N} \right)^{-1} \frac{\hat{X}_{fs}}{N} \left( \hat{\tau} - \hat{X}_{fs} - X \right) \Gamma^p \] (3.57)

which implies that

\[
\hat{\Gamma}_{fs} - \begin{bmatrix} \gamma_0 + d_0 \\ \gamma_{1f} + d_1 \\ 0_{K_s} \end{bmatrix} = \left( \frac{\hat{X}_{fs} \hat{X}_{fs}}{N} \right)^{-1} \frac{\hat{X}_{fs}}{N} \left( \hat{\tau} - \hat{X}_{fs} - X \right) \Gamma^p + \begin{bmatrix} \gamma_0 + d_0 \\ \gamma_{1f} + d_1 \\ 0_{K_s} \end{bmatrix} = \left( \frac{\hat{X}_{fs} \hat{X}_{fs}}{N} \right)^{-1} \left( \frac{\hat{X}_{fs}}{N} \left( \hat{\tau} - \hat{X}_{fs} - X \right) \Gamma^p - \begin{bmatrix} 0 \\ d_0 \\ -\gamma_{1f} \end{bmatrix} \right) \] (3.58)

Consider first \( \frac{\hat{X}_{fs}}{N} \hat{\tau} \). We have

\[
\frac{\hat{X}_{fs}}{N} \hat{\tau} = \begin{bmatrix} \frac{\hat{\gamma}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} \\ \frac{\hat{\beta}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} \\ 0_{K_s} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{\hat{p}_s^{\alpha f}}{N} \frac{\hat{\imath}_f}{N} \\ \frac{\hat{p}_s^{\alpha f}}{N} \frac{\hat{\imath}_f}{N} \end{bmatrix} = \begin{bmatrix} \frac{\hat{\gamma}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} + \frac{\hat{p}_s^{\alpha f}}{N} \frac{\hat{\imath}_f}{N} \\ \frac{\hat{\beta}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} + \frac{\hat{p}_s^{\alpha f}}{N} \frac{\hat{\imath}_f}{N} \end{bmatrix} \] (3.59)

Moreover

\[
-\frac{\hat{X}_{fs}}{N} \left( \hat{X}_{fs} - X \right) \Gamma^p = \begin{bmatrix} 0 & \frac{\hat{\gamma}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} & \frac{\hat{\gamma}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} \\ 0_{K_f} & -\frac{\hat{\beta}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} & -\frac{\hat{\beta}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} \end{bmatrix} \Gamma^p - \begin{bmatrix} 0 \\ 0_{K_s} \end{bmatrix} \begin{bmatrix} \hat{P}_f \hat{\gamma}_f \\ P_s \hat{\gamma}_s \end{bmatrix} \] (3.60)

Therefore

\[
\frac{\hat{X}_{fs}}{N} \left( \hat{\tau} - \hat{X}_{fs} - X \right) \Gamma^p = \begin{bmatrix} \frac{\hat{\gamma}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} + \frac{\hat{p}_s^{\alpha f}}{N} \frac{\hat{\imath}_f}{N} - \frac{\hat{\gamma}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} \hat{P}_f \hat{\gamma}_f - \frac{\hat{\beta}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} \hat{P}_s \hat{\gamma}_s \\ \frac{\hat{\beta}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} \hat{P}_f \hat{\gamma}_f - \frac{\hat{\beta}_s \hat{\alpha}_f}{N} \frac{\hat{\imath}_f}{N} \hat{P}_s \hat{\gamma}_s - \frac{\hat{p}_s^{\alpha f}}{N} \frac{\hat{\imath}_f}{N} \hat{P}_s \hat{\gamma}_s \\ 0_{K_s} \end{bmatrix} \] (3.61)
Using (3.64), we have

\[
\left( \frac{\hat{X}_{f}^t \hat{X}_{f}}{N} \right) \left[ \begin{array}{c} d_0 \\ d_1 \\ \gamma_{f_s}^p \end{array} \right] = \left[ \begin{array}{ccc} 1 & \frac{\gamma_{f_s}^p}{N} & \frac{\gamma_{f_s}^p}{N} \\ \frac{\hat{\beta}_{1}^{f_s}}{N} & \frac{\hat{\beta}_{1}^{f_s}}{N} & \frac{\hat{\beta}_{1}^{f_s}}{N} \\ \frac{\hat{\beta}_{2}^{f_s}}{N} & \frac{\hat{\beta}_{2}^{f_s}}{N} & \frac{\hat{\beta}_{2}^{f_s}}{N} \end{array} \right] \left[ \begin{array}{c} d_0 \\ d_1 \\ \gamma_{f_s}^p \end{array} \right] = \left[ \begin{array}{c} d_0 + \frac{\gamma_{f_s}^p}{N} d_1 - \frac{\gamma_{f_s}^p}{N} \gamma_{f_s}^p \\ \frac{\hat{\beta}_{1}^{f_s}}{N} d_0 + \frac{\hat{\beta}_{1}^{f_s}}{N} d_1 - \frac{\hat{\beta}_{1}^{f_s}}{N} \gamma_{f_s}^p \\ \frac{\hat{\beta}_{2}^{f_s}}{N} d_0 + \frac{\hat{\beta}_{2}^{f_s}}{N} d_1 - \frac{\hat{\beta}_{2}^{f_s}}{N} \gamma_{f_s}^p \end{array} \right]. \tag{3.62}
\]

Using (3.61), (3.62), remembering that \( \hat{B}_g = \epsilon' P_g \) and noticing that \( \frac{\hat{\beta}_1^{f_s}}{N} \gamma_{f_s}^p - \frac{\hat{\beta}_1^{f_s}}{N} P_f \gamma_{f_s}^p - \frac{\hat{\beta}_2^{f_s}}{N} P_g \gamma_{f_s}^p = 0_{K_f} \), then (3.135) can be rewritten as

\[
\hat{\Gamma}_{f_s} = \left( \begin{array}{c} \gamma_{f_s}^0 + d_1 \\ \gamma_{f_s}^p \end{array} \right) = \left( \frac{\hat{X}_{f}^t \hat{X}_{f}}{N} \right)^{-1} \left[ \begin{array}{cc} \frac{1}{T} - \frac{\gamma_{f_s}^p}{N} P_f \gamma_{f_s}^p - d_0 - \frac{\gamma_{f_s}^p}{N} d_1 \\ \frac{1}{T} \end{array} \right], \tag{3.63}
\]

Consider first the term \( I_1 \). Using (3.56) and defining

\[
Q_f = \frac{1}{T} - P_f D^{-1} \left( \Sigma_{\beta_f} - \mu_{\beta_f} \mu_{\beta_f}' \right) \gamma_{f_s}^p
\]

we have

\[
I_1 = \frac{1}{N} \left[ \frac{1}{T} - \frac{\gamma_{f_s}^p}{N} P_f \gamma_{f_s}^p - d_0 - \frac{\gamma_{f_s}^p}{N} d_1 \right]
\]

\[
= \frac{1}{N} \left[ \frac{1}{T} - \frac{\gamma_{f_s}^p}{N} P_f \gamma_{f_s}^p - \sigma^2 \mu_{\beta_f} D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right] + \frac{1}{N} \left( \frac{1}{T} - \frac{\gamma_{f_s}^p}{N} P_f \gamma_{f_s}^p - \sigma^2 \mu_{\beta_f} D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left[ \frac{1}{T} - \frac{\gamma_{f_s}^p}{N} P_f \gamma_{f_s}^p - \sigma^2 \mu_{\beta_f} D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right] + \frac{1}{N} \left( \frac{1}{T} - \frac{\gamma_{f_s}^p}{N} P_f \gamma_{f_s}^p - \sigma^2 \mu_{\beta_f} D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left[ \frac{1}{T} - P_f \left( \gamma_{f_s}^p + d_1 \right) \right] + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left[ \frac{1}{T} - P_f \gamma_{f_s}^p \right] + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left[ \frac{1}{T} - P_f \gamma_{f_s}^p \right] + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left( \frac{1}{T} - P_f \gamma_{f_s}^p \right) + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left[ \frac{1}{T} - P_f \gamma_{f_s}^p \right] + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left( \frac{1}{T} - P_f \gamma_{f_s}^p \right) + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left( \frac{1}{T} - P_f \gamma_{f_s}^p \right) + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left( \frac{1}{T} - P_f \gamma_{f_s}^p \right) + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left( \frac{1}{T} - P_f \gamma_{f_s}^p \right) + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left( \frac{1}{T} - P_f \gamma_{f_s}^p \right) + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left( \frac{1}{T} - P_f \gamma_{f_s}^p \right) + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]

\[
= \frac{1}{N} \left( \frac{1}{T} - P_f \gamma_{f_s}^p \right) + \sigma^2 \left( \frac{1}{N} B_f \gamma_{f_s}^p - \mu_{\beta_f} \right) \left( D^{-1} (F' M_T F)^{-1} \gamma_{f_s}^p \right)
\]
Moreover, using \( d_0 = \sigma^2 \mu_{\beta_j} D^{-1} (F'M_T F)^{-1} \gamma_{i_j}^P \), and \( d_1 = -\sigma^2 D^{-1} (F'M_T F)^{-1} \gamma_{i_j}^P \),

\[
I_2 = \frac{B'_j}{N} \frac{\gamma_{i_j}^T}{T} + \frac{P'_j \epsilon' \gamma_{i_j}^T}{T} - \frac{B'_j}{N} P_f \gamma_{i_j}^P - \frac{P'_j \epsilon' \gamma_{i_j}^P}{N} - \frac{B'_j \gamma_{i_j}^T}{N} d_0 - \frac{B'_j \hat{B}_j}{N} d_1 
\]

(3.65)

\[
= \frac{B'_j}{N} \frac{\gamma_{i_j}^T}{T} + \frac{P'_j \epsilon' \gamma_{i_j}^T}{T} - \frac{B'_j}{N} P_f \gamma_{i_j}^P - \frac{P'_j \epsilon' \gamma_{i_j}^P}{N} - \frac{B'_j \gamma_{i_j}^T}{N} (\sigma^2 \mu_{\beta_j} D^{-1} P_f \gamma_{i_j}^P) 
\]

(3.66)

\[
- \frac{B'_j \gamma_{i_j}^T}{N} \frac{(\sigma^2 \mu_{\beta_j} D^{-1} P_f \gamma_{i_j}^P)}{N} - \frac{P'_j \epsilon' \gamma_{i_j}^T}{N} \frac{(\sigma^2 \mu_{\beta_j} D^{-1} P_f \gamma_{i_j}^P)}{N} 
\]

(3.67)

\[
+ \frac{B'_j \gamma_{i_j}^T}{N} \frac{(\sigma^2 D^{-1} P_f \gamma_{i_j}^P)}{N} + \frac{P'_j \epsilon' \gamma_{i_j}^T}{N} \frac{(\sigma^2 D^{-1} P_f \gamma_{i_j}^P)}{N} 
\]

(3.68)

\[
+ \frac{B'_j \gamma_{i_j}^T}{N} \frac{(\sigma^2 D^{-1} P_f \gamma_{i_j}^P)}{N} - \frac{P'_j \gamma_{i_j}^P}{N} \frac{(\sigma^2 D^{-1} P_f \gamma_{i_j}^P)}{N} 
\]

(3.69)

\[
= \frac{B'_j}{N} \left( \frac{\gamma_{i_j}^T}{T} - P_f \gamma_{i_j}^P + P_f (\sigma^2 D^{-1} P_f \gamma_{i_j}^P) \right) 
\]

(3.70)

\[
+ \frac{P'_j \epsilon' \gamma_{i_j}^T}{T} - P_f \epsilon' \gamma_{i_j}^T + P'_j \epsilon' \gamma_{i_j}^P - \frac{P'_j \epsilon' \gamma_{i_j}^P}{N} (\sigma^2 D^{-1} P_f \gamma_{i_j}^P) 
\]

(3.71)

\[
- \frac{B'_j \gamma_{i_j}^T}{N} (\sigma^2 \mu_{\beta_j} D^{-1} P_f \gamma_{i_j}^P) + \frac{B'_j \gamma_{i_j}^T}{N} (\sigma^2 D^{-1} P_f \gamma_{i_j}^P) 
\]

(3.72)

\[
+ \left( \frac{P'_j \gamma_{i_j}^P}{N} - \frac{P'_j \epsilon' \gamma_{i_j}^P}{N} \right) (\sigma^2 D^{-1} P_f \gamma_{i_j}^P) 
\]

(3.73)

using \( Q_f = \frac{1}{T} - P_f \left( \gamma_{i_j}^P + d_1 \right) \), then

\[
(3.70) = \frac{B'_j}{N} Q_f, 
\]

and

\[
(3.71) = P'_f \left( \frac{\epsilon' \gamma_{i_j}^T}{N} - \sigma^2 I_T \right) Q_f - \sigma^2 DD^{-1} P_f \gamma_{i_j}^P + \sigma^2 P_f \gamma_{i_j}^P = P'_f \left( \frac{\epsilon' \gamma_{i_j}^T}{N} - \sigma^2 I_T \right) Q_f - \sigma^2 (\hat{F}' \hat{F})^{-1} (\gamma_{i_j}^P + d_1) 
\]

Moreover, rewriting (3.72) as

\[
(3.72) = \left( \frac{B'_j B_f}{N} - \frac{B'_j \gamma_{i_j}^T}{N} \mu_{\beta_j} \right) (\sigma^2 D^{-1} P_f \gamma_{i_j}^P) 
\]

\[
= \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 \mu_{\beta_j} D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
- \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \frac{B'_j B_f}{N} \right) \left( \frac{B'_j \gamma_{i_j}^T}{N} \right) \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

\[
+ \left( \sigma^2 D^{-1} P_f \gamma_{i_j}^P \right) 
\]

then,
\[
I_2 = \frac{B'_f \epsilon'_f}{N} Q_f + P_f' \left( \frac{\epsilon'_f}{N} - \sigma^2 I_T \right) Q_f - \sigma^2 (\bar{F}' \bar{F})^{-1} (\gamma'_f + d_1) \\
- \left( \frac{B'_f B_f}{N} - \frac{B'_f 1 \epsilon'_f}{N} \mu_{d_T} + \mu_{d_T} \mu_{d_T} \right) d_1 + \sigma^2 (\bar{F}' \bar{F})^{-1} (\gamma'_f + d_1) \\
+ \left( \frac{P_f' \epsilon_f B_f}{N} - \frac{P_f' 1 \epsilon_f}{N} \mu_{d_T} \right) (\sigma^2 D^{-1} P_f P_f' \gamma_f) \\
= \frac{B'_f \epsilon'_f}{N} Q_f + P_f' \left( \frac{\epsilon'_f}{N} - \sigma^2 I_T \right) Q_f \\
- \left( \frac{B'_f B_f}{N} - \frac{B'_f 1 \epsilon'_f}{N} \mu_{d_T} + \mu_{d_T} \mu_{d_T} \right) d_1 \\
- \left( \frac{P_f' \epsilon_f B_f}{N} - \frac{P_f' 1 \epsilon_f}{N} \mu_{d_T} \right) d_1
\]

Finally, using Assumption 3.29, we have

\[
I_3 = \frac{P'_s \epsilon'_s 1 \epsilon'_f}{N T} - \frac{P'_s \epsilon'_s}{N T} P_f' \gamma_f - \frac{P'_s 1 \epsilon'_f}{N} (\sigma^2 \mu_{d_T} D^{-1} P_f' P_f' \gamma_f) \\
+ \frac{P'_s \epsilon'_s B_f}{N} (\sigma^2 D^{-1} P_f' P_f' \gamma_f) + \frac{P'_s \epsilon'_s P_f'}{N} (\sigma^2 D^{-1} P_f' P_f' \gamma_f) \\
= \frac{P'_s \epsilon'_s}{N} \left( \frac{\epsilon'_s}{N} - \sigma^2 I_T \right) Q_f - \left( \frac{P'_s \epsilon_f B_f}{N} - \frac{P'_s 1 \epsilon_f}{N} \mu_{d_T} \right) d_1 \\
\text{(3.74)}
\]

Hence, putting all together

\[
\sqrt{N} \left( \tilde{F}_{fg} - \begin{bmatrix} \gamma_0 & d_0 \\ \gamma'_f & d_1 \\ 0_{K_s} \end{bmatrix} \right) = \\
\left( \frac{\hat{X}_{fg}}{N} \right)^{-1} \begin{bmatrix} \frac{\sqrt{N}}{N} Q_f - \sqrt{N} \left( \frac{P'_s \epsilon_f B_f}{N} - \frac{P'_s 1 \epsilon_f}{N} \mu_{d_T} \right) d_1 + P_f' \sqrt{N} \left( \frac{\epsilon'_s}{N} - \sigma^2 I_T \right) Q_f \\
\frac{P'_s \sqrt{N} \left( \frac{\epsilon'_s}{N} - \sigma^2 I_T \right) Q_f - \sqrt{N} \left( \frac{P'_s \epsilon_f B_f}{N} - \frac{P'_s 1 \epsilon_f}{N} \mu_{d_T} \right) d_1} \end{bmatrix} + o_p(1) \\
\text{(3.75)}
\]

Notice that (3.75) has zero mean. Moreover,

\[
\begin{bmatrix} \frac{\sqrt{N}}{N} Q_f - \sqrt{N} \left( \frac{P'_s \epsilon_f B_f}{N} - \frac{P'_s 1 \epsilon_f}{N} \mu_{d_T} \right) d_1 + P_f' \sqrt{N} \left( \frac{\epsilon'_s}{N} - \sigma^2 I_T \right) Q_f \\
\frac{P'_s \sqrt{N} \left( \frac{\epsilon'_s}{N} - \sigma^2 I_T \right) Q_f - \sqrt{N} \left( \frac{P'_s \epsilon_f B_f}{N} - \frac{P'_s 1 \epsilon_f}{N} \mu_{d_T} \right) d_1} \end{bmatrix} = \\
\begin{bmatrix} \frac{\sqrt{N}}{N} Q_f - \sqrt{N} \left( \frac{P'_s \epsilon_f B_f}{N} - \frac{P'_s 1 \epsilon_f}{N} \mu_{d_T} \right) d_1 \\
\frac{P'_s \sqrt{N} \left( \frac{\epsilon'_s}{N} - \sigma^2 I_T \right) Q_f - \sqrt{N} \left( \frac{P'_s \epsilon_f B_f}{N} - \frac{P'_s 1 \epsilon_f}{N} \mu_{d_T} \right) d_1} \end{bmatrix} + o_p(1) \\
\text{(3.76)}
\]

\[
\begin{bmatrix} \frac{\sqrt{N}}{N} Q_f - \sqrt{N} \left( \frac{P'_s \epsilon_f B_f}{N} - \frac{P'_s 1 \epsilon_f}{N} \mu_{d_T} \right) d_1 \\
\frac{P'_f \sqrt{N} \left( \frac{\epsilon'_s}{N} - \sigma^2 I_T \right) Q_f - \sqrt{N} \left( \frac{P'_s \epsilon_f B_f}{N} - \frac{P'_s 1 \epsilon_f}{N} \mu_{d_T} \right) d_1} \end{bmatrix} + \begin{bmatrix} 0 \\
\frac{P'_f \sqrt{N} \left( \frac{\epsilon'_s}{N} - \sigma^2 I_T \right) Q_f - \sqrt{N} \left( \frac{P'_s \epsilon_f B_f}{N} - \frac{P'_s 1 \epsilon_f}{N} \mu_{d_T} \right) d_1} \end{bmatrix},
\]
which implies that

\[
\begin{align*}
\text{Var} \left[ \frac{B'_{f'} Z_{f}^T}{\sqrt{N}} Q_f + P_f' \sqrt{N} \left( \frac{\alpha'_f}{\sqrt{N}} - \sigma^2 I_T \right) Q_f \right] & \\
= \text{Var} \left[ \frac{B'_{f'} Z_{f}^T}{\sqrt{N}} Q_f - \sqrt{N} \left( \frac{P_f' \beta_f}{N} \right) \mu_{\beta_f} d_1 \right] + \text{Var} \left[ P_f' \sqrt{N} \left( \frac{\alpha'_f}{\sqrt{N}} - \sigma^2 I_T \right) Q_f \right]
\end{align*}
\]

(3.3.77)

by Assumption 3.4(vi). Let us start by considering the first term on the right-hand side of (3.3.77). Using the fact that \( P_f' Q_f = 0_{K_f} \) and assuming that \( F' \tilde{G} = 0_{K_f \times K_s} \), we have

\[
\begin{align*}
\text{Var} \left[ \frac{B'_{f'} Z_{f}^T}{\sqrt{N}} Q_f & - \sqrt{N} \left( \frac{P_f' \beta_f}{N} \right) \mu_{\beta_f} d_1 \right] = \text{Var} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i' Q_f \right] \right] \\
\rightarrow_p & \left[ \begin{array}{cccc}
\sigma^2 Q_f & 0 & 0 & 0'_{K_s} \\
0 & \sigma^2 (Q_f' \mu_{\beta_f}) Q_f & 0_{K_f \times K_s} & 0_{K_f 	imes K_s} \\
0 & 0_{K_s 	imes K_f} & 0_{K_s} & 0_{K_s 	imes K_f} \\
0 & 0_{K_s 	imes K_f} & 0_{K_s 	imes K_f} & 0_{K_s 	imes K_f} \\
\end{array} \right] \\
& + \left[ \begin{array}{cccc}
0 & 0_{K_f} & 0_{K_s} & 0'_{K_s} \\
0_{K_f} & \sigma^2 d_1' \Sigma_{\beta_f} d_1 (F' \tilde{F})^{-1} - \sigma^2 \Sigma_{\beta_f} d_1 Q_f d_1' \Sigma_{\beta_f} & 0_{K_f 	imes K_s} & 0_{K_f 	imes K_s} \\
0_{K_s} & 0_{K_s} & 0_{K_s} & 0_{K_s} \\
0_{K_s} & 0_{K_s} & 0_{K_s} & 0_{K_s} \\
\end{array} \right] \\
& = \sigma^2 Q_f' Q_f \left[ \begin{array}{cccc}
0 & 0_{(K_f + 1) \times K_s} & 0_{K_f \times K_f} \\
0_{K_s \times (K_f + 1)} & 0_{K_s \times K_f} & 0_{K_s \times K_f} \\
\end{array} \right] + \sigma^2 \Omega
\end{align*}
\]

(3.3.78)

setting

\[
\Sigma_{f_g} = \left[ \begin{array}{cccc}
\Sigma_{X_f} & 0_{(K_f + 1) \times K_s} & 0_{K_f \times K_f} \\
0_{K_s \times (K_f + 1)} & 0_{K_s \times K_f} & 0_{K_s \times K_f} \\
\end{array} \right], \quad \text{with} \quad \Sigma_{X_f} = \left[ \begin{array}{cccc}
1 & \mu_{\beta_f} & 0'_{K_s} \\
0_{K_f} & d_1' \Sigma_{\beta_f} d_1 (F' \tilde{F})^{-1} - \Sigma_{\beta_f} d_1 Q_f d_1' \Sigma_{\beta_f} & 0_{K_f \times K_s} & 0_{K_f \times K_s} \\
0_{K_s} & 0_{K_s} & 0_{K_s} & 0_{K_s} \\
\end{array} \right]
\]

\[
\Omega = \left[ \begin{array}{cccc}
0 & 0_{K_f} & 0_{K_s} & 0'_{K_s} \\
0_{K_f} & d_1' \Sigma_{\beta_f} d_1 (F' \tilde{F})^{-1} - \Sigma_{\beta_f} d_1 Q_f d_1' \Sigma_{\beta_f} & 0_{K_f \times K_s} & 0_{K_f \times K_s} \\
0_{K_s} & 0_{K_s} & 0_{K_s} & 0_{K_s} \\
\end{array} \right]
\]

Consider now the second term on the right-hand side of (3.3.77). We have
\[ \text{Var} \left[ \frac{0}{P_f' \sqrt{N} \left( \frac{\hat{r}_f}{N} - \sigma^2 I_T \right) Q_f} \frac{0}{P_g' \sqrt{N} \left( \frac{\hat{r}_g}{N} - \sigma^2 I_T \right) Q_f} \right] \]
\[ = E \left[ \begin{bmatrix} 0 & 0_{K_f} \\ 0_{K_f} & P_f' \left( \frac{\hat{r}_f}{N} - \sigma^2 I_T \right) Q_f Q_f' \left( \frac{\hat{r}_f}{N} - \sigma^2 I_T \right) P_f \end{bmatrix} \begin{bmatrix} 0_{K_g} \\ 0_{K_g} \end{bmatrix} \right] \]
\[ \rightarrow_p \left[ \begin{bmatrix} 0 & 0_{K_f} \\ 0_{K_f} & (Q_f' \otimes P_f') U_f (Q_f \otimes P_f) \\ 0_{K_g} & (Q_f' \otimes P_f') U_f (Q_f \otimes P_f) \end{bmatrix} \right] \]  

(3.79)

This concludes the proof of part (ii). ■

**Proof of Theorem 3.7 (i)** To derive the asymptotic distribution of \( t_{k_g} \), it is convenient to rewrite it as

\[ t_{k_g} = \frac{\sqrt{N} \hat{r}_{k_g}}{\sqrt{\hat{s}_{fg}^{(2)} \left( \frac{\hat{r}_f \hat{r}_g}{N} \right)^{-1}} \left[ \hat{r}_f \hat{r}_g \right]} \]  

(3.80)

where \( \left( \frac{\hat{r}_f \hat{r}_g}{N} \right)_{[k_g]} \) denotes the \((k_g, k_g)\)-th element of the matrix \( \frac{\hat{r}_f \hat{r}_g}{N} \). Consider first \( \hat{s}_{fg}^{(2)} \) and notice that

\[ \hat{r}_f \hat{r}_g = \left( \hat{R} - \hat{X}_{fg} \hat{\Gamma}_{fg} \right)' \left( \hat{R} - \hat{X}_{fg} \hat{\Gamma}_{fg} \right) \]
\[ = \hat{R}' \hat{R} + \hat{\Gamma}_{fg}' \hat{X}_{fg} \hat{X}_{fg} \hat{\Gamma}_{fg} - 2 \hat{R}' \hat{X}_{fg} \hat{\Gamma}_{fg} \]
\[ = \hat{R}' \hat{R} + \hat{\Gamma}_{fg}' \hat{X}_{fg} \hat{X}_{fg} \hat{\Gamma}_{fg} - 2 \hat{\Gamma}_{fg}' \hat{X}_{fg} \hat{X}_{fg} \hat{\Gamma}_{fg} \]
\[ = \hat{R}' \hat{R} - \hat{\Gamma}_{fg}' \hat{X}_{fg} \hat{X}_{fg} \hat{\Gamma}_{fg}, \]

so that

\[ \frac{\hat{r}_f \hat{r}_g}{N} = \frac{\hat{R}' \hat{R}}{N} - \frac{\hat{\Gamma}_{fg}' \hat{X}_{fg} \hat{X}_{fg} \hat{\Gamma}_{fg}}{N}. \]  

(3.81)

Consider the first term on the right-hand-side of (3.81). Using the fact that \( \hat{R} = \hat{X}_{fg} \hat{\Gamma}_{fg} + \hat{\epsilon} \), then

\[ \frac{\hat{R}' \hat{R}}{N} = \frac{1}{N} (X_{fg} \hat{\Gamma}_{fg} + \hat{\epsilon})' (X_{fg} \hat{\Gamma}_{fg} + \hat{\epsilon}) \]
\[ = \frac{\hat{\epsilon}' \hat{\epsilon}}{N} + \hat{\Gamma}_{fg}' \hat{X}_{fg} \hat{X}_{fg} \hat{\epsilon} - 2 \hat{\Gamma}_{fg}' \hat{X}_{fg} \hat{\epsilon} \]
\[ = \frac{\hat{\epsilon}' \hat{\epsilon}}{N} + \hat{\Gamma}_{fg}' \hat{X}_{fg} \hat{X}_{fg} \hat{\epsilon} - 2 \hat{\Gamma}_{fg}' \hat{X}_{fg} \hat{\epsilon} = O_p \left( \frac{1}{\sqrt{N}} \right) \]
\[ \rightarrow_p \frac{\sigma^2}{T} + \gamma_0 + \gamma_1 \hat{r}_f \mu_{\beta_f} + 2 \gamma_0 \gamma_1 \hat{\epsilon} \]

This completes the proof.
since

\[
\frac{\tau'}{N} = \frac{\tau'}{T} \frac{\tau'}{T}
\]

\[\rightarrow_P \frac{\sigma^2}{T},\]

and

\[
\Gamma_{f_g}^P X_{f_g} X_{f_g} \Gamma_{f_g}^P = \begin{bmatrix}
\gamma_0, & \gamma_1, & \gamma_{1s}^P \\
\end{bmatrix}
\begin{bmatrix}
1 & \frac{B_1^T B_1}{N} & 0_{K_s} \\
\frac{B_1^T B_1}{N} & \frac{B_2^T B_2}{N} & 0_{K_s \times K_s} \\
0_{K_s \times K_s} & 0_{K_s \times K_s} & 0_{K_s \times K_s}
\end{bmatrix}
\begin{bmatrix}
\gamma_0 \\
\gamma_1 \\
\gamma_{1s}^P
\end{bmatrix}
\]

\[= \gamma_0^2 + \gamma_1^2 \frac{B_1^T B_1}{N} \gamma_1 + 2 \gamma_0 \gamma_1 \frac{B_1^T B_1}{N} \]

\[\rightarrow_P \gamma_0^2 + \gamma_1^2 \Sigma_{\beta g} \gamma_{1s}^P + 2 \gamma_0 \gamma_1 \gamma_{1s}^P \mu_{\beta g}.
\]

Now consider the second term on the right-hand-side of (3.81). Then, Theorem 3.6 and (3.48) imply that

\[
\tilde{\Gamma}_{f_g} X_{f_g} X_{f_g} \tilde{\Gamma}_{f_g} = \begin{bmatrix}
\gamma_0, & \gamma_1, & \gamma_{1s}^P \\
\gamma_0, & \gamma_1, & \gamma_{1s}^P \\
\end{bmatrix}
\begin{bmatrix}
1 & \gamma_1^P & 0_{K_s} \\
\gamma_1^P & \Sigma_{\beta g} + \sigma^2 (F' M T F)^{-1} & 0_{K_s \times K_s} \\
0_{K_s \times K_s} & 0_{K_s \times K_s} & 0_{K_s \times K_s}
\end{bmatrix}
\begin{bmatrix}
\gamma_0 + d_0 \\
\gamma_1^P + d_1 \\
0_{K_s}
\end{bmatrix}
\]

\[= (\gamma_0 + d_0)^2 + (\gamma_1^P + d_1)(\Sigma_{\beta g} + \sigma^2 (F' M T F)^{-1})(\gamma_1^P + d_1) + 2 (\gamma_0 + d_0)(\gamma_1^P + d_1) \mu_{\beta g}.
\]

Hence, substituting in (3.81), we get

\[
\frac{\tilde{\epsilon}_{f_g} \tilde{e}_{f_g}}{N} \rightarrow_P \frac{\sigma^2}{T} + \gamma_0^2 + \gamma_1^2 \Sigma_{\beta g} \gamma_{1s}^P + 2 \gamma_0 \gamma_1 \gamma_{1s}^P \mu_{\beta g}
\]

\[= (\gamma_0 + d_0)^2 - (\gamma_1^P + d_1)^2 (\Sigma_{\beta g} + \sigma^2 (F' M T F)^{-1})(\gamma_1^P + d_1) - 2 (\gamma_0 + d_0)(\gamma_1^P + d_1) \mu_{\beta g}
\]

\[= \frac{\sigma^2}{T} + \gamma_1^P \sigma^2 (F' F)^{-1} D^{-1} (\Sigma_{\beta g} - \mu_{\beta g} \mu_{\beta g}) \gamma_1^P
\]

\[= \frac{\sigma^2}{T} + \gamma_1^P \sigma^2 (F' F)^{-1} D^{-1} \Sigma_{\beta g} \gamma_1^P
\]

(3.82)

Then, by results in (3.48),

\[
\tilde{\sigma}_{f_g}^2 \left( \frac{\tilde{X}_{f_g} \tilde{X}_{f_g}}{N} \right)^{-1} \rightarrow_P \left( \frac{\sigma^2}{T} + \gamma_1^P \sigma^2 (F' F)^{-1} D^{-1} \Sigma_{\beta g} \gamma_1^P \right) \frac{G' G}{\sigma^2}
\]

(3.83)

Consider now the numerator of (3.80). From Theorem 3.6 we have that
\[ \sqrt{N} \hat{\gamma}_{k_y} \rightarrow_d (0_{K_y}, \mathcal{V}_g) \]  

(3.C.84)

where

\[
\mathcal{V}_g = \frac{\tilde{G}' \tilde{G}}{\sigma^2} \left( \sigma^2 d_1^T \Sigma_{\beta_f} \sigma_1 (\tilde{G}' \tilde{G})^{-1} + (Q_f' \otimes P_g') U_c (Q_f \otimes P_g) \right) \frac{\tilde{G}' \tilde{G}}{\sigma^2}
\]

\[
= \frac{d_1^T \Sigma_{\beta_f} \sigma_1}{\sigma^2} \frac{\tilde{G}' \tilde{G}}{\sigma^2} + \frac{\tilde{G}' \tilde{G}}{\sigma^2} (Q_f' \otimes P_g') U_c (Q_f \otimes P_g) \frac{\tilde{G}' \tilde{G}}{\sigma^2}
\]

\[
= \frac{d_1^T \Sigma_{\beta_f} \sigma_1}{\sigma^2} \frac{\tilde{G}' \tilde{G}}{\sigma^2} + \frac{1}{\sigma^2} (Q_f' \otimes G) U_c (Q_f \otimes \tilde{G})
\]

Combining (3.C.83) together with (3.C.85), then

\[
t_{g,k_y} \rightarrow_d \left( 0, \frac{d_1^T \Sigma_{\beta_f} \sigma_1 + \sigma^{-2} W_{[k_x,k_y]}}{\sigma^2 + \gamma_{i_f}^2 \sigma^2 (F'F)^{-1} D^{-1} \Sigma_{\beta_f} \gamma_{i_f}^T} \right)
\]

(3.C.85)

where \( W_{[k_x,k_y]} \) denotes the \((k_x,k_y)\)-th element of the matrix \((Q_f' \otimes \tilde{G}) U_c (Q_f \otimes \tilde{G})\). This concludes the proof of part (i).

(ii) Starting from the definition in (3.33), we can write

\[
R_{CRS_{ fs}}^2 = 1 - \frac{\hat{e}_{fs} \hat{e}_{fs}}{R' \mathcal{M}_N \hat{R}/N}.
\]

(3.C.86)

Using the result in (3.C.82), we have

\[
\frac{\hat{e}_{fs} \hat{e}_{fs}}{N} \rightarrow_p \frac{\sigma^2}{T} + \gamma_{i_f}^2 \sigma^2 (F'F)^{-1} D^{-1} \Sigma_{\beta_f} \gamma_{i_f}^T.
\]

(3.C.87)

Moreover, notice that

\[
\frac{R' \mathcal{M}_N \hat{R}}{N} = \frac{R' \hat{R}}{N} - \frac{R' 1_N 1_N^T \hat{R}}{N}
\]

where, using (3.C.82),

\[
\frac{R' \hat{R}}{N} \rightarrow_p \frac{\sigma^2}{T} + \gamma_{i_f}^2 + \gamma_{i_f}^2 \Sigma_{\beta_f} \gamma_{i_f}^T + 2 \gamma_{i_f} \gamma_{i_f}^T \mu_{\beta_f}.
\]

(3.C.88)

while
\[ \frac{\dot{R}^1_{1N}}{N} = \frac{(\dot{\gamma}_f + \dot{\hat{\gamma}}_{fg} \hat{X}_{fg})^1_{1N}}{N} \]
\[ = \frac{\dot{\gamma}'_{fg} \hat{X}_{fg}}{N} \]
\[ = 0 + [\bar{\gamma}_0, \bar{\gamma}'_{lg}, \bar{\gamma}'_{lg}] \left[ \begin{array}{c} 1 \\ \frac{\hat{B}_{1N}}{B_{1N}} \\ \frac{\hat{P}_{1N}}{P_{1N}} \end{array} \right] \]
\[ = [\bar{\gamma}_0, \bar{\gamma}'_{lg}, \bar{\gamma}'_{lg}] \left[ \begin{array}{c} 1 \\ \frac{\hat{B}_{1N}}{B_{1N}} \\ \frac{\hat{P}_{1N}}{P_{1N}} \end{array} \right] \rightarrow_p \left( (\gamma_0 + d_0), (\gamma'_l + d_1)'_{\mu_{\beta_j}} \right) \left[ \begin{array}{c} 1 \\ \mu_{\beta_j} \\ 0 \end{array} \right] \]
\[ = (\gamma_0 + d_0) + (\gamma'_l + d_1)'_{\mu_{\beta_j}} \]
\[ = \gamma_0 + \gamma'_l \mu_{\beta_j} \quad (3.89) \]

Therefore,
\[ \frac{\dot{R}^1_{1N}}{N} \rightarrow_p \gamma_0 + \gamma'_l \mu_{\beta_j} \]

and hence
\[ \frac{\dot{R}'_{MN\bar{R}}}{N} \rightarrow_p \frac{\sigma^2}{T} + \gamma'_l \Sigma_{\beta_j} \gamma'_l \quad (3.90) \]

Then, substituting (3.87) and (3.91) into (3.86), we get
\[ R_{GRS_{fg}}^2 \rightarrow_p 1 - \frac{\sigma^2}{T} - \frac{\gamma'_l \Sigma_{\beta_j} \gamma'_l}{T} \]
\[ = \gamma'_l \Sigma_{\beta_j} - D^{-1} \Sigma_{\beta_j} \gamma'_l \]

which concludes the proof of part (ii).

(iii) By definition we have
\[ \dot{e}_{fg} = \dot{R} - \dot{X}_{fg} \tilde{\hat{R}}_{fg} \quad (3.92) \]

where
\[ \tilde{\hat{R}}_{fg} = \left[ \begin{array}{c} \hat{\gamma}_0 \\ \hat{\gamma}'_l \\ \hat{\gamma}'_{lg} \end{array} \right] = (\hat{X}_{fg} \hat{X}_{fg})^{-1} \hat{X}_{fg} \tilde{R} \quad (3.93) \]
with $\hat{X}_g = [1_N, \hat{B}_f, \hat{B}_g]$ and where $\hat{B}_f$ and $\hat{B}_g$ are normalized in such a way that Assumption in (3.29) holds true. That is

$$
\hat{B}_f = R' \mathcal{M}_T F (F' \mathcal{M}_T F)^{-1} \\
\quad = B_f + \epsilon' \mathcal{M}_T F (F' \mathcal{M}_T F)^{-1} \\
\quad = B_f + \epsilon' P_f
$$

(3.94)

and

$$
\hat{B}_g = R' \mathcal{M}_T G (G' \mathcal{M}_T G)^{-1} \\
\quad = B_g + \epsilon' \mathcal{M}_T G (G' \mathcal{M}_T G)^{-1} \\
\quad = B_g + \epsilon' P_g \\
\quad = 0_{N \times K_a} + \epsilon' P_g.
$$

(3.95)

Similarly, the OLS residuals from the restricted model in (3.34) have the following form

$$
\hat{e}_f = R - \hat{X}_f \hat{\Gamma}_f
$$

(3.96)

where

$$
\hat{\Gamma}_f = \begin{bmatrix}  \hat{\gamma}_0^* \\ \hat{\gamma}_t^* \end{bmatrix} = (\hat{X}_f \hat{X}_f)^{-1} \hat{X}_f R
$$

(3.97)

with $\hat{X}_f = [1_N, \hat{B}_f]$. Notice that

$$
\hat{e}'_f \hat{e}_f = (R - \hat{X}_f \hat{\Gamma}_f)' (R - \hat{X}_f \hat{\Gamma}_f) \\
\quad = R'R + \hat{\Gamma}_f' \hat{X}_f \hat{X}_f \hat{\Gamma}_f - 2 \hat{\Gamma}_f' \hat{X}_f \hat{\Gamma}_f \\
\quad = R'R - \hat{\Gamma}_f' \hat{X}_f \hat{\Gamma}_f \\
\quad = R'R - \hat{\Gamma}_f' \hat{\Gamma}_f.
$$

(3.98)

and

$$
\hat{e}_f' \hat{e}_f = (R - \hat{X}_f \hat{\Gamma}_f)' (R - \hat{X}_f \hat{\Gamma}_f) \\
\quad = R'R + \hat{\Gamma}_f' \hat{X}_f \hat{\Gamma}_f - 2 \hat{\Gamma}_f' \hat{\Gamma}_f \\
\quad = R'R + \hat{\Gamma}_f' \hat{\Gamma}_f \\
\quad = R'R - \hat{\Gamma}_f' \hat{\Gamma}_f.
$$

(3.99)

since, by the first-order-conditions,

$$
R' \hat{X}_f \hat{\Gamma}_f = (\hat{e}_f + \hat{X}_f \hat{\Gamma}_f)' \hat{X}_f \hat{\Gamma}_f = \hat{e}'_f \hat{X}_f \hat{\Gamma}_f + \hat{\Gamma}_f' \hat{X}_f \hat{\Gamma}_f = 0 + \hat{\Gamma}_f' \hat{\Gamma}_f.
$$

(3.100)
Therefore,

\[
\hat{e}_j^* \hat{e}_j - \hat{e}_{j_f} \hat{e}_{j_g} = \hat{R}' \hat{R} - \hat{\Gamma}_{j_f} X_{j_f} \hat{X}_j \hat{\Gamma}_j - \hat{R} \hat{R} + \hat{\Gamma}_{j_g} \hat{X}_{j_g} \hat{X}_j \hat{\Gamma}_j
\]

\[
= \hat{\Gamma}_{j_g} \hat{X}_{j_g} \hat{X}_j \hat{\Gamma}_j - \hat{\Gamma}_{j_f} \hat{X}_{j_f} \hat{X}_j \hat{\Gamma}_j
\]

\[
= \hat{\Gamma}_{j_g} \hat{X}_{j_g} \hat{X}_j \hat{\Gamma}_j - \frac{1}{N} \hat{\Gamma}_{j_g} X_{j_g} \hat{X}_j \hat{\Gamma}_j - \hat{\Gamma}_{j_f} \hat{X}_{j_f} \hat{X}_j \hat{\Gamma}_j + \frac{1}{N} \hat{\Gamma}_{j_f} X_{j_f} \hat{X}_j \hat{\Gamma}_j
\]

(3.C.100)

since, by construction, \(\frac{1}{N} \hat{\Gamma}_{j_g} X_{j_g} \hat{X}_j \hat{\Gamma}_j = \frac{1}{N} \hat{\Gamma}_{j_f} X_{j_f} \hat{X}_j \hat{\Gamma}_j\). In particular, dividing by \(N\), and defining the following matrices

\[
\left( \begin{array}{c} \hat{B}_f' \hat{B}_f \\ \frac{1}{N} \end{array} \right) = \left( \begin{array}{c} \hat{B}_f' \hat{B}_f \\ \frac{1}{N} \end{array} \right) = \left( \begin{array}{c} \hat{B}_f' \hat{B}_f \\ \frac{1}{N} \end{array} \right),
\]

\[
\left( \begin{array}{c} \hat{B}_g' \hat{B}_g \\ \frac{1}{N} \end{array} \right) = \left( \begin{array}{c} \hat{B}_g' \hat{B}_g \\ \frac{1}{N} \end{array} \right) = \left( \begin{array}{c} \hat{B}_g' \hat{B}_g \\ \frac{1}{N} \end{array} \right),
\]

we have

\[
\frac{\hat{e}_j^* \hat{e}_j}{N} - \frac{\hat{e}_{j_f} \hat{e}_{j_g}}{N} = \hat{\Gamma}_{j_g} \hat{X}_{j_g} \hat{X}_j \hat{\Gamma}_j - \hat{\Gamma}_{j_f} \hat{X}_{j_f} \hat{X}_j \hat{\Gamma}_j + \frac{1}{N} \hat{\Gamma}_{j_g} X_{j_g} \hat{X}_j \hat{\Gamma}_j - \hat{\Gamma}_{j_f} \hat{X}_{j_f} \hat{X}_j \hat{\Gamma}_j + \frac{1}{N} \hat{\Gamma}_{j_f} X_{j_f} \hat{X}_j \hat{\Gamma}_j
\]

(3.C.101)

\[
- \hat{\Gamma}_{j_g} \hat{X}_{j_g} \hat{X}_j \hat{\Gamma}_j + \hat{\Gamma}_{j_f} \hat{X}_{j_f} \hat{X}_j \hat{\Gamma}_j + \frac{1}{N} \hat{\Gamma}_{j_g} X_{j_g} \hat{X}_j \hat{\Gamma}_j - \hat{\Gamma}_{j_f} \hat{X}_{j_f} \hat{X}_j \hat{\Gamma}_j + \frac{1}{N} \hat{\Gamma}_{j_f} X_{j_f} \hat{X}_j \hat{\Gamma}_j
\]

(3.C.102)

\[
\hat{\Gamma}_{j_g} \hat{X}_{j_g} \hat{X}_j \hat{\Gamma}_j - \hat{\Gamma}_{j_f} \hat{X}_{j_f} \hat{X}_j \hat{\Gamma}_j + \frac{1}{N} \hat{\Gamma}_{j_g} X_{j_g} \hat{X}_j \hat{\Gamma}_j - \hat{\Gamma}_{j_f} \hat{X}_{j_f} \hat{X}_j \hat{\Gamma}_j + \frac{1}{N} \hat{\Gamma}_{j_f} X_{j_f} \hat{X}_j \hat{\Gamma}_j
\]

(3.C.103)

\[
\hat{\Gamma}_{j_f} \hat{X}_{j_f} \hat{X}_j \hat{\Gamma}_j - \hat{\Gamma}_{j_g} \hat{X}_{j_g} \hat{X}_j \hat{\Gamma}_j + \frac{1}{N} \hat{\Gamma}_{j_f} X_{j_f} \hat{X}_j \hat{\Gamma}_j - \hat{\Gamma}_{j_g} \hat{X}_{j_g} \hat{X}_j \hat{\Gamma}_j + \frac{1}{N} \hat{\Gamma}_{j_g} X_{j_g} \hat{X}_j \hat{\Gamma}_j
\]

(3.C.104)

Notice that, since we are not interested in estimating \(\gamma_0\), we have
\[
\begin{bmatrix}
\hat{\gamma}_{1,t} \\
\hat{\gamma}_{1,s}
\end{bmatrix}
= \left( \begin{bmatrix}
\hat{B}'_f \\
\hat{B}'_g
\end{bmatrix} M_N \begin{bmatrix}
\hat{B}_f \\
\hat{B}_g
\end{bmatrix} \right)^{-1}
\begin{bmatrix}
\hat{B}'_f \\
\hat{B}'_g
\end{bmatrix} M_N \hat{R}
= \begin{bmatrix}
\hat{\gamma}'_{1,t} \\
\hat{\gamma}'_{1,s}
\end{bmatrix} + \begin{bmatrix}
\hat{B}'_f \hat{B}_f' \\
\hat{B}'_g \hat{B}_g'
\end{bmatrix}^{-1}
\begin{bmatrix}
\hat{B}'_f \\
\hat{B}'_g
\end{bmatrix} \left( \bar{\tau} - (\bar{X}_{fg} - X_{fg}) \Gamma_{fg}' \right)
= \begin{bmatrix}
\hat{\gamma}'_{1,t} \\
\hat{\gamma}'_{1,s}
\end{bmatrix} + \begin{bmatrix}
\hat{B}'_f \hat{B}_f' \\
\hat{B}'_g \hat{B}_g'
\end{bmatrix}^{-1}
\begin{bmatrix}
\hat{B}'_f \\
\hat{B}'_g
\end{bmatrix} \eta_{fg}
\end{align}
\]

(3.C.106)

where \( M_N = I_N - \frac{1}{N} \bar{X}_{fg} \), \( \eta_{fg} = \left( \bar{\tau} - (\bar{X}_{fg} - X_{fg}) \Gamma_{fg}' \right) \), \( X_{fg} = \left[ 1_{N}, B_f, 0_{N \times K_d} \right] \), \( \bar{X}_{fg} = \left[ 1_{N}, \hat{B}_f, \hat{B}_g \right] \), \( \Gamma_{fg}' = \left[ \gamma_0, \gamma_{1,t}', \gamma_{1,s}' \right]' \). Moreover, using the expression of the inverse of a partitioned matrix, then

\[
\begin{bmatrix}
\hat{B}'_f \hat{B}_f' \\
\hat{B}'_g \hat{B}_g'
\end{bmatrix}^{-1} = \begin{bmatrix}
(\hat{B}'_f \hat{B}_f)^{-1} + (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_f \hat{B}_g) \hat{D}_{fg}^{-1}(\hat{B}'_f \hat{B}_f)^{-1} & -(\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_f \hat{B}_g) \hat{D}_{fg}^{-1} \\
\hat{B}'_f \hat{B}_g & \hat{D}_{fg}^{-1}(\hat{B}'_f \hat{B}_f)^{-1} \hat{B}'_f \hat{B}_g \hat{D}_{fg}^{-1}
\end{bmatrix}
\]

(3.C.107)

with \( \hat{D}_{fg} = (\hat{B}'_f \hat{B}_f) - (\hat{B}'_f \hat{B}_f)(\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_f \hat{B}_g) \). Then, substituting (3.C.107) in (3.C.106), we have

\[
\hat{\gamma}_{1,t} = \hat{\gamma}'_{1,t} + (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_f \hat{B}_f) \hat{D}_{fg}^{-1}(\hat{B}'_f \hat{B}_f)(\hat{B}'_f \hat{B}_f)^{-1} \hat{B}'_f (\bar{\tau} - (\bar{X}_{fg} - X_{fg}) \Gamma_{fg}')
= \hat{\gamma}'_{1,t} + \Delta_{1,t}
\]

(3.C.108)

where we define

\[
\Delta_{1,t} = (\hat{B}'_f \hat{B}_f)^{-1} + (\hat{B}'_f \hat{B}_f) \hat{D}_{fg}^{-1}(\hat{B}'_f \hat{B}_f)(\hat{B}'_f \hat{B}_f)^{-1} \hat{B}'_f \eta_{fg} - (\hat{B}'_f \hat{B}_f)^{-1}(\hat{B}'_f \hat{B}_g) \hat{D}_{fg}^{-1} \hat{B}_g \eta_{fg}
\]

(3.C.109)

Using the same procedure for \( \hat{\gamma}_{1,s} \), and define \( \eta^*_t = \left( \bar{\tau} - (\bar{X}_f - X_f) \Gamma_f' \right) \) we have

\[
\hat{\gamma}_{1,s} = \hat{\gamma}'_{1,s} + (\hat{B}'_f \hat{B}_f)^{-1} \hat{B}'_f \eta^*_t
= \hat{\gamma}'_{1,s} + \Delta_{1,s}
\]

(3.C.110)

where

\[
\Delta_{1,s} = (\hat{B}'_f \hat{B}_f)^{-1} \hat{B}_f \eta^*_t
\]

(3.C.111)

Moreover, notice that, using (3.C.109) and (3.C.111), the following holds

\[
\hat{d}_{1,t} = \hat{d}_{1,t}^* + \hat{\Delta}_{fg}
\]

(3.C.112)
where

\[
\hat{\Delta}_{fs} = \left(\tilde{B}'_f \tilde{B}_f\right)^{-1}\tilde{B}'_f (B_s - \tilde{B}_s) \gamma_s^f + \left(\tilde{B}'_f \tilde{B}_f\right)^{-1}(\tilde{B}'_f \tilde{B}_f)\left(\tilde{B}'_f \tilde{B}_f\right)^{-1}\tilde{B}'_f \eta_{fs} + \left(\tilde{B}'_f \tilde{B}_f\right)^{-1}(\tilde{B}'_f \tilde{B}_f)\left(\tilde{B}'_f \tilde{B}_f\right)^{-1}\tilde{B}'_f \eta_{fs}.
\]

Therefore, using (3.C.112) and rearranging, (3.C.105) can be written as

\[
\frac{\tilde{c}_f^* \tilde{c}_f}{N} - \frac{\tilde{c}_f^* \tilde{c}_f}{N} = \hat{\Delta}_{fs} \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) \hat{\Delta}_{fs} + 2 \hat{\Delta}_{fs} \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) (\gamma_s^f + \tilde{d}_s^f) + \gamma_s^f \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) \tilde{\gamma}_s + 2 (\tilde{d}_s^f + \hat{\Delta}_{fs} + \gamma_s^f) \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) \tilde{\gamma}_s.
\]

Moreover, using Lemmas 3.11 and 3.12, together with Theorem 3.6, we can also write

\[
\frac{\tilde{c}_f^* \tilde{c}_f}{N} - \frac{\tilde{c}_f^* \tilde{c}_f}{N} = 2 \hat{\Delta}_{fs} \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) (\gamma_s^f + \tilde{d}_s^f) + \gamma_s^f \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) \tilde{\gamma}_s + 2 (\tilde{d}_s^f + \gamma_s^f) \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) \tilde{\gamma}_s + O_p \left(\frac{1}{N^2}\right)
\]

which implies that

\[
\tilde{c}_f^* \tilde{c}_f - \tilde{c}_f^* \tilde{c}_f = 2 N \hat{\Delta}_{fs} \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) (\gamma_s^f + \tilde{d}_s^f) + \sqrt{N} \gamma_s^f \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) \sqrt{N} \tilde{\gamma}_s + 2 (\tilde{d}_s^f + \gamma_s^f) \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) \sqrt{N} \tilde{\gamma}_s + O_p \left(\frac{1}{N}\right)
\]

setting

\[
E_1 = 2 N \hat{\Delta}_{fs} \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) (\gamma_s^f + \tilde{d}_s^f), \quad E_2 = \sqrt{N} \gamma_s^f \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) \sqrt{N} \tilde{\gamma}_s,
\]

and

\[
E_3 = 2 (\tilde{d}_s^f + \gamma_s^f) \left(\frac{\tilde{B}'_f \tilde{B}_f}{N}\right) \sqrt{N} \tilde{\gamma}_s.
\]
Let us start with $E_2$. Using (3.3.75), we have

$$\sqrt{N} \hat{\gamma}_{i_j} = \hat{G}' \hat{G} \left( P_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma_i^2 I_T) Q_f - P_g' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i (\beta_{f_i} - \mu_{\beta_f}) d_1 \right),$$

(3.3.113)

and

$$\left( \frac{B_{g}' \tilde{B}_{g}}{N} - \frac{B_{g}' 1_N 1_N' \tilde{B}_{g}}{N} \right) \rightarrow_p \sigma^2 (\hat{G}' \hat{G})^{-1}.$$ 

It implies that

$$E_2 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}'(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \left( Q_f Q_f' \otimes P_g \frac{\hat{G}'}{\sigma^2} \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i'(\beta_{f_i} - \mu_{\beta_f}) \frac{\hat{G}'}{\sigma^2} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i'(\beta_{f_i} - \mu_{\beta_f}) \epsilon_i - 2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}'(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \left( Q_f \otimes P_g \frac{\hat{G}'}{\sigma^2} \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i'(\beta_{f_i} - \mu_{\beta_f}) \epsilon_i$$

Consider now $E_3$. First notice that, by Lemma 3.12,

$$\gamma_{i_j}^{\epsilon_{g}} + d_{i_j}^{\epsilon_{g}} \rightarrow_p \gamma_{i_j}^{\epsilon_{g}} + d_{i_j}$$

(3.3.114)

Moreover

$$\frac{\tilde{B}_g' \tilde{B}_g}{\sqrt{N}} = B_{f} \epsilon_{g} P_g + P_f' \sqrt{N} \left( \frac{\epsilon \epsilon'}{N} - \sigma^2 I_T \right) P_g - \frac{B_{f} 1_N \epsilon_{g}'}{\sqrt{N}} P_g + P_f' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{f_i} - \mu_{\beta_f}) \epsilon_i P_g + O_p \left( \frac{1}{\sqrt{N}} \right)$$

(3.3.115)

Hence, using (3.3.113), we have
\[ E_3 = 2(\gamma_{1f}^P + d_1)' \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{fi} - \mu_{\beta_f}) \epsilon_i' P_g + P_f' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i' - \sigma_i^2 I_T) P_g \right) \times \]
\[ \times \frac{\tilde{G}' \tilde{G}}{\sigma^2} \left( P_g' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i' - \sigma_i^2 I_T) Q_f - P_g' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i(\beta_{fi} - \mu_{\beta_f})' d_1 \right) \]
\[ = 2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i' (\beta_{fi} - \mu_{\beta_f})' (\gamma_{1f}^P + d_1) \left( Q_f' \otimes P_g \tilde{G}' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i' - \sigma_i^2 I_T) \right) \]
\[ - 2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i' (\beta_{fi} - \mu_{\beta_f})' (\gamma_{1f}^P + d_1) P_g \tilde{G}' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i(\beta_{fi} - \mu_{\beta_f}) \epsilon_i \]
\[ + 2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i' - \sigma_i^2 I_T) \left( P_f (\gamma_{1f}^P + d_1) Q_f' \otimes P_g \tilde{G}' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i' - \sigma_i^2 I_T) \right) \]
\[ - 2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i' - \sigma_i^2 I_T) \left( P_g \tilde{G}' \oslash P_f (\gamma_{1f}^P + d_1) \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i(\beta_{fi} - \mu_{\beta_f}) \epsilon_i \]

Finally, consider \( N\Delta_{fg} \). Using the result of Lemma 3.11

\[ N\Delta_{fg} = -\left( \frac{\tilde{B}_f' \tilde{B}_g}{N} \right)^{-1} \left( \frac{\tilde{B}_f' \tilde{B}_g}{\sqrt{N}} \left( \frac{1}{T} - P_f \gamma_{1f}^P \right) \right)^{-1} \left( \frac{\tilde{B}_g}{\sqrt{N}} \right)^{-1} \left( \frac{\tilde{B}_f}{\sqrt{N}} \right) \left( \frac{\tilde{B}_g}{\sqrt{N}} \right)^{-1} \left( \frac{\tilde{B}_f' \tilde{B}_f}{N} \right)^{-1} \sigma^2 (\tilde{F}' \tilde{F})^{-1} \gamma_{1f}^P \]

where, using (3.C.115) and (3.29),

\[ \Delta_1 = -\left( \frac{\tilde{B}_f' \tilde{B}_g}{N} \right)^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{fi} - \mu_{\beta_f}) \epsilon_i' P_g + P_f' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i' - \sigma_i^2 I_T) P_g \right) \left( \frac{\tilde{B}_g}{N} \right)^{-1} \times \]
\[ \times P_g' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i' - \sigma_i^2 I_T) \left( \frac{1}{T} - P_f \gamma_{1f}^P \right) \]

and

\[ \Delta_2 = -\sigma^2 \left( \frac{\tilde{B}_f' \tilde{B}_f}{N} \right)^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{fi} - \mu_{\beta_f}) \epsilon_i' P_g + P_f' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i' - \sigma_i^2 I_T) P_g \right) \left( \frac{\tilde{B}_g}{N} \right)^{-1} \times \]
\[ \times \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{fi} - \mu_{\beta_f}) \epsilon_i' P_g + P_f' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i' - \sigma_i^2 I_T) P_g \right) \left( \frac{\tilde{B}_f' \tilde{B}_f}{N} \right)^{-1} \left( \tilde{F}' \tilde{F} \right)^{-1} \gamma_{1f}^P \]
Therefore,

\[
E_1 = -2 \left( \gamma^p_{\beta_f} + \hat{d}_f \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{fi} - \mu_{\beta_f}) \epsilon_i P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) \left( \frac{1}{T} - P_f \gamma^p_{\beta_f} \right)
\]

(3.116)

\[
- 2 \left( \gamma^p_{\beta_f} + \hat{d}_f \right) P'_f \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) \left( \frac{1}{T} - P_f \gamma^p_{\beta_f} \right)
\]

(3.117)

\[
- 2 \sigma^2 (\gamma^p_{\beta_f} + \hat{d}_f) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{fi} - \mu_{\beta_f}) \epsilon'_i P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i (\beta_{fi} - \mu_{\beta_f})' \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} \left( \hat{F}' \hat{F} \right)^{-1} \gamma^p_{\beta_f}
\]

(3.118)

\[
- 2 \sigma^2 (\gamma^p_{\beta_f} + \hat{d}_f) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{fi} - \mu_{\beta_f}) \epsilon'_i P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_f \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} \left( \hat{F}' \hat{F} \right)^{-1} \gamma^p_{\beta_f}
\]

(3.119)

\[
- 2 \sigma^2 (\gamma^p_{\beta_f} + \hat{d}_f) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_f \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} \left( \hat{F}' \hat{F} \right)^{-1} \gamma^p_{\beta_f}
\]

(3.120)

\[
- 2 \sigma^2 (\gamma^p_{\beta_f} + \hat{d}_f) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_f \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} \left( \hat{F}' \hat{F} \right)^{-1} \gamma^p_{\beta_f}
\]

(3.121)

Then, collecting (3.116) with (3.119), and (3.117) with (3.121), and defining the T-vector

\[
\hat{Q}_f = \left( \frac{1}{T} - P_f \gamma^p_{\beta_f} + \sigma^2 P_f \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} \left( \hat{F}' \hat{F} \right)^{-1} \gamma^p_{\beta_f} \right)
\]

(3.122)

we have

\[
E_1 = -2 \left( \gamma^p_{\beta_f} + \hat{d}_f \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{fi} - \mu_{\beta_f}) \epsilon_i P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) \hat{Q}_f
\]

(3.123)

\[
- 2 \left( \gamma^p_{\beta_f} + \hat{d}_f \right) P'_f \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) \hat{Q}_f
\]

(3.124)

\[
- 2 \sigma^2 (\gamma^p_{\beta_f} + \hat{d}_f) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{fi} - \mu_{\beta_f}) \epsilon'_i P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i (\beta_{fi} - \mu_{\beta_f})' \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} \left( \hat{F}' \hat{F} \right)^{-1} \gamma^p_{\beta_f}
\]

(3.125)

\[
- 2 \sigma^2 (\gamma^p_{\beta_f} + \hat{d}_f) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\beta_{fi} - \mu_{\beta_f}) \epsilon'_i P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_f \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} \left( \hat{F}' \hat{F} \right)^{-1} \gamma^p_{\beta_f}
\]

(3.126)

\[
- 2 \sigma^2 (\gamma^p_{\beta_f} + \hat{d}_f) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_f \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} \left( \hat{F}' \hat{F} \right)^{-1} \gamma^p_{\beta_f}
\]

(3.127)

\[
- 2 \sigma^2 (\gamma^p_{\beta_f} + \hat{d}_f) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_g \left( \frac{\hat{D}_{fg}}{N} \right)^{-1} P'_g \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon'_i - \sigma^2 I_T) P_f \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} \left( \hat{F}' \hat{F} \right)^{-1} \gamma^p_{\beta_f}
\]

(3.128)

Now, using the fact that \( \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} \to_p D^{-1} \), then \( \hat{Q}_f \to_p Q_f \) and also \( -\sigma^2 \left( \frac{\hat{E}_f \hat{E}_f}{N} \right)^{-1} P_f P_f \gamma^p_{\beta_f} \to_p d_1 \).
Moreover, notice that

\[
\frac{\hat{D}_{f_2}}{N} = \frac{\hat{B}_{f_2}}{N} - \frac{\hat{B}_{f_2}^T (\hat{B}_{f_2} \hat{B}_{f_2})^{-1} \hat{B}_{f_2}}{\sqrt{N}} = \frac{\hat{B}_{f_2}}{N} + O_{f} \left( \frac{1}{N} \right)
\to_p \sigma^2 (\hat{G}' \hat{G})^{-1}
\]

Hence, using the properties of the vec operator, we can write

\[
E_1 = -2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i (\beta_{f_i} - \mu_{\beta_f})' (\gamma_{1,f_i}^P + d_1) \left(Q_f \otimes P_g \frac{\bar{G}'}{\sigma^2} \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon_i' - \sigma^2 I_{T_T})
\]

\[
- 2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}'(\epsilon_i \epsilon_i' - \sigma^2 I_{T_T}) \left(P_f (\gamma_{1,f_i}^P + d_1) Q_f \otimes P_g \frac{\bar{G}'}{\sigma^2} \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon_i' - \sigma^2 I_{T_T})
\]

\[
+ 2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i (\beta_{f_i} - \mu_{\beta_f})' (\gamma_{1,f_i}^P + d_1) P_g \frac{\bar{G}'}{\sigma^2} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i (\beta_{f_i} - \mu_{\beta_f}) \epsilon_i
\]

\[
+ 2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}'(\epsilon_i \epsilon_i' - \sigma^2 I_{T_T}) \left(P_g \frac{\bar{G}'}{\sigma^2} \otimes P_f (\gamma_{1,f_i}^P + d_1) \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i (\beta_{f_i} - \mu_{\beta_f}) \epsilon_i
\]

It implies that \( E_1 + E_3 = 0 \) and hence

\[
\hat{e}_f^T \hat{e}_f^* - \hat{e}_f' \hat{e}_f = E_2
\]

\[
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}'(\epsilon_i \epsilon_i' - \sigma^2 I_{T_T}) \left(Q_f Q_f' \otimes P_g \frac{\bar{G}'}{\sigma^2} \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon_i' - \sigma^2 I_{T_T})
\]

\[
+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i (\beta_{f_i} - \mu_{\beta_f}) \epsilon_i P_g \frac{\bar{G}'}{\sigma^2} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i (\beta_{f_i} - \mu_{\beta_f}) \epsilon_i
\]

\[
- 2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}'(\epsilon_i \epsilon_i' - \sigma^2 I_{T_T}) \left(Q_f \otimes P_g \frac{\bar{G}'}{\sigma^2} \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i (\beta_{f_i} - \mu_{\beta_f}) \epsilon_i
\]

Let \( Z_1 \equiv N(0_{T_T}, U_{T_T}) \) and \( Z_2 \equiv N \left( 0_T, \sigma^2 d_1^T \Sigma_{\beta_f} d_1 I_T \right) \) be two normally distributed vectors of dimension \( T^2 \times 1 \) and \( T \times 1 \), respectively, such that

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon_i' - \sigma^2 I_{T_T}) \xrightarrow{d} Z_1
\]

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_i (\beta_{f_i} - \mu_{\beta_f}) \epsilon_i \xrightarrow{d} Z_2
\]

Then,

\[
\hat{e}_f^T \hat{e}_f^* - \hat{e}_f' \hat{e}_f \xrightarrow{d} \begin{bmatrix} Z_1 \mid Z_2 \end{bmatrix} \mathcal{W}_{f_2} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}
\]

(3.C.123)
where

\[
W_{fg} = \frac{1}{\sigma^2} \begin{bmatrix}
Q_f Q_f' \otimes P_g \tilde{G}' & -Q_f \otimes P_g \tilde{G}' \\
-Q_f' \otimes P_g \tilde{G}' & P_g \tilde{G}'
\end{bmatrix}
\]

(3.C.124)

Consider now the denominator of the F-test in (3.35). Using (3.C.82), we have

\[
\frac{\hat{c}_{fg}^2 \hat{c}_{fg}'}{N} = \frac{\sigma^2}{T} + \gamma_{fg}' \sigma^2 (\tilde{F}' \tilde{F})^{-1} D_1 \Sigma_{\beta_{fg}'} \gamma_{fg}
= \frac{\sigma^2}{T} - \hat{d}' \Sigma_{\beta_{fg}'} \gamma_{fg}
\]

Therefore,

\[
F_{GSR_{fg}} \Rightarrow \begin{bmatrix} Z_1', & Z_2' \end{bmatrix} \begin{bmatrix} W_{fg}/K_g \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}
\]

\[
\frac{\hat{c}_{fg}^2 \hat{c}_{fg}'}{N} \begin{bmatrix} Z_1', & Z_2' \end{bmatrix} \begin{bmatrix} W_{fg}/K_g \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}
\]

which concludes the proof of part (iii).

**Proof of Proposition 3.2.** For the proof of this theorem we can use most of the results of Theorem 3.6(i).

In particular, \(\frac{\hat{\beta}_{fg}'}{N} \to_p \mu_{\beta_{fg}'}\), \(\frac{\hat{\beta}_{fg}'}{N} \to_p 0_{K_g}\), \(\frac{\hat{\beta}_{fg}'}{N} \to_p \sigma^2 D^{-1}\). Now,

\[
\frac{\hat{B}_{fg}'}{N} \to_p \sigma^2 (\tilde{F}' \tilde{F})^{-1} \tilde{F}' \tilde{G}D^{-1} \equiv Q_{fg},
\]

Using the same arguments of Theorem 3.6 and defining \(\Sigma_{\beta_{fg}} = \text{plim} \frac{1}{N} \sum_{i=1}^N \beta_i \beta_i'\), we have

\[
\frac{\hat{B}_{fg}'}{N} \to_p \Sigma_{\beta_{fg}} + \sigma^2 \left( (\tilde{F}' \tilde{F})^{-1} + (\tilde{F}' \tilde{F})^{-1} (\tilde{F}' \tilde{G}) (\tilde{G}' \tilde{G})^{-1} (\tilde{G}' \tilde{F}) (\tilde{F}' \tilde{F})^{-1} \right).
\]

(3.C.125)

This implies that

\[
\frac{\hat{X}_{fg}' \hat{X}_{fg}}{N} \to_p \begin{bmatrix} 1 & \mu_{\beta_{fg}} \\ \mu_{\beta_{fg}} & 0_{K_g} \end{bmatrix} + \sigma^2 \begin{bmatrix} (\tilde{F}' \tilde{F})^{-1} & 0_{K_g} \\ 0_{K_g} & Q_{fg} \end{bmatrix}.
\]

Consider now \(\frac{\hat{X}_{fg}' R}{N}\). Using (3.27), we have

\[
\frac{\hat{X}_{fg}' \hat{R}}{N} = \left( \frac{\hat{X}_{fg}' \hat{X}_{fg}}{N} \right) \Gamma + \frac{\hat{X}_{fg}'}{N} \left( \hat{\xi} - (\hat{X}_{fg} - X) \Gamma \right)
\]

where

\[
\hat{X}_{fg}' \hat{\xi} = O_p \left( \frac{1}{\sqrt{N}} \right).
\]
and where

\[-\frac{\dot{X}_{f_g}'}{N}(\dot{X}_{f_g} - X)\Gamma_P = \frac{X_{f_g}'}{N}(X - \dot{X}_{f_g})\Gamma_P + \frac{(X - \dot{X}_{f_g})'(\dot{X}_{f_g} - X)}{N}\Gamma_P\]

\[-\gamma_{\rho} = \begin{bmatrix} 0 & 0'_{K_s} \\ 0_{K_s} & \sigma^2 \left((\ddot{F}'\ddot{F})^{-1} + (\dot{F}'\dot{F})(\dddot{G}'\dddot{G})^{-1}(\ddot{G}'\ddot{G})(\dddot{F}'\dddot{F})^{-1}\right) \sigma^2 D^{-1} \end{bmatrix} \Gamma_P.\]

This implies that

\[\Gamma_P \rightarrow_{\rho} \Gamma_P = \begin{bmatrix} 1 & \mu_{\rho}' \\ \mu_{\rho} & \Sigma_{\rho} + \sigma^2 \left((\ddot{F}'\ddot{F})^{-1} + (\dot{F}'\dot{F})(\dddot{G}'\dddot{G})^{-1}(\ddot{G}'\ddot{G})(\dddot{F}'\dddot{F})^{-1}\right) \sigma^2 D^{-1} \end{bmatrix}^{-1} \times \]

\[\begin{bmatrix} 0 & 0'_{K_s} \\ 0_{K_s} & \sigma^2 \left((\ddot{F}'\ddot{F})^{-1} + (\dot{F}'\dot{F})(\dddot{G}'\dddot{G})^{-1}(\ddot{G}'\ddot{G})(\dddot{F}'\dddot{F})^{-1}\right) \sigma^2 D^{-1} \end{bmatrix} \Gamma_P.\]

Notice that

\[\begin{bmatrix} 1 & \mu_{\rho}' \\ \mu_{\rho} & \Sigma_{\rho} + \sigma^2 \left((\ddot{F}'\ddot{F})^{-1} + (\dot{F}'\dot{F})(\dddot{G}'\dddot{G})^{-1}(\ddot{G}'\ddot{G})(\dddot{F}'\dddot{F})^{-1}\right) \sigma^2 D^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} 1 + (\mu_{\rho}', 0'_{K_s})P^{-1}(\mu_{\rho}', 0'_{K_s})' & -(\mu_{\rho}', 0'_{K_s})P^{-1} \\ -(\mu_{\rho}', 0'_{K_s})P^{-1} & P^{-1} \end{bmatrix} \]

setting

\[P = \begin{bmatrix} \Sigma_{\rho} + \sigma^2 \left((\ddot{F}'\ddot{F})^{-1} + (\dot{F}'\dot{F})(\dddot{G}'\dddot{G})^{-1}(\ddot{G}'\ddot{G})(\dddot{F}'\dddot{F})^{-1}\right) \sigma^2 D^{-1} \end{bmatrix},\]

\[C = \begin{bmatrix} \sigma^2 \left((\ddot{F}'\ddot{F})^{-1} + (\dot{F}'\dot{F})(\dddot{G}'\dddot{G})^{-1}(\ddot{G}'\ddot{G})(\dddot{F}'\dddot{F})^{-1}\right) \sigma^2 D^{-1} \end{bmatrix}.\]

This yields

\[\Gamma_P \rightarrow_{\rho} \Gamma_P = \begin{bmatrix} 1 + (\mu_{\rho}', 0'_{K_s})P^{-1}(\mu_{\rho}', 0'_{K_s})' & -(\mu_{\rho}', 0'_{K_s})P^{-1} \\ -P^{-1}(\mu_{\rho}', 0'_{K_s})' & P^{-1} \end{bmatrix} \begin{bmatrix} 0 & 0'_{K_{f-g}} + C \\ 0_{K_{f-g}} & C \end{bmatrix} \Gamma_P.\]

Now

\[P^{-1} = \begin{bmatrix} E^{-1} & -E^{-1}Q_{f_g}P \sigma^2 \\ -\frac{P}{\sigma^2}Q_{f_g}E^{-1} & \frac{P}{\sigma^2} + \frac{P}{\sigma^2}Q_{f_g}E^{-1} \end{bmatrix} \]

\[= \begin{bmatrix} -\frac{P}{\sigma^2}Q_{f_g}E^{-1} & \frac{P}{\sigma^2}Q_{f_g}E^{-1} \\ -\frac{P}{\sigma^2}Q_{f_g}E^{-1} & \frac{P}{\sigma^2} + \frac{P}{\sigma^2}Q_{f_g}E^{-1} \end{bmatrix}.\]
setting
\[ E = \Sigma_{\beta} + \sigma^2 \left( (\hat{F}'\hat{F})^{-1} + (\hat{F}'\hat{G})(\hat{G}'\hat{G})^{-1}(\hat{G}'\hat{F})(\hat{F}'\hat{F})^{-1} \right) - Q_{f,g} \frac{D}{\sigma^2} Q'_{f,g}, \]

implying
\[ P^{-1}G = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_{K_f} - E^{-1}\Sigma_{\beta} & 0_{K_f \times K_g} \\ \frac{D}{\sigma^2} Q_{f,g} E^{-1}\Sigma_{\beta} & I_{K_g} \end{bmatrix}, \]

where
\begin{align*}
A_{11} & = E^{-1}\sigma^2 \left( (\hat{F}'\hat{F})^{-1} + (\hat{F}'\hat{G})(\hat{G}'\hat{G})^{-1}(\hat{G}'\hat{F})(\hat{F}'\hat{F})^{-1} \right) - Q_{f,g} \frac{D}{\sigma^2} Q'_{f,g} = I_{K_f} - E^{-1}\Sigma_{\beta}, \\
A_{12} & = E^{-1}Q_{f,g} - E^{-1}Q_{f,g} = 0_{K_f \times K_g}, \\
A_{21} & = -\frac{D}{\sigma^2} Q_{f,g} E^{-1}\sigma^2 \left( (\hat{F}'\hat{F})^{-1} + (\hat{F}'\hat{G})(\hat{G}'\hat{G})^{-1}(\hat{G}'\hat{F})(\hat{F}'\hat{F})^{-1} \right) + \left( \frac{D}{\sigma^2} + \frac{D}{\sigma^2} Q_{f,g} E^{-1}Q_{f,g} \frac{D}{\sigma^2} \right) Q_{f,g} \\
& = \frac{D}{\sigma^2} Q_{f,g} E^{-1}\Sigma_{\beta}, \\
A_{22} & = -\frac{D}{\sigma^2} Q_{f,g} E^{-1}Q_{f,g} + \left( \frac{D}{\sigma^2} + \frac{D}{\sigma^2} Q_{f,g} E^{-1}Q_{f,g} \frac{D}{\sigma^2} \right) \sigma^2 D^{-1} = I_{K_g}.
\end{align*}

Thus,
\[ \hat{\Gamma}_{f,g}^P \rightarrow_p \Gamma^P = \begin{bmatrix} 0 & \mu_{\beta}' (I_{K_f} - E^{-1}\Sigma_{\beta}) & 0_{\gamma_{f,g}'} \\ 0_{K_f} & (I_{K_f} - E^{-1}\Sigma_{\beta}) & 0_{K_f \times K_g} \\ 0_{K_g} & \frac{D}{\sigma^2} Q'_{f,g} E^{-1}\Sigma_{\beta} & I_{K_g} \end{bmatrix} \Gamma^P = \begin{bmatrix} \gamma_0 - \mu_{\beta}' (I_{K_f} - E^{-1}\Sigma_{\beta}) \gamma_{f,g}'^P \\ E^{-1}\Sigma_{\beta} \gamma_{f,g}'^P \\ -\frac{D}{\sigma^2} Q'_{f,g} E^{-1}\Sigma_{\beta} \gamma_{f,g}'^P \end{bmatrix}, \]

which concludes the proof of the theorem. ■

**Proof of Theorem 3.8.** Starting from the definition in (3.44), the feasible OLS CSR estimator can be written as
\[ \hat{\Gamma}_{f,g} = \left( \frac{\hat{X}_{f,g}' \hat{X}_{f,g}}{N} \right)^{-1} \frac{\hat{X}_{f,g}' \hat{R}}{N}. \] (3.C.126)

Now, using the definition in (3.40) and under (3.29), we have
\[ \frac{\hat{X}_{f,g}' \hat{X}_{f,g}}{N} = \begin{bmatrix} 1 & \hat{\beta}_{[1]}' & 1' \hat{\beta}_{[2]}' \\ \hat{\beta}_{[1]}' & \hat{\beta}_{[1]}' \times \hat{\beta}_{[1]}' & \hat{\beta}_{[2]}' \times \hat{\beta}_{[2]}' \\ \hat{\beta}_{[1]}' \times \hat{\beta}_{[1]}' & \hat{\beta}_{[2]}' \times \hat{\beta}_{[2]}' & \hat{\beta}_{[3]}' \times \hat{\beta}_{[3]}' \end{bmatrix} \rightarrow_p \begin{bmatrix} 1 & \mu_{\beta}' \Sigma_{\beta} + \sigma^2 (\hat{F}'[1] \hat{F}'[1])^{-1} \\ \mu_{\beta}' \Sigma_{\beta} + \sigma^2 (\hat{F}'[1] \hat{F}'[1])^{-1} & 0_{K_f \times K_g} \\ 0_{K_g \times K_f} & 0_{K_g \times K_f} \end{bmatrix} = \Sigma_{K_f \times K_g} + \Lambda_{f,g}' \] (3.C.127)

It implies that
\[
\left( \frac{\hat{X}^{\prime}_{f,3}}{N} \right)^{-1} \rightarrow_p \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right]^{-1} \begin{array}{c}
0'_{K_x} \\
0_{K_x \times K_f}
\end{array} \left( \frac{\hat{G}_0 \hat{G}}{\sigma^2} \right)
\]
\[
\begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}
\begin{bmatrix}
0'_{K_x} \\
0_{K_x \times K_f}
\end{bmatrix}
\left( \frac{\hat{G}_0 \hat{G}}{\sigma^2} \right)
\] (3.C.129)

where
\[
A_{11} = 1 + \frac{\mu_{\beta_f}^{[1]}}{\mu_{\beta_f}^{[1]}} D^{-1} \mu_{\beta_f}^{[1]}
\] (3.C.130)
\[
A_{22} = D^{-1}
\] (3.C.132)
\[
D = \Sigma_{\beta_f}^{[1]} - \mu_{\beta_f}^{[1]} \mu_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\] (3.C.133)

Moreover, using the definition in (3.40) and rewriting (3.43) as \( \hat{R} = X_f \Gamma_f^p + \tau \), then

\[
\frac{\hat{X}^{\prime}_{f,3}}{N} \hat{R} = \frac{\hat{X}^{\prime}_{f,3}}{N} (X_f \Gamma_f^p + \tau)
\]
\[
= \frac{\hat{X}^{\prime}_{f,3}}{N} X_f \left( \frac{1}{\mu_{\beta_f}^{[1]}} + \frac{\hat{F}^{[1]} \hat{F}^{[1]}}{\sigma^2} \right)
\]
\[
= \frac{\hat{X}^{\prime}_{f,3}}{N} X_f \left( \frac{1}{\mu_{\beta_f}^{[1]}} + \frac{\hat{F}^{[1]} \hat{F}^{[1]}}{\sigma^2} \right)
\]
\[
= \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right] \Gamma_f^p + \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right]
\]
\[
= \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right] \Gamma_f^p + \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right]
\]
\[
= \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right] \Gamma_f^p + \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right]
\]
\[
= \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right] \Gamma_f^p + \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right]
\]
\[
= \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right] \Gamma_f^p + \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right]
\]
\[
\rightarrow_p \left[ \begin{array}{c}
\frac{1}{\mu_{\beta_f}^{[1]}} \\
\Sigma_{\beta_f}^{[1]} + \sigma^2 (\hat{F}^{[1]} \hat{F}^{[1]})^{-1}
\end{array} \right] \Gamma_f^p .
\] (3.C.134)

Then, using (3.C.129) and (3.C.134), we get
\[ \hat{\Gamma}_{f,g} \rightarrow P \left[ \begin{array}{ccc} A_{11} & A_{12} & 0_{K_s} \\ A_{21} & A_{22} & 0_{K_s \times K_f} \\ 0_{K_s} & 0_{K_s \times K_f} & 0_{K_s \times K_f} \end{array} \right] \left[ \begin{array}{ccc} 1 & \mu_\beta^{[1]} & \mu_\beta^{[2]} \\ \mu_\beta^{[1]} & \Sigma_\beta^{[1]} & \Sigma_\beta^{[1,2]} \\ \mu_\beta^{[2]} & \Sigma_\beta^{[1,2]} & \Sigma_\beta^{[2]} \end{array} \right] \Gamma_f^P \\
= \left[ \begin{array}{ccc} A_{11} + A_{12} \mu_\beta^{[1]} & A_{11} \mu_\beta^{[1]} & A_{11} \mu_\beta^{[1,2]} \\ A_{21} + A_{22} \mu_\beta^{[1]} & A_{21} \mu_\beta^{[1]} & A_{21} \mu_\beta^{[1,2]} \\ 0_{K_s} & 0_{K_s \times K_f} & 0_{K_s \times K_f} \end{array} \right] \Gamma_f^P \\
= \left[ \begin{array}{ccc} 1 & \sigma^2 \mu_\beta^{[1]} D^{-1}(\tilde{F}^{[1]}M_T F^{[1]})^{-1} & \mu_\beta^{[2]} - \mu_\beta^{[1]} D^{-1}(\tilde{F}^{[1]} - \mu_\beta^{[1]} \mu_\beta^{[2]} \tilde{F}^{[2]}) \\ 0_{K_s} & D^{-1}(\Sigma_\beta^{[1]} - \mu_\beta^{[1]} \mu_\beta^{[2]} \Sigma_\beta^{[2]}) & 0_{K_s \times K_f} \end{array} \right] \Gamma_f^P \\
= \left[ \begin{array}{ccc} 0_{K_s} & 0_{K_s \times K_f} & 0_{K_s \times K_f} \end{array} \right] \\
= \left[ \begin{array}{ccc} \gamma_0 & \gamma_0 \end{array} \right] + \tilde{d}_0 \\
= D^{-1}(\Sigma_\beta^{[1]} - \mu_\beta^{[1]} \mu_\beta^{[2]} \Sigma_\beta^{[2]}) \gamma_1^{P[1]} + D^{-1}(\Sigma_\beta^{[1,2]} - \mu_\beta^{[1]} \mu_\beta^{[2]} \gamma_1^{P[2]} \\
0_{K_s} \end{array} \right] \\

Setting \( \tilde{d}_0 = \sigma^2 \mu_\beta^{[1]} D^{-1}(\tilde{F}^{[1]} - \mu_\beta^{[1]} \mu_\beta^{[2]} \tilde{F}^{[2]}) \gamma_1^{P[1]} + \mu_\beta^{[2]} \gamma_1^{P[2]} - \mu_\beta^{[1]} D^{-1}(\tilde{F}^{[1]} - \mu_\beta^{[1]} \mu_\beta^{[2]} \gamma_1^{P[2]}) \end{array} \right], \\
\tilde{d}_{11} = D^{-1}(\Sigma_\beta^{[1]} - \mu_\beta^{[1]} \mu_\beta^{[2]} \gamma_1^{P[1]} \gamma_1^{P[2]}), \text{ and } \tilde{d}_{12} = D^{-1}(\Sigma_\beta^{[1,2]} - \mu_\beta^{[1]} \mu_\beta^{[2]} \gamma_1^{P[1]} \gamma_1^{P[2]}), \end{array} \right]
\]

concludes the proof of part (i).

(ii) Notice that
\[ \hat{\mathbf{f}}_{1,9} = \left( \frac{\hat{\mathbf{X}}_{1,9} \hat{\mathbf{X}}_{1,9}}{N} \right)^{-1} \frac{\hat{\mathbf{X}}_{1,9} \hat{\mathbf{R}}}{N} \]

\[ = \left( \frac{\hat{\mathbf{X}}_{1,9} \hat{\mathbf{X}}_{1,9}}{N} \right)^{-1} \times \left\{ \begin{array}{c}
\frac{B^{[1]}}{N} + \frac{P^{[1]} \epsilon \gamma}{N} + \frac{1/\epsilon + 1/\gamma}{N} \\
\frac{B^{[2]}}{N} + \frac{P^{[2]} \epsilon \gamma}{N} + \frac{1/\epsilon + 1/\gamma}{N}
\end{array} \right\} \gamma P_f + \left\{ \begin{array}{c}
\frac{1/\epsilon + 1/\gamma}{N} \\
\frac{1/\epsilon + 1/\gamma}{N}
\end{array} \right\} \}

\[ = \left( \frac{\hat{\mathbf{X}}_{1,9} \hat{\mathbf{X}}_{1,9}}{N} \right)^{-1} \times \left\{ \begin{array}{c}
\frac{B^{[1]}}{N \gamma} + \frac{P^{[1]} \epsilon \gamma}{N \gamma} + \frac{1/\epsilon + 1/\gamma}{N \gamma} \\
\frac{B^{[2]}}{N \gamma} + \frac{P^{[2]} \epsilon \gamma}{N \gamma} + \frac{1/\epsilon + 1/\gamma}{N \gamma}
\end{array} \right\} \gamma P_f + \left\{ \begin{array}{c}
\frac{1/\epsilon + 1/\gamma}{N \gamma} \\
\frac{1/\epsilon + 1/\gamma}{N \gamma}
\end{array} \right\} \}

\[ = \left( \frac{\hat{\mathbf{X}}_{1,9} \hat{\mathbf{X}}_{1,9}}{N} \right)^{-1} \times \left\{ \begin{array}{c}
\frac{B^{[1]}}{N \gamma} + \frac{P^{[1]} \epsilon \gamma}{N \gamma} + \frac{1/\epsilon + 1/\gamma}{N \gamma} \\
\frac{B^{[2]}}{N \gamma} + \frac{P^{[2]} \epsilon \gamma}{N \gamma} + \frac{1/\epsilon + 1/\gamma}{N \gamma}
\end{array} \right\} \gamma P_f + \left\{ \begin{array}{c}
\frac{1/\epsilon + 1/\gamma}{N \gamma} \\
\frac{1/\epsilon + 1/\gamma}{N \gamma}
\end{array} \right\} \}

\[ = \left( \frac{\hat{\mathbf{X}}_{1,9} \hat{\mathbf{X}}_{1,9}}{N} \right)^{-1} \times \left\{ \begin{array}{c}
\frac{B^{[1]}}{N \gamma} + \frac{P^{[1]} \epsilon \gamma}{N \gamma} + \frac{1/\epsilon + 1/\gamma}{N \gamma} \\
\frac{B^{[2]}}{N \gamma} + \frac{P^{[2]} \epsilon \gamma}{N \gamma} + \frac{1/\epsilon + 1/\gamma}{N \gamma}
\end{array} \right\} \gamma P_f + \left\{ \begin{array}{c}
\frac{1/\epsilon + 1/\gamma}{N \gamma} \\
\frac{1/\epsilon + 1/\gamma}{N \gamma}
\end{array} \right\} \}

\]

Therefore,
\[
\hat{F}_{f, g} - \left[ \begin{array}{c}
\dd_1^2 \gamma_{1, f}^1 + \dd_2 \gamma_{1, f}^2 \\
0_{K_g}
\end{array} \right]
= \left( \frac{\hat{X}'_{f, g} \hat{X}_{f, g}}{N} \right)^{-1} \times \\
\times \left[ \begin{array}{c}
\gamma_0 + \dd_0 \\
\dd_1^2 \gamma_{1, f}^1 + \dd_2 \gamma_{1, f}^2
\end{array} \right] \\
= \left( \frac{\hat{X}'_{f, g} \hat{X}_{f, g}}{N} \right)^{-1} \times \\
\times \left[ \begin{array}{c}
\gamma_0 + \dd_0 \\
\dd_1^2 \gamma_{1, f}^1 + \dd_2 \gamma_{1, f}^2
\end{array} \right]
\]

where
\[
\left( \frac{\hat{X}_{f_{1,2}} \hat{X}_{f_{1,2}}}{N} \right) \begin{bmatrix}
\gamma_0 + \tilde{d}_0 \\
\tilde{d}_{11} \gamma_{1,f} \\
\tilde{d}_{12} \gamma_{2,f} \\
0_{K_s}
\end{bmatrix} = \\
= \begin{bmatrix}
\frac{1}{\tilde{B}_1 N} \frac{1}{\tilde{b}_1 N} \\
\frac{\tilde{B}_2}{\tilde{b}_2 N} \\
\frac{\tilde{B}_3}{\tilde{b}_3 N} \\
0_{K_s}
\end{bmatrix} \begin{bmatrix}
\frac{1}{\tilde{B}_1 N} + \frac{1}{\tilde{b}_1 N} \\
\frac{\tilde{B}_2}{\tilde{b}_2 N} \\
\frac{\tilde{B}_3}{\tilde{b}_3 N} \\
0_{K_s}
\end{bmatrix} \begin{bmatrix}
\gamma_0 + \tilde{d}_0 \\
\gamma_{1,f} \\
\gamma_{2,f} \\
0_{K_s}
\end{bmatrix} \\
= \begin{bmatrix}
\frac{1}{\tilde{B}_1 N} + \frac{1}{\tilde{b}_1 N} \\
\frac{\tilde{B}_2}{\tilde{b}_2 N} \\
\frac{\tilde{B}_3}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N}
\end{bmatrix} \begin{bmatrix}
\gamma_0 + \tilde{d}_0 \\
\gamma_{1,f} \\
\gamma_{2,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f}
\end{bmatrix} \\
= \begin{bmatrix}
\frac{1}{\tilde{B}_1 N} + \frac{1}{\tilde{b}_1 N} \\
\frac{\tilde{B}_2}{\tilde{b}_2 N} \\
\frac{\tilde{B}_3}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N}
\end{bmatrix} \begin{bmatrix}
\gamma_0 + \tilde{d}_0 \\
\gamma_{1,f} \\
\gamma_{2,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f}
\end{bmatrix}
\times
\begin{bmatrix}
\frac{1}{\tilde{B}_1 N} + \frac{1}{\tilde{b}_1 N} \\
\frac{\tilde{B}_2}{\tilde{b}_2 N} \\
\frac{\tilde{B}_3}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N}
\end{bmatrix} \begin{bmatrix}
\gamma_0 + \tilde{d}_0 \\
\gamma_{1,f} \\
\gamma_{2,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f}
\end{bmatrix}
\times
\begin{bmatrix}
\frac{1}{\tilde{B}_1 N} + \frac{1}{\tilde{b}_1 N} \\
\frac{\tilde{B}_2}{\tilde{b}_2 N} \\
\frac{\tilde{B}_3}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N} \\
\frac{1}{\tilde{b}_3 N} + \frac{1}{\tilde{b}_3 N}
\end{bmatrix} \begin{bmatrix}
\gamma_0 + \tilde{d}_0 \\
\gamma_{1,f} \\
\gamma_{2,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f}
\end{bmatrix}
\end{equation}

setting \( \gamma_0 = \gamma_0 + \tilde{d}_0 \) and \( \gamma_{1,f} = D^{-1} \left( \Sigma_{\beta_f}^{[1]} - \mu_{\beta_f}^{[1]} \mu_{\beta_f}^{[1]}' \right) \gamma_{1,f}^{[1]} + D^{-1} \left( \Sigma_{\beta_f}^{[2]} - \mu_{\beta_f}^{[2]} \mu_{\beta_f}^{[2]}' \right) \gamma_{1,f}^{[2]} \). This implies that

\[
\hat{F}_{f_{1,2}} = \begin{bmatrix}
\frac{1}{\tilde{B}_1 N} + \frac{1}{\tilde{b}_1 N} \\
\frac{\tilde{B}_2}{\tilde{b}_2 N} \\
\frac{\tilde{B}_3}{\tilde{b}_3 N} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f}
\end{bmatrix} \begin{bmatrix}
\gamma_0 + \tilde{d}_0 \\
\gamma_{1,f} \\
\gamma_{2,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f} \\
\gamma_{1,f}
\end{bmatrix} = \left( \frac{\hat{X}_{f_{1,2}} \hat{X}_{f_{1,2}}}{N} \right)^{-1} \begin{bmatrix}
E_1 \\
E_2 \\
E_3
\end{bmatrix}
\end{equation}

(3.C.135)
where

\[
E_1 = \gamma_0 + \frac{1}{N} \sum_{i=1}^{N} B_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} D_{ij}^{[2]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} C_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} D_{ij}^{[2]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} E_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} F_{ij}^{[1]} \gamma_{1j}
\]

\[
= \gamma_0 + \frac{1}{N} \sum_{i=1}^{N} B_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} D_{ij}^{[2]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} E_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} F_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} G_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} H_{ij}^{[1]} \gamma_{1j}
\]

\[
= \gamma_0 + \frac{1}{N} \sum_{i=1}^{N} B_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} D_{ij}^{[2]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} E_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} F_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} G_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} H_{ij}^{[1]} \gamma_{1j}
\]

\[
= \gamma_0 + \frac{1}{N} \sum_{i=1}^{N} B_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} D_{ij}^{[2]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} E_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} F_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} G_{ij}^{[1]} \gamma_{1j} + \frac{1}{N} \sum_{i=1}^{N} H_{ij}^{[1]} \gamma_{1j}
\]

\[
\text{since}
\]

\[
\left( \frac{1}{N} \sum_{i=1}^{N} B_{ij}^{[1]} \gamma_{1j} \right) = \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right)
\]

and

\[
\left( \frac{1}{N} \sum_{i=1}^{N} D_{ij}^{[2]} \gamma_{1j} \right) = \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right)
\]

and where we define the T-vector

\[
Q_{ij}^{[1,2]} = \left( \frac{1}{T} - P_{ij}^{[1]} D^{-1} \left( \Sigma_{ij}^{[1]} - \mu_{ij}^{[1]}' \mu_{ij}^{[1]} \right) \gamma_{1j}^{[1]} - P_{ij}^{[2]} D^{-1} \left( \Sigma_{ij}^{[2]} - \mu_{ij}^{[2]}' \mu_{ij}^{[2]} \right) \gamma_{1j}^{[2]} \right).
\]

Moreover,
$$E_2 = \frac{B_j^{[1]'}}{N} \gamma_0 + \frac{P_j^{[1]'}}{N} \epsilon \gamma_N \gamma_0 + \frac{P_j^{[1]'}}{N} \epsilon B_j^{[1]} \gamma_j \gamma_1 \gamma^{[1]} + \frac{P_j^{[1]'}}{N} \epsilon B_j^{[1]} \gamma_j \gamma_1 \gamma^{[1]}$$

$$+ \frac{B_j^{[2]'}}{N} \gamma_1 \gamma_2 \gamma^{[2]} + \frac{P_j^{[2]'}}{N} \epsilon B_j^{[2]} \gamma_1 \gamma_2 \gamma^{[2]} + \frac{B_j^{[2]'}}{N} \epsilon \gamma_1 \gamma_2 \gamma^{[2]} + \frac{P_j^{[2]'}}{N} \epsilon \gamma_1 \gamma_2 \gamma^{[2]}$$

$$- \frac{B_j^{[1]'}}{N} \gamma_0 + \frac{P_j^{[1]'}}{N} \epsilon \gamma_N \gamma_0 + \frac{B_j^{[1]'}}{N} \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$- \frac{P_j^{[1]'}}{N} \epsilon \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$= \frac{B_j^{[1]'}}{N} \gamma_0 + \frac{P_j^{[1]'}}{N} \epsilon \gamma_N \gamma_0 + \frac{B_j^{[1]'}}{N} \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$+ \frac{P_j^{[2]'}}{N} \epsilon \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$- \frac{B_j^{[1]'}}{N} \gamma_0 + \frac{P_j^{[1]'}}{N} \epsilon \gamma_N \gamma_0 + \frac{B_j^{[1]'}}{N} \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$- \frac{P_j^{[1]'}}{N} \epsilon \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$= \frac{B_j^{[1]'}}{N} \gamma_0 + \frac{P_j^{[1]'}}{N} \epsilon \gamma_N \gamma_0 + \frac{B_j^{[1]'}}{N} \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$+ \frac{P_j^{[2]'}}{N} \epsilon \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$- \frac{B_j^{[1]'}}{N} \gamma_0 + \frac{P_j^{[1]'}}{N} \epsilon \gamma_N \gamma_0 + \frac{B_j^{[1]'}}{N} \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$- \frac{P_j^{[1]'}}{N} \epsilon \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$= \frac{B_j^{[1]'}}{N} \gamma_0 + \frac{P_j^{[1]'}}{N} \epsilon \gamma_N \gamma_0 + \frac{B_j^{[1]'}}{N} \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$+ \frac{P_j^{[2]'}}{N} \epsilon \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$- \frac{B_j^{[1]'}}{N} \gamma_0 + \frac{P_j^{[1]'}}{N} \epsilon \gamma_N \gamma_0 + \frac{B_j^{[1]'}}{N} \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$- \frac{P_j^{[1]'}}{N} \epsilon \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$= \frac{B_j^{[1]'}}{N} \gamma_0 + \frac{P_j^{[1]'}}{N} \epsilon \gamma_N \gamma_0 + \frac{B_j^{[1]'}}{N} \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

$$+ \frac{P_j^{[2]'}}{N} \epsilon \gamma_1 \gamma_2 \gamma^{[2]} \gamma_j$$

Now, notice that
\[
\frac{B_{f}^{(1)}' B_{f}^{(1)} / \gamma_{1f}^{P[1]}}{N} - \frac{B_{f}^{(1)}' 1N / \gamma_{1f}^{P[1]}}{N} \sigma^2 \mu_{\beta_f}^{(1)} D^{-1} (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} \gamma_{1f}^{P[1]}
\]
\[
- \frac{B_{f}^{(1)}' B_{f}^{(1)} / \gamma_{1f}^{P[1]}}{N} D^{-1} \left( \Sigma_{\beta_f}^{(1)} - \mu_{\beta_f}^{(1)} \mu_{\beta_f}^{(1)'} \right) \gamma_{1f}^{P[1]} - \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} D^{-1} \left( \Sigma_{\beta_f}^{(1)} - \mu_{\beta_f}^{(1)} \mu_{\beta_f}^{(1)'} \right) \gamma_{1f}^{P[1]}
\]
\[
= \frac{B_{f}^{(1)}' B_{f}^{(1)}}{N} D^{-1} \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} \gamma_{1f}^{P[1]} - \frac{B_{f}^{(1)}' 1N / \gamma_{1f}^{P[1]}}{N} \sigma^2 \mu_{\beta_f}^{(1)} D^{-1} (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} \gamma_{1f}^{P[1]}
\]
\[
- \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} D^{-1} \left( \Sigma_{\beta_f}^{(1)} - \mu_{\beta_f}^{(1)} \mu_{\beta_f}^{(1)'} \right) \gamma_{1f}^{P[1]}
\]
\[
= \left( \frac{B_{f}^{(1)}' B_{f}^{(1)}}{N} - \frac{B_{f}^{(1)}' 1N \mu_{\beta_f}^{(1)}}{N} \right) D^{-1} \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} \gamma_{1f}^{P[1]}
\]
\[
- \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} D^{-1} \left( \Sigma_{\beta_f}^{(1)} - \mu_{\beta_f}^{(1)} \mu_{\beta_f}^{(1)'} \right) \gamma_{1f}^{P[1]}
\]
\[
= \left( \frac{B_{f}^{(1)}' B_{f}^{(1)}}{N} - \frac{B_{f}^{(1)}' 1N \mu_{\beta_f}^{(1)}}{N} + \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} \right) D^{-1} \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} \gamma_{1f}^{P[1]} - \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} \gamma_{1f}^{P[1]}
\]
\[
= DD^{-1} \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} \gamma_{1f}^{P[1]} - \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} \gamma_{1f}^{P[1]} + o_p \left( \frac{1}{\sqrt{N}} \right)
\]
\[
= 0_{K_{t1}} + o_p(1),
\]

and

\[
\frac{B_{f}^{(2)}' B_{f}^{(2)} / \gamma_{1f}^{P[2]}}{N} - \frac{B_{f}^{(2)}' 1N / \gamma_{1f}^{P[2]}}{N} \mu_{\beta_f}^{(2)} \gamma_{1f}^{P[2]}
\]
\[
+ \frac{B_{f}^{(2)}' 1N / \gamma_{1f}^{P[2]}}{N} \mu_{\beta_f}^{(2)} D^{-1} \left( \Sigma_{\beta_f}^{(2)} - \mu_{\beta_f}^{(2)} \mu_{\beta_f}^{(2)'} \right) \gamma_{1f}^{P[2]} - \frac{B_{f}^{(2)}' B_{f}^{(2)}}{N} D^{-1} \left( \Sigma_{\beta_f}^{(2)} - \mu_{\beta_f}^{(2)} \mu_{\beta_f}^{(2)'} \right) \gamma_{1f}^{P[2]}
\]
\[
- \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} D^{-1} \left( \Sigma_{\beta_f}^{(2)} - \mu_{\beta_f}^{(2)} \mu_{\beta_f}^{(2)'} \right) \gamma_{1f}^{P[2]}
\]
\[
= \left( \frac{B_{f}^{(2)}' B_{f}^{(2)}}{N} - \frac{B_{f}^{(2)}' 1N \mu_{\beta_f}^{(2)}}{N} \right) \gamma_{1f}^{P[2]}
\]
\[
- \left( \frac{B_{f}^{(2)}' B_{f}^{(2)}}{N} - \frac{B_{f}^{(2)}' 1N \mu_{\beta_f}^{(2)}}{N} + \sigma^2 (\tilde{F}^{(1)}' \tilde{F}^{(1)})^{-1} \right) D^{-1} \left( \Sigma_{\beta_f}^{(2)} - \mu_{\beta_f}^{(2)} \mu_{\beta_f}^{(2)'} \right) \gamma_{1f}^{P[2]}
\]
\[
= \left( \Sigma_{\beta_f}^{(2)} - \mu_{\beta_f}^{(2)} \mu_{\beta_f}^{(2)'} \right) \gamma_{1f}^{P[2]} - DD^{-1} \left( \Sigma_{\beta_f}^{(2)} - \mu_{\beta_f}^{(2)} \mu_{\beta_f}^{(2)'} \right) \gamma_{1f}^{P[2]} + o_p \left( \frac{1}{\sqrt{N}} \right)
\]
\[
= 0_{K_{t1}} + o_p(1),
\]

Moreover,
\[
\frac{P_j^{[1]'}}{N} eB_j^{[1]}_{\gamma_j} P_j^{[1]} = \frac{P_j^{[1]'}}{N} eB_j^{[1]}_{\gamma_j} D^{-1} \left( \Sigma_{\beta_j}^{[1],} - \mu_{\beta_j}^{[1],} \right) \gamma_j^{[1]} - \frac{P_j^{[1]'}}{N} e1N_{\mu_{\beta_j}^{[1],} D^{-1} (\tilde{F}^{[1]'})^{-1} \gamma_j^{[1]}} = P_j^{[1]'} e \left( \frac{B_j^{[1]}}{N} - \frac{1}{N} \mu_{\beta_j}^{[1]} \right) D^{-1} \sigma^2 (\tilde{F}^{[1]'})^{-1} \gamma_j^{[1]},
\]

and

\[
\frac{P_j^{[1]'}}{N} eB_j^{[2]}_{\gamma_j} P_j^{[2]} = \frac{P_j^{[1]'}}{N} e1N_{\mu_{\beta_j}^{[2],} \gamma_j^{[2]}} = P_j^{[1]'} e \left( \frac{B_j^{[2]}}{N} - \frac{1}{N} \mu_{\beta_j}^{[2]} \right) \gamma_j^{[2]},
\]

This implies that

\[
E_2 = \frac{B_j^{[1]'}}{N} e' Q_j^{[1,2]} + \left( P_j^{[1]'} eB_j^{[1]}_{\gamma_j} - P_j^{[1]'} e1N_{\mu_{\beta_j}^{[1],} D^{-1} (\tilde{F}^{[1]'})^{-1} \gamma_j^{[1]}} \right) \sigma^2 (\tilde{F}^{[1]'})^{-1} \gamma_j^{[1]} + \left( \frac{P_j^{[1]'}}{N} eB_j^{[2]}_{\gamma_j} P_j^{[2]} - P_j^{[1]'} \frac{1}{N} \mu_{\beta_j}^{[2]} \right) \gamma_j^{[2]} + P_j^{[1]'} e \left( \frac{e'}{N} - \sigma^2 I_T \right) Q_j^{[1,2]} + o_p \left( \frac{1}{\sqrt{N}} \right).
\]

Finally,
\[ E_3 = \frac{P_s' e_1 N}{N \gamma_0} + \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} + \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} + \frac{P_s' e_1 T}{N T} \]

\[ - \left( \frac{P_s' e_1 N}{N \gamma_0} - \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \right) \left( \frac{\epsilon_0'}{N} - \frac{\epsilon^2 I_T}{T} \right) \]

\[ \frac{P_s' e_1 N}{N \gamma_0} - \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 N}{N \gamma_0} - \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

since

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]

\[ \frac{P_s' e_1 B_f^1}{N \gamma_1 f'} - \frac{P_s' e_1 B_f^2}{N \gamma_1 f'} - \frac{P_s' e_1 T}{N T} \times \frac{\gamma_0}{\gamma_1 f'} \]
Now, notice that

\[ \gamma_{1_f}^{P[1]} = \tilde{d}_{11} \gamma_{1_f}^{P[1]} + \tilde{d}_{12} \gamma_{2_f}^{P[2]} \]
\[ = D^{-1} \Sigma_{\beta_f}^{P[1]} \gamma_{1_f}^{P[1]} + D^{-1} \Sigma_{\beta_f}^{P[2]} \gamma_{2_f}^{P[2]} \]
\[ = \gamma_{1_f}^{P[1]} - D^{-1} \sigma^2 (\bar{E}_{f}^{[1]} - \bar{E}_{f}^{[2]})^{-1} \gamma_{1_f}^{P[1]} + D^{-1} \Sigma_{\beta_f}^{1,2} \gamma_{2_f}^{P[2]} \]
\[ = \gamma_{1_f}^{P[1]} + \tilde{d}_1 \]

where we define \( \tilde{d}_1 = -D^{-1} \sigma^2 (\bar{E}_{f}^{[1]} - \bar{E}_{f}^{[2]})^{-1} \gamma_{1_f}^{P[1]} + D^{-1} \Sigma_{\beta_f}^{1,2} \gamma_{2_f}^{P[2]} \). It implies that

\[ E_2 = \frac{B_f^{[1]'} \epsilon^{[2]} Q_f^{[1,2]}}{N} + \left( \frac{P_f^{[1]'} \epsilon B_f^{[1]} - P_f^{[1]'} \epsilon \Sigma_{\beta_f}^{P[1]}}{N} \right) \tilde{d}_1 + \left( \frac{P_f^{[1]'} \epsilon B_f^{[2]} - P_f^{[1]'} \epsilon \Sigma_{\beta_f}^{P[2]}}{N} \right) \gamma_{2_f}^{P[2]} \]
\[ + P_g^{[1]'} \left( \frac{\epsilon \Sigma_{\beta_f}^{P[1]} - \sigma^2 I_T}{N} \right) Q_f^{[1,2]} + o_p \left( \frac{1}{\sqrt{N}} \right), \]

and

\[ E_3 = \left( \frac{P_g^{[1]'} \epsilon B_f^{[1]} - P_g^{[1]'} \epsilon \Sigma_{\beta_f}^{P[1]}}{N} \right) \tilde{d}_1 + \left( \frac{P_g^{[1]'} \epsilon B_f^{[2]} - P_g^{[1]'} \epsilon \Sigma_{\beta_f}^{P[2]}}{N} \right) \gamma_{2_f}^{P[2]} \]
\[ + P_g^{[1]'} \left( \frac{\epsilon \Sigma_{\beta_f}^{P[1]} - \sigma^2 I_T}{N} \right) Q_f^{[1,2]} + o_p \left( \frac{1}{\sqrt{N}} \right) \]

Therefore, putting all together, we can write

\[ \sqrt{N} \left( \tilde{\Gamma}_{f_{1g}} - \left[ D^{-1} \left( \Sigma_{\beta_f}^{[1]} - \mu_{\beta_f}^{[1]} \mu_{\beta_f}^{[1]'} \right) \gamma_{1_f}^{P[1]} + \tilde{d}_1 \right] \right) \]
\[ = \left( \frac{\hat{X}_{f_{1g}} \hat{X}_{f_{1g}}}{N} \right)^{-1} \left[ \begin{array}{c} \frac{B_f^{[1]'} \epsilon^{[2]} Q_f^{[1,2]}}{\sqrt{N}} - \left( \frac{P_f^{[1]'} \epsilon B_f^{[1]} - P_f^{[1]'} \epsilon \Sigma_{\beta_f}^{P[1]}}{N} \right) \tilde{d}_1 + \left( \frac{P_f^{[1]'} \epsilon B_f^{[2]} - P_f^{[1]'} \epsilon \Sigma_{\beta_f}^{P[2]}}{N} \right) \gamma_{2_f}^{P[2]} \\
- \left( \frac{P_g^{[1]'} \epsilon B_f^{[1]} - P_g^{[1]'} \epsilon \Sigma_{\beta_f}^{P[1]}}{N} \right) \tilde{d}_1 + \left( \frac{P_g^{[1]'} \epsilon B_f^{[2]} - P_g^{[1]'} \epsilon \Sigma_{\beta_f}^{P[2]}}{N} \right) \gamma_{2_f}^{P[2]} \\
\end{array} \right] + o_p(1) \]
\[ + \left[ \begin{array}{c} \frac{P_f^{[1]'} \epsilon^{[2]} Q_f^{[1,2]}}{\sqrt{N}} \left( \frac{\epsilon^{[2]} - \sigma^2 I_T}{N} \right) Q_f^{[1,2]} \\
\frac{P_g^{[1]'} \epsilon^{[2]} Q_f^{[1,2]}}{\sqrt{N}} \left( \frac{\epsilon^{[2]} - \sigma^2 I_T}{N} \right) Q_f^{[1,2]} \\
\end{array} \right] + o_p(1) \]
\[ = \left( \frac{\hat{X}_{f_{1g}} \hat{X}_{f_{1g}}}{N} \right)^{-1} (I_1 + I_2) + o_p(1) \]  

(3.C.138)

Now, using (3.C.128), then

\[ \left( \frac{\hat{X}_{f_{1g}} \hat{X}_{f_{1g}}}{N} \right)^{-1} \rightarrow_p \left( \Sigma_{\tilde{X}_{f_{1g}}} + \Lambda_{f_{1g}}^{[1]} \right)^{-1} \]
Consider now $I_1$ and notice that it has zero mean. Therefore

$$\text{Var}(I_1) =$$

$$= \text{Var} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \frac{Q_f^{[1]} \otimes \beta_f^{[1]}}{\sqrt{N}} - \left( d_i (\beta_f^{[1]} - \mu_{\beta_f^{[1]}}) - \gamma_f^{[2]} (\beta_f^{[2]} - \mu_{\beta_f^{[2]}}) \right) \epsilon_i \right) \right]$$

$$\rightarrow_p \begin{bmatrix}
\sigma^2 Q_f^{[1,2]} \otimes \mu_{\beta_f^{[1]}} & \sigma^2 Q_f^{[1,2]} \otimes \mu_{\beta_f^{[2]}} \\
\sigma^2 Q_f^{[1,2]} \otimes \mu_{\beta_f^{[2]}} & \sigma^2 Q_f^{[1,2]} \otimes \mu_{\beta_f^{[2]}} \\
0_{K_s} & 0_{K_s} \\
0_{K_s \times K_{f_1}} & 0_{K_s \times K_{f_1}} \\
0_{K_s} & 0_{K_s} \\
0_{K_s \times K_{f_1}} & 0_{K_s \times K_{f_1}}
\end{bmatrix}
$$

$$+ \sigma^2 \begin{bmatrix}
0_{K_{f_1}} & 0_{K_{f_1}} & 0_{K_{f_1} \times K_s} & 0_{K_{f_1} \times K_s} \\
0_{K_s} & 0_{K_s} & 0_{K_s \times K_{f_1}} & 0_{K_s \times K_{f_1}} \\
0_{K_s} & 0_{K_s} & 0_{K_s \times K_{f_1}} & 0_{K_s \times K_{f_1}}
\end{bmatrix}$$

$$= \sigma^2 Q_f^{[1,2]} \otimes \mu_{\beta_f^{[1]}} + \sigma^2 \Omega_{f g}$$

$$= \sigma^2 \left( \frac{\bar{d}_{11} \gamma_f^{[1]} + \bar{d}_{12} \gamma_f^{[2]}}{T} (\bar{F} \bar{F})^{-1} (\bar{d}_{11} \gamma_f^{[1]} + \bar{d}_{12} \gamma_f^{[2]}) \right) \Sigma_{K_{f g}}^{[1]} + \sigma^2 \Omega_{f g} = V_{f g}$$

where

$$\Sigma_{K_{f g}}^{[1]} = \begin{bmatrix}
1 & \mu_{\beta_f^{[1]}} \\
\mu_{\beta_f^{[2]}} & \Sigma_{\beta_f^{[1]}}^{[1]} \\
0_{K_s} & 0_{K_s \times K_{f_1}}
\end{bmatrix}$$

and

$$\Omega_{f g} = \begin{bmatrix}
0_{K_{f_1}} & 0_{K_{f_1}} & 0_{K_{f_1} \times K_s} & 0_{K_{f_1} \times K_s} \\
0_{K_s} & 0_{K_s} & 0_{K_s \times K_{f_1}} & 0_{K_s \times K_{f_1}}
\end{bmatrix}$$

with

$$\varphi = \left( \bar{d}_{1} \Sigma_{\beta_f^{[1]}} \bar{d}_1 + \gamma_f^{[2]} \bar{\Sigma}_{\beta_f^{[2]}} \gamma_f^{[2]} - 2 \gamma_f^{[2]} \Sigma_{\beta_f^{[2]}} \Sigma_{\beta_f^{[1]}} \bar{d}_1 \right).$$

Let us analyze the term

$$\begin{bmatrix}
P_f^{[1]} \sqrt{N} \left( \mu_{\beta_f^{[1]}} - \Sigma_{\beta_f^{[1]}} I_T \right) Q_f^{[1,2]} \\
P_f^{[2]} \sqrt{N} \left( \mu_{\beta_f^{[2]}} - \Sigma_{\beta_f^{[2]}} I_T \right) Q_f^{[1,2]}
\end{bmatrix}.$$
where

\[
E[W_{11}] = \mathbb{E} \left[ P_{f}^{1,2} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2_i I_T) Q_f^{1,2} \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (\epsilon_j \epsilon_j' - \sigma^2_j I_T) P_{f}^{1,2} \right] \\
= \mathbb{E} \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) \mathbb{E} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon_i' - \sigma^2_i I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \text{vec}'(\epsilon_j \epsilon_j' - \sigma^2_j I_T) \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) \right] \\
= \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) U_i \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) + o_p(1)
\]

\[
\rightarrow_p \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) U_i \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right), \tag{3.C.139}
\]

\[
E[W_{21}] = \mathbb{E} \left[ P_{g}^{1,2} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2_i I_T) Q_f^{1,2} \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (\epsilon_j \epsilon_j' - \sigma^2_j I_T) P_{f}^{1,2} \right] \\
= \mathbb{E} \left( Q_f^{1,2} \otimes P_{g} \right) \mathbb{E} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon_i' - \sigma^2_i I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \text{vec}'(\epsilon_j \epsilon_j' - \sigma^2_j I_T) \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) \right] \\
= \left( Q_f^{1,2} \otimes P_{g} \right) E \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon_i' - \sigma^2_i I_T) \text{vec}'(\epsilon_i \epsilon_i' - \sigma^2_i I_T) \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) \right] + o_p(1) \\
\rightarrow_p \left( Q_f^{1,2} \otimes P_{g} \right) U_i \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right), \tag{3.C.140}
\]

and

\[
E[W_{22}] = \mathbb{E} \left[ P_{g}^{1,2} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma^2_i I_T) Q_f^{1,2} \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (\epsilon_j \epsilon_j' - \sigma^2_j I_T) P_{g} \right] \\
= \mathbb{E} \left( Q_f^{1,2} \otimes P_{g} \right) \mathbb{E} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon_i' - \sigma^2_i I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \text{vec}'(\epsilon_j \epsilon_j' - \sigma^2_j I_T) \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) \right] \\
= \left( Q_f^{1,2} \otimes P_{g} \right) E \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i \epsilon_i' - \sigma^2_i I_T) \text{vec}'(\epsilon_i \epsilon_i' - \sigma^2_i I_T) \right] \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) + o_p(1) \\
\rightarrow_p \left( Q_f^{1,2} \otimes P_{g} \right) U_i \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right), \tag{3.C.141}
\]

It implies that

\[
\text{Var} \left( \begin{bmatrix} 0 & \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) U_i \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) \\ 0 & \left( Q_f^{1,2} \otimes P_{g} \right) U_i \left( Q_f^{1,2} \otimes P_{f}^{1,2} \right) \end{bmatrix} \right) \\
\rightarrow_p \begin{bmatrix} 0_{K_1} & 0_{K_2} \\ 0_{K_1} & 0_{K_2} \end{bmatrix} \equiv W_{fg}
\]
Assumption 3.4(vi) concludes the proof of part (ii). ■.
Bibliography


263


