Imperial College Business School
Imperial College London

Essays in Financial Economics

Adelina Georgiana Barbalau

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Imperial College London

May 2019

Abstract

The thesis investigates the way that decision-makers' understanding of risks impacts real and financial markets. The first two chapters explore theoretically how decision-makers learn about the risks that they are exposed to, and what are the implications of their information choices in terms of the transmission of shocks to the real economy. The last chapter proposes an econometric framework for empirically testing whether the compensation that decision-makers require for being exposed to systematic sources of risk in financial markets is in line with economic theory.

Declaration of originality

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

Adelina Barbalau

Copyright

Acknowledgements

I thank my supervisors, Franklin Allen and Savitar Sundaresan, for their invaluable guidance and support. Franklin, thank you for always believing in me. Savi, thank you for always being available.

I thank my family and friends for their love.

I thank myself, without whom this thesis would not have been possible.

Contents

List of Tables

List of Figures

Overview of the Thesis

The first chapter proposes an information based theory of shock transmission to explain how the impact of a shock can decrease with exposure to it. The chapter is motivated by evidence from the 2007-2008 financial crisis which shows that countries which were relatively more exposed to the crisis epicenter, the United States, were among the least affected. This evidence points to the existence of a negative correlation between the degree of exposure to a shock and the impact of that shock which has not been rationalised using existing theories of contagion. I propose a model in which decision-makers learn about the risk factors that they are exogenously exposed to, but have limited capacity to process information. I find that decision-makers optimally choose to learn more about the risk factors they are more exposed to. This informational advantage mitigates the direct impact of shocks to risk factors that decision-makers are relatively more exposed to because it enables them to take better informed investment decisions and thus minimize the loss due to suboptimal action. By the same token, the impact of shocks to risk factors that decision-makers are relatively less exposed to is amplified through their poorly informed investment decisions because they result in a higher loss due to suboptimal action. Relative to an exogenous information benchmark, the endogenous information model proposed here predicts that shocks to risk factors that decision-makers are relatively less exposed to are amplified, while shocks to risk factors that decision-makers are relatively more exposed to are attenuated.

The second chapter proposes an information based theory of state-dependent cautious behavior. The chapter is motivated by evidence that certain states of the world, such as economic downturns and financial crises in particular, are characterized by stronger reactions to negative news by economic agents, as well as stronger correlations between markets. This state-dependent cautious behavior, whereby negative news affect conditional actions more than good news, has been rationalized by assuming ambiguity-aversion. In this chapter, I endogenize ambiguity-averse behavior using costly information acquisition. I propose a model in which decision-makers can invest in information about future states of the world such that upon the occurrence of any state, they receive ambiguous signals: the precision of a given signal is not exactly know but it is only known to lie in a range or interval of possible signal precisions. The degree of ambiguity of these signals is determined by their informational investment decision. I find that the ambiguity of information in a state of nature varies inversely with the ex-ante degree of anticipation of that state. Uncertainty regarding the interpretation of information increases endogenously in highly unexpected states of nature because decision-makers optimally choose to learn less about events that are deemed to be unlikely. This causes ambiguity-averse decision-makers to behave cautiously by reacting more strongly to bad news than to good news. However, in highly anticipated states of the world there is no uncertainty regarding the interpretation of signals. Decision-makers behaviour no longer exhibits ambiguity-aversion, and good and bad news affect conditional actions in a symmetric fashion. The model explains why and how the behavior of decision-makers changes during crises, and delivers predictions that are in line with observed market outcomes. Relative to a no-ambiguity benchmark, the model I propose predicts that the transmission of negative shocks is amplified, while the transmission of positive shocks is attenuated as their degree of anticipation decreases.

The third chapter proposes a multivariate inequality testing framework to assess the consistency of risk-factor asset pricing models with the theoretical restrictions imposed by the intertemporal CAPM (ICAPM). Multifactor asset pricing models seek to explain cross-sectional differences in expected returns in terms of exposures to systematic risk factors. The ICAPM posits that investors should be compensated for being exposed to the risk of unexpected changes in the future investment opportunity set, which is captured by unidentified state variables. The fact that these state variables are not explicitly identified has allowed applied researchers to choose from a wide range of potential risk factors and use the ICAPM as a theoretical justification for relatively ad-hoc empirical specifications. However, the ICAPM imposes several theoretical restrictions on the time-series and cross-sectional behaviour of the candidate state variables. This chapter develops an inequality constraints testing framework for assessing the consistency of several multifactor models with the time-series and cross-sectional restrictions imposed by the ICAPM. The proposed test of joint sign restrictions takes into account the estimation error in the model parameters as well as the uncertainty arising from potential model misspecification. The framework is applied to test the consistency of several popular multifactor models with the ICAPM restrictions when using size and book-to-market, and size and momentum sorted portfolios as test assets. Results indicate that with a few exceptions, the null of consistency cannot be rejected. In other words, these theoretical restrictions are rarely violated in practice, suggesting that the ICAPM may be a fishing license as argued by Fama (1991).

Chapter 1

Information Choice, Shock Transmission and Contagion

1.1 Introduction

During the 2007-2008 financial crisis, countries that were relatively more exposed (in terms of assets, debt, exports or trade) to the crisis epicenter, the United States, were among the least affected in real terms.[1] In financial markets, equity portfolios that were relatively more exposed to a United States-specific factor experienced a drop in returns that was lower than the one predicted by their pre-crisis exposures.[2] This evidence points to the existence of a negative correlation between the degree of exposure to a shock and the impact of that shock.

The existing literature on international financial contagion does not explain how the impact of a shock can decrease with exposure to it.[3] Specifically, theories of contagion predict a monotonically increasing relationship between the degree of exposure to a risk factor and the impact of a shock to that risk factor. This chapter seeks to fill this gap in the literature. I propose an information based theory of contagion and provide a characterization of the conditions under which the impact of shocks decreases with exposure. The notion of contagion adopted in this chapter is one where the transmission of shocks is unexplained by the observable measure of exposure to those shocks.[4]

I introduce a framework in which decision-makers learn about the risk factors that they are

[1] See Rose and Spiegel (2010, 2011).

[2] See Bekaert, Ehrmann, Fratzscher, and Mehl (2014).

[3] Contagion is concerned with the transmission of shocks and can be most broadly described by the idea that shocks can spread and cause a great deal more damage than the original impact (Allen and Gale, 2009).

[4] This definition is in line with a large literature which defines contagion as a change in shock transmission mechanism that cannot be explained by "fundamentals", or co-movements that are deemed to be "excessive" (King and Wadhwani, 1990; Forbes and Rigobon, 2002; Karolyi, 2003; Pericoli and Sbracia, 2003; Jotikasthira, Lundblad, and Ramadorai, 2012; Bekaert et al., 2014).

exogenously exposed to, but have limited capacity to process information. I focus on studying the real consequences of shocks to these risk factors by embedding my framework in a simple model of corporate investment. The baseline model features a representative firm which undertakes investment to maximize profits. The return on investment depends on the fundamentals of the economy in which the firm operates. Economic fundamentals, in turn, are modelled as a sum of risk factor exposures. Before investing, the firm chooses how much information to observe about the risk factors that fundamentals are exposed to. Importantly, the firm has limited resources to acquire and process information.

I find that learning about a risk factor optimally increases with exposure to it. This informational advantage gives rise to a non-monotonic relation between the degree of exposure to a risk factor and the impact of a shock to that risk factor. Specifically, the firm optimally chooses to learn more about the risks that it is relatively more exposed to. The reduction in uncertainty achieved through learning mitigates the impact of shocks to risk factors that the firm is relatively more exposed to, by enabling it to take a better informed investment decision. By the same token, the impact of shocks to risk factors that the firm is relatively less exposed to is amplified through the poorly informed investment decision of the firm. The interpretation of these prediction in terms of the motivating evidence is that countries which were relatively less exposed to the United States shock were more affected because decision-makers operating in these countries had a poorer understanding of the shock and, as a consequence, took actions that aggravated their circumstances.

My model shows that the actions of decision-makers can amplify or attenuate the direct impact of shocks and thus contribute to their transmission. The model builds on the intuition that the consequences of events depend not only on decision-makers' direct exposure to events, but also on the decision-makers' degree of understanding of and their reactions to those events. Risk perception research supports the notion that reactions to events depend importantly on the degree of understanding of those events.[5] Hence, understanding how decision-makers learn is central to understanding how their actions contribute to the transmission of shocks. In this chapter, I explore how endogenous information choice affects decision-makers' responses to shocks and as a consequence the impact of those shocks.

The information based shock transmission mechanism I propose in this chapter works through decision-makers' uncertainty.[6] Information choice reduces the uncertainty that the firm faces when

[5] See Slovic (1987); Kasperson, Renn, Slovic, Brown, Emel, Goble, Kasperson, and Ratick (1988); Renn, Burns, Kasperson, Kasperson, and Slovic (1992).

[6] Uncertainty refers to the variance of beliefs about the realization of a particular shock.

choosing optimal investment. Consider a benchmark model in which the firm is endowed with an exogenous amount of information that is equally allocated among all risk factors. Relative to this exogenous information benchmark, uncertainty is lower when the firm can optimally choose which risks to learn about. Importantly, uncertainty is convex in exposure when the firm is exogenously endowed with information, but it is concave in exposure when the firm optimally learns about risk factor exposures. Alternatively stated, uncertainty increases with exposure to relatively important risk factors when information is given, but it decreases with exposure when information is optimally chosen.[7] In the endogenous learning model I propose, a change in exposure has two competing effects. On the one hand, higher exposure to a risk factor increases uncertainty mechanically, through an *exposure channel*. On the other hand, higher exposure to a risk factor entails a reduction in uncertainty via learning, which operates through an *information channel*. For relatively low levels of exposure to a risk factor, the exposure channel is stronger than the information channel, but for relatively high levels of exposure to a risk factor, the information channel is stronger than the exposure channel. As a consequence, uncertainty is concave in exposure to a risk factor. Thus, my model highlights a trade-off between the cost of being highly exposed to a risk and the benefit of having a better understanding of it.

Reducing uncertainty will enable the firm to accurately incorporate the shocks affecting fundamentals into investment decisions. To the extent that the firm's investment deviates from the first-best optimum obtained under perfect information, such deviations are suboptimal. In the case of positive shocks the deviation from optimality will manifest as under-investment, while in the case of negative shocks it will manifest as over-investment. It is thus convenient to define the loss due to suboptimal investment as the squared deviation from the perfect information optimum. The loss due to suboptimal investment essentially measures the contribution of actions to the transmission of shocks. I find that the loss due to suboptimal investment decreases with exposure when information is endogenously chosen, and it increases with exposure when information is exogenous. Relative to the exogenous information benchmark, the loss due to suboptimal investment is higher if a shock hits a risk factor that the firm is relatively less exposed to. On the other hand, if a shock hits a risk factor that is relatively more important in terms of exposure, the loss due to suboptimal investment is lower under the endogenous learning model than under the benchmark.

My model explains contagion manifested as shock amplification.[8] In other words it explains

[7] The importance of a risk factor is given by the interaction between the exposure to and the volatility of the factor.

[8] Amplification refers to situations in which small shocks can have disproportionately large effects (Allen and Gale, 2004; Krishnamurthy, 2010; Benoit, Colliard, Hurlin, and Pérignon, 2017)

how shocks can have disproportionately large effects. Importantly, it sheds light on two important dimensions of shock amplification, namely: (i) which of the shocks that an entity is exposed to are likely to amplify, and (ii) which of the entities exposed to a shock are likely to be more affected. The model predicts that the transmission of shocks is intensified as exposure to shocks decreases.[9] This prediction implies that: (i) the shocks that an entity is less exposed to are amplified, and (ii) the entities that are less exposed to a shock can be more affected.

I consider a number of extensions to the baseline model. First, I extend the baseline model to account for the degree of anticipation of shocks, and I am thus also able to explain why unanticipated crises are more contagious.[10] Specifically, the extended model predicts that the transmission of shocks is intensified as the degree of anticipation of shocks decreases. This prediction implies that unanticipated shocks are more likely to amplify and have more severe consequences. Second, I allow for strategic interactions and find that the amplification of shocks increases with the degree of strategic complementary in investment. Third, I relax the assumption that the risk factors affecting fundamentals are independent and find that the loss due to suboptimal investment decreases with the degree of correlation between the risks. Finally, I relax the assumption that exposures to the risk factors are exogenous and find that it is optimal for the firm to specialize in learning about one risk factor and to be relatively more exposed to that factor.

The rest of the chapter proceeds as follows. Section 1.2 formally introduces the mechanism through which the impact of shocks can decrease with exposure. Section 1.3 discusses the baseline results, and Section 1.4 considers a number of extensions to the baseline model. 1.5 outlines some concluding remarks. All proofs and derivations can be found in the Appendix.

1.1.1 Related literature

This chapter is mainly related and contributes to the literature on contagion, and the literature on rational inattention.

The contagion literature is vast and fraught with disagreement over how exactly to define contagion, how to measure it or what are the channels through which it operates. Contagion concerns itself with studying the transmission of shocks and can be most broadly described by the idea that shocks can spread and cause a great deal more damage than the original impact (Allen and Gale, 2009). The main dimensions of disagreement in the literature regarding what constitutes

[9] Although my baseline model is a representative agent model, comparative statics with respect to exposure are informative about a cross-section of agents which are exposed to the same set of risk factors, but which vary in their degree of exposure to these factors.

[10] See evidence in Kaminsky, Reinhart, and Vegh (2003); Rigobon and Wei (2003).

contagion have been focused around the types of linkages through which shocks are transmitted (in the presence or absence of linkages) and the types of shocks being transmitted (systemic or idiosyncratic shocks). In contrast to the contagion literature that studies shock transmission in the absence of linkages between the crisis epicenter and the entities subsequently affected, I study the transmission of shocks that occurs in the presence of linkages but which is unexplained by the observable shock transmission mechanism. In contrast to contagion literature that studies the transmission of idiosyncratic shocks across entities, I focus on studying the transmission a systemic or common shock that occurs differentially across the entities exposed to it.[11]

This chapter can be framed in the context of the literature that studies the transmission of shocks in the presence of linkages (i.e. exposure), but disproportionate to the objective measure of these linkages.[12] These include real channels such as trade linkages (Eichengreen, Rose, and Wyplosz, 1996; Glick and Rose, 1999; Forbes, 2004), as well as financial channels such as interbank linkages (Allen and Gale, 2000; Dasgupta, 2004; Freixas, Parigi, and Rochet, 2000; Iyer and Peydro, 2011) and portfolio linkages (Yuan, 2005; Pavlova and Rigobon, 2008; Jotikasthira et al., 2012; Manconi, Massa, and Yasuda, 2012). The basic idea underlying this literature is that decision-makers transmit shocks by directly altering the linkages or, alternatively stated, by changing their exposure to risks. My model offers an alternative explanation by showing that information choices and decisions about learning can effectively alter their risk exposures, and hence the transmission of shocks, even when decision-makers do not directly change their exposures.

The chapter is most related and contributes to the literature on information based models of financial crises and contagion (King and Wadhwani, 1990; Calvo and Mendoza, 2000; Kodres and Pritsker, 2002; Calvo, 2004; Acharya and Yorulmazer, 2008; Allen, Babus, and Carletti, 2012). The basic idea underlying this literature is that idiosyncratic shocks that should not be transmitted across entities if observable, are in fact transmitted because of imperfect information. There is no role for information choice in most of these models. In contrast, I study how endogenous information choices results in the differential transmission of a systemic shock in the cross-section of entities that are exposed to it. Mondria and Quintana-Domeque (2013) also explain financial contagion using fluctuations in attention allocation and information choice. However, they focus on a financial channel of contagion and study how international investors transmit an idiosyncratic shock by optimally reallocating their attention to the risk being shocked and away from the other risks in their portfolio. In contrast, I focus on a real channel of contagion and study how firms

[11]The terms entities is used generically to encompass countries, financial markets or financial institutions.

[12]This is to be contrasted with the theories of "pure contagion" which studies the transmission of shocks in the absence of direct linkages between the crisis epicentre and the entities being affected (Forbes, 2012).

undertaking investment optimally allocate their attention across the risks that they are exogenously exposed to and what this attention allocation implies for the transmission of subsequent shocks to these risk factors.[13] Also related is the paper by Ahnert and Kakhbod (2017), who propose an amplification mechanism of financial crises based on the information choice of investors. They propose a global game of regime change in which changes in the public signal affects the incentives of investors to acquire private information and thus amplify the probability of a crisis. The key mechanism in their model relies on the interplay between private and public information, whereas my mechanism relies on the fact that learning about one risk affect the ability to learn about other risks and there is no distinction between public and private information. In other words, I distinguish between information about different risks rather than information from different sources.

The chapter is also related to the rational inattention literature popularized by Sims (2003), which builds on the idea that attention, rather than information is a scarce resource. An increasing number of rational inattention applications focus on attention allocation across many risks.[14] A more recent strand of literature focuses on allocation across states of the world.[15] I micro-found a model of attention allocation across fixed exposures to risk factors, which I then also extend to account for attention allocation across states and thus provide a unified treatment of these two frameworks. Maćkowiak and Wiederholt (2015) also explore the idea that the degree to which decision-makers are prepared for events can exacerbate their consequences. While their model focuses on degree of anticipation of shocks to study how economic agents make state-contingent plans, my baseline model focuses on the degree of exposure to shocks to study how the transmission of a common shock varies across the agents exposed to it. My baseline model is extended to account for the degree of anticipation of shocks as well, so my work complements theirs in that I study attention allocation both across risky exposures (type of risk) and across states of the world (degree of anticipation of risk). The result from my model that diversified exposures are suboptimal under endogenous learning relates to the under-diversification result noted in the portfolio allocation literature (Van Nieuwerburgh and Veldkamp, 2010). While in a

[13]Alternatively stated, rather than studying how a shock to one risk factor affects the attention allocated to other factors, I focus on studying how exposure to one risk factor affects the attention allocated to other risk factors.

[14]Rational inattention applications to setups in which agents learn about many risks include asset pricing and portfolio choice (Mondria, 2010; Van Nieuwerburgh and Veldkamp, 2010; Kacperczyk, Van Nieuwerburgh, and Veldkamp, 2016), monetary policy (Woodford, 2001, 2009; Paciello and Wiederholt, 2013; Alvarez, Lippi, and Paciello, 2015), consumption dynamics (Luo, 2008; Tutino, 2013), price setting (Mackowiak and Wiederholt, 2009; Stevens, 2015; Matějka, 2015).

[15]Rational inattention applications to setups in which agents learn about future states of the world has been explored in a static setting by Maćkowiak and Wiederholt (2015) and in a dynamic setting by Sundaresan (2018); Nimark and Sundaresan (2018); Ilut and Valchev (2017).

portfolio allocation context decision-makers choose both their information about and exposure to risk factors, in my baseline model decision-makers only choose how much information to acquire about the risks they are exogenously exposed to. My model captures situations in which decision-makers are exposed to risk that are beyond their immediate control, with a focus on understanding how attention allocation affects the transmission of shocks.

1.2 Model

This section formally introduces the mechanism through which the impact of shocks can decrease with exposure. It outlines a baseline model of learning about exogenous risk factor exposures when the capacity to process information is limited. The basic result is that it is optimal to learn more about higher risk exposures and this information advantage can mitigate the negative impact or consequences of shocks.

1.2.1 Structure of the economy

I illustrate the mechanism in the context of a simple canonical model of investment. There is a risk-neutral, representative firm in the economy. The firm undertakes investment with an aim to maximize expected profits. Realized profits are given by

$$\pi = \lambda\theta - C(\lambda) \tag{1.1}$$

where λ is the chosen level of investment, $C(\lambda)$ is the cost of investment and θ is the exogenous gross return to investment. The return on investment is thus parametrized by an unknown exogenous state variable θ. Following Angeletos and Pavan (2004) and in line with the motivation, I interpret the random variable θ as the underlying fundamentals in the economy, but it can also be thought of as exogenous productivity or production technology. My preferred interpretation captures the idea that real investment returns are affected by the state of the economy in which the firm operates. The cost function $C(\lambda)$ is increasing and convex in investment, and assumed to take the form $\frac{\lambda^2}{2}$.

Economic fundamentals θ are a function of exogenous exposures to independent risk factors. More specifically, economic fundamentals are modelled as an exhaustive sum of independent risks

$$\theta = \alpha f_1 + (1-\alpha) f_2 \tag{1.2}$$

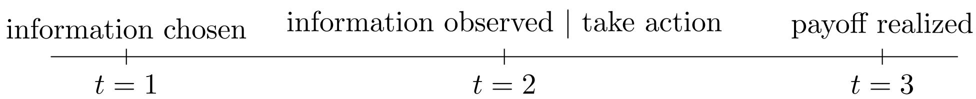
where f_1 and f_2 are exogenous risk factors which affect fundamentals in proportion to exogenous exposures or factor loadings α and $1-\alpha$ respectively. The risk factors f_i are given by

$$f_i = \mu_i + \epsilon_i, \ i = 1, 2$$

where μ_i are constants and ϵ_i are independently distributed normal random variables $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, which will be further referred to as shocks.

This simple factor structure accommodates a number of interpretations. The interpretation preferred here is that they are country-specific risks. This captures the intuition that aggregate economic outcomes in a country, such as GDP, are affected both by events occurring within the country i.e. domestic risks, and, to the extent that it engages in economic relations with other countries, by events occurring within those other countries i.e. foreign risks. Alternatively stated, fundamentals in a country are determined by risks that are specific to that country and risks that are specific to other countries that the country has links with. The extent to which these risks affect economic fundamentals is captured by the exposure parameters α and $1-\alpha$. For instance, the first factor f_1 can be thought of as capturing domestic risks and the second factor f_2 can be thought of as capturing foreign risks; a relatively high exposure parameter $\alpha > 0.5$ would thus describe a relatively closed economy for which domestic risks are more important, while a relatively low parameter $\alpha < 0.5$ would describe a relatively open economy that is more exposed to foreign risks.

Fundamentals are realized but unknown when the firm chooses its investment and this introduces uncertainty about the optimal level of investment. To reduce this uncertainty, the firm chooses how much information to observe about the risk factors affecting fundamentals, before investing. Reducing the uncertainty about the unobserved fundamentals will enable the firm to reduce the loss due to suboptimal investment and thus achieve a higher payoff and utility. The sequence of events is illustrated below.



The model can thus be broken down into three periods 1, 2 and 3. In the first period the representative firm chooses its information. In the second period, the firm observes the chosen information and optimally decides on a level of investment. In the third period the payoff of the investment is realized and utility is consumed. The firm's objective function is to maximize date-1 utility given by

$$U_1 \equiv E_1[E_2[\pi]] \tag{1.3}$$

where $E_i[\cdot]$ and $U_i[\cdot]$ denote the expected value and expected utility, respectively, conditional on the information available at time i.

1.2.2 Information structure

The firm takes the structure of risk factor exposures as given and decides how much to reduce uncertainty about each risk through learning. However, the firm has limited capacity to process information, meaning that its choice of how much information to observe about each risk factor is subject to a constraint on the total amount of information that it can observe.

The firm devotes limited information processing resources to learn about the factor-specific shocks affecting fundamentals. It is endowed with the prior beliefs that the shocks follow $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ and acquires noisy signals about each shock

$$s_i = \epsilon_i + \epsilon_{s_i}, \quad i = 1, 2 \tag{1.4}$$

where the signal noise is normally distributed $\epsilon_{s_i} \sim \mathcal{N}(0, \sigma_{s_i}^2)$ and uncorrelated with the other signals. The firm combines the prior beliefs with the acquired signals and forms posterior beliefs according to Bayes' law. Let $\hat{\theta}$ and $\hat{\sigma}^2$ denote the posterior mean and variance of fundamentals, respectively, conditional on the information available at time 2

$$\hat{\theta} \equiv E[\theta|s_1, s_2] = \alpha \left(\mu_1 + \frac{\sigma_{s_1}^{-2}}{\sigma_1^{-2} + \sigma_{s_1}^{-2}} s_1 \right) + (1-\alpha) \left(\mu_2 + \frac{\sigma_{s_2}^{-2}}{\sigma_2^{-2} + \sigma_{s_2}^{-2}} s_2 \right) \tag{1.5}$$

$$\hat{\sigma}^2 \equiv V[\theta|s_1, s_2] = \alpha^2 \frac{1}{\sigma_1^{-2} + \sigma_{s_1}^{-2}} + (1-\alpha)^2 \frac{1}{\sigma_2^{-2} + \sigma_{s_2}^{-2}}. \tag{1.6}$$

Denoting the factor-specific posterior variance by $\hat{\sigma}_i^2 \equiv \left(\sigma_i^{-2} + \sigma_{s_i}^{-2}\right)^{-1}$, $i = 1, 2$, the conditional mean and variance of fundamentals can be re-written as

$$\hat{\theta} \equiv E[\theta|s_1, s_2] = \alpha \left(\mu_1 + \frac{\hat{\sigma}_1^2}{\sigma_{s_1}^2} s_1 \right) + (1-\alpha) \left(\mu_2 + \frac{\hat{\sigma}_2^2}{\sigma_{s_2}^2} s_2 \right) \tag{1.7}$$

$$\hat{\sigma}^2 \equiv V[\theta|s_1, s_2] = \alpha^2 \hat{\sigma}_1^2 + (1-\alpha)^2 \hat{\sigma}_2^2. \tag{1.8}$$

The firm has limited resources or capacity to process information about the risk factors that fundamentals are exposed to. Let K denote this total capacity to process information and let k_i denote the amount of capacity devoted to learning about risk factor $i = 1, 2$. The information processing constraint can be generically formulated as

$$k_1 + k_2 \leq K. \tag{1.9}$$

Essentially, this condition tells us that the firm's choice of how much information to observe about each risk factor is subject to a constraint on the total amount of information it can observe. It also implies that for a given total capacity, learning more about one risk reduces the resources that can be devoted to learning about the other risk.

The capacity devoted to learning about a risk factor k_i, henceforth referred to as factor-specific information processing capacity, is essentially a measure of the reduction in uncertainty that can be achieved through learning. I employ the rational inattention framework proposed by Sims (2003), and model this reduction in uncertainty using tools from information theory, namely entropy and mutual information. Entropy is the standard measure of information in information theory, and it measures the amount of uncertainty in a distribution. Mutual information is the difference between the entropy of an unconditional and a conditional distribution, and it measures the amount of uncertainty resolved by conditioning on information. The factor-specific information processing capacity k_i is defined as the difference between the entropy of prior and posterior beliefs. The higher the factor-specific information processing capacity k_i, the higher the uncertainty resolved by a signal and the more informative or precise the signal is said to be. Given the assumption of normally distributed priors and signals, the factor-specific information processing capacity k_i is

$$k_i \equiv \frac{1}{2} \ln \frac{\sigma_i^2}{\hat{\sigma}_i^2}, \quad i = 1, 2 \tag{1.10}$$

where $\hat{\sigma}_i^2$ is the factor-specific posterior variance.

The entropy-based learning technology essentially imposes a bound on the product of posterior precisions.[16] This has two implications that make this information processing technology suitable for the setup considered here. First, it has a form of increasing returns to learning built into it, which means that it is less costly to learn about risk factors that are already well understood.[17] This captures the intuition that learning about risk factor exposures that are fixed, stable or sticky, and which take a long time to build or terminate, such as the cross-country real or financial linkages considered in the motivating example, is a process of refined learning. Secondly, the

[16] This follows from combining (1.9) and (1.10) and re-writing it as $\prod_{i=1}^{2} \sigma_i^2 \hat{\sigma}_i^{-2} \leq e^{2K}$. Thus, more information capacity implies a higher product of (weighted) posterior precisions $\hat{\sigma}_i^{-2}$.

[17] This is due to the fact that an increase in signal precision when prior precision is high increases the product by less (since posterior precision is the sum of prior precision and signal precision).

entropy-based learning technology accounts for the fact that the number and composition of risks affecting fundamentals are relevant for learning, which means that it is costly to learn about all the risk factors.[18] It captures the intuition that learning about one risk factor will affect the ability to learn about the other factors and it is less costly to specialize in learning about one rather than all the risk factors.

In addition to the capacity constraint, the firm also faces a no-forgetting constraint which rules out the possibility of forgetting information about one risk in order to obtain more information about another one, without violating the capacity constraint:

$$k_i \geq 0, \quad i = 1, 2. \tag{1.11}$$

This is essentially a condition that the precision of each signal must be non-negative and it captures the intuition that learning about a risk factor should not increase uncertainty i.e. posterior variance should not exceed prior variance.

1.2.3 Solving the model

Given a level of capacity K, a solution to the model is: a choice of factor-specific capacity k_i to maximize date-1 utility (1.3), subject to the capacity constraint (1.9), the no-forgetting constraint (1.11), and rational expectations about the date-2 (conditional) investment; posterior beliefs which are formed according to Bayes' law (1.5) and (1.6), given a signal about the risk factors; a choice of investment that maximizes expected utility, given the signal realization.

The model is solved using backward induction. First, given an arbitrary information choice, the firm decides the optimal investment. Then, given the optimal investment for each information choice, the firm decides the optimal information choice.

1.3 Results

In this section, I derive the equilibrium allocation of information capacity across risk factors. Then I discuss the implications of these information choices in terms of the uncertainty faced by the firm when investing. Finally, I discuss the implications in terms of the chosen level of investment and the loss due to suboptimal investment.

[18]This comes from the fact that the marginal cost of an increase in precision about one risk factor is proportional to the precision about the other risk factors (since the constraint applies to the product of factor-specific posterior precisions).

1.3.1 Optimal information choice

The date-2 problem consists of the firm choosing an investment level to maximize expected profits (1.1), while taking information choice as given

$$\max_{\lambda} U_2 \equiv E_2[\pi] = \lambda E_2[\theta] - \frac{\lambda^2}{2} \tag{1.12}$$

where $E_2[\cdot]$ denotes the expected value conditional on the information available at date 2.

The first order condition with respect to λ yields best investment response $\lambda = E_2[\theta] = \hat{\theta}$. Thus, for any given information choice, the optimal investment level is the expected level of economic fundamentals. Substituting this optimal investment choice into the objective (1.12) delivers the indirect date-2 utility of having any posterior beliefs and investing optimally

$$U_2 = \frac{(E_2[\theta])^2}{2} = \frac{\hat{\theta}^2}{2}. \tag{1.13}$$

The date-1 problem consists of choosing the optimal level of capacity devoted to learning about each risk factor to maximize the expected value of (1.13), and subject to the capacity constraint (1.9) and the no-forgetting constraint (1.11). The posterior mean belief $\hat{\theta}$ is unknown at date-1. It is a normally distributed random variable, $\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2 - \hat{\sigma}^2)$, so date-1 utility is given by

$$U_1 \equiv E_1[U_2] = \frac{E_1[\hat{\theta}^2]}{2} = \frac{E_1[\hat{\theta}]^2 + V_1[\hat{\theta}]}{2} = \frac{\theta^2 + \sigma^2 - \hat{\sigma}^2}{2}. \tag{1.14}$$

Since date-1 utility is decreasing in posterior uncertainty $\hat{\sigma}^2$ and all other terms are exogenous variables, maximizing date-1 utility in equation (1.14) is equivalent to minimizing posterior variance

$$\max_{k_1,k_2} \quad -\hat{\sigma}^2 = \alpha^2 \hat{\sigma}_1^2 + (1-\alpha)^2 \hat{\sigma}_2^2$$

$$\text{s.t. } \hat{\sigma}_i^2 = \sigma_i^2 e^{-2k_i} \quad \text{and} \quad \sum_{i=1}^{2} k_i \leq K \quad \text{and} \quad 0 \leq k_i, \quad i = 1, 2.$$

The unique solution to this problem delivers the following optimal factor-specific capacity allocation

$$k_1 = \begin{cases} 0 & \text{if } \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} < e^{-K} \\ \frac{1}{2}\left(K + \ln \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right) & \text{if } e^{-K} \leq \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} \leq e^{K} \\ K & \text{if } \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} > e^{K} \end{cases} \tag{1.15}$$

and $k_2 = K - k_1$.

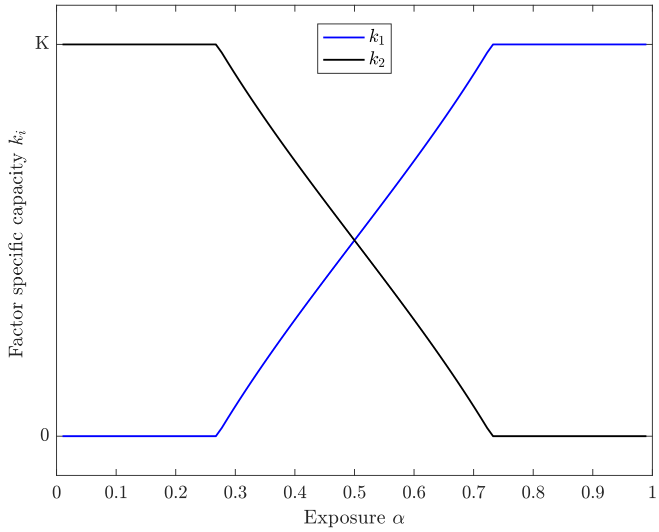


Figure 1.1 Optimal Information Choice.
The figure illustrates the relationship between the degree of exposure to factor 1, α, and the optimal level of information processing capacity allocated to factor 1, k_1, and factor 2, k_2. The parameter values are $\sigma_1 = \sigma_2 = 1$ and $K = 1$

At the interior optimum, the optimal level of capacity devoted to learning about a risk factor increases with factor-specific exposure, factor-specific prior uncertainty and total information processing capacity. In other words, the firm will optimally choose to learn more about the risk factors that fundamentals are more exposed to and which are ex-ante more uncertain. Note that corner solutions are possible. For a given level of capacity, no (all) capacity is allocated to a risk factor if factor-specific exposure and prior uncertainty are low (high) enough relative to the other factor.

The higher the capacity to process information K the smaller the range of (exposure and uncertainty) parameter values for which corner solutions are obtained, in line with the intuition that less capacity constrained agents are able to attend to more sources of risk. Note that for any limited information processing capacity, the firm will stop learning about one of the risk factor exposures. In other words, for any finite level of capacity, $\forall\ K < \infty$, there exists a level of exposure $0 < \alpha < 1$ for which the conditions in (1.15) hold and corner solutions are obtained.

Figure 1.1 plots the optimal level of information processing capacity allocated to the two factors against exposure to factor 1. If exposure to factor 1 is very low, then the firm optimally chooses to pay no attention to it and instead devotes all information processing resources to factor 2. The economic intuition is that when exposure to factor 1 is very low, the marginal benefit of learning about factor 1 is lower than the benefit of learning about factor 2, whose exposure is relatively higher. Consequently, the firm would like to forget information about factor 1 in order to obtain more information about factor 2 but the no-forgetting constraint prevents it from doing so and the zero corner solution is obtained. A similar reasoning is applied when exposure to factor 1 is very high, leading to the full capacity corner solution. At the interior optimum, factor-specific capacity increases with factor-specific exposure and the firm optimally learns more about the risk factor fundamentals are relatively more exposed to.

1.3.2 Implications for uncertainty

Information choice is a tool to reduce the uncertainty that the firm faces when deciding on the level of investment. Relative to an equal capacity, exogenous information benchmark, uncertainty is lower when the firm can optimally choose what risk factors to learn about. Importantly, uncertainty is convex in exposure when the firm is exogenously endowed with information, but it is concave in exposure when the firm optimally learns about risk factor exposures.

Given the optimal information processing capacity allocated to each factor in (1.15), uncertainty can be backed out using results (1.10) and (1.8), and it is given by

$$\hat{\sigma}^2 = \begin{cases} \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2 e^{-2K} & \text{if } \quad \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} < e^{-K} \\ 2\alpha(1-\alpha)\sigma_1\sigma_2 e^{-K} & \text{if } \quad e^{-K} \leq \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} \leq e^{K} \\ \alpha^2\sigma_1^2 e^{-2K} + (1-\alpha)^2\sigma_2^2 & \text{if } \quad \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} > e^{K} \end{cases} \tag{1.16}$$

The uncertainty expressions for the corner solutions reflect the intuition that when there is no learning about a factor its factor-specific posterior uncertainty is equal to its prior uncertainty ($\hat{\sigma}_i^2 = \sigma_i^2$), while the uncertainty about the other factor is reduced in proportion to the total capacity ($\hat{\sigma}_i^2 = \sigma_i^2 e^{-2K}$). At the interior optimum, uncertainty increases with the attention grabbing attributes of the two risk factors, namely exposure and prior uncertainty, and decreases with total capacity.

In order to assess the implications of learning for uncertainty, and thus conditional investment, a suitable benchmark is needed for comparison. I consider as benchmark a model in which the firm is endowed with an exogenous amount of information which is equally allocated among the risk factors. This model will be further referred to as the exogenous information benchmark. The firm in this model has the same priors as the endogenous learning firm in my model but it is now endowed with signals with exogenous noise $\tilde{\epsilon}_{s_i} \sim \mathcal{N}(0, \tilde{\sigma}_{s_i}^2),\ i = 1, 2$. Posterior beliefs are formed according to Bayes' rule, such that benchmark uncertainty is

$$\tilde{\sigma}_B^2 = \alpha^2\tilde{\sigma}_1^2 + (1-\alpha)^2\tilde{\sigma}_2^2 \tag{1.17}$$

where $\tilde{\sigma}_i^2 \equiv \left(\sigma_i^{-2} + \tilde{\sigma}_{s_i}^{-2}\right)^{-1},\ i = 1, 2$.

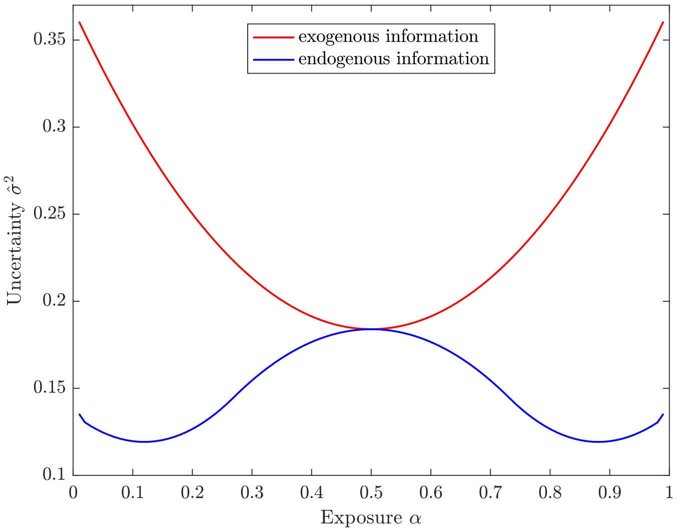


Figure 1.2 Implications for Uncertainty.
The figure illustrates the relationship between the degree of exposure to factor 1, α, and the uncertainty faced by the firm that is implied by the exogenous information model (red line) and the endogenous information model (blue line). The parameter values are $\sigma_1 = \sigma_2 = 1$, $\tilde{\sigma}_{s_1} = \tilde{\sigma}_{s_2} = 0.75$ and the total capacity implied by these parameters is $K = \frac{1}{2}\ln\sigma_1^2\left(\sigma_1^{-2} + \tilde{\sigma}_{s_1}^{-2}\right) + \frac{1}{2}\ln\sigma_1^2\left(\sigma_2^{-2} + \tilde{\sigma}_{s_2}^{-2}\right) = 1$.

Figure 1.2 depicts the relationship between the degree of exposure to a risk factor and the uncertainty implied by the endogenous information model $\hat{\sigma}^2$ (blue line) and the exogenous information benchmark model $\tilde{\sigma}^2$ (red line). This exercise is informative of the uncertainty faced by firms operating in economies whose fundamentals share the same factor structure but vary in the degree of exposure to a risk factor (in this case the exposure to factor 1, measured by α). To enable meaningful comparisons, the total capacity in the endogeous learning model is set equal to the capacity implied by exogenous signal precisions in the benchmark model when the capacity constraint is binding i.e. $\frac{1}{2}\ln\frac{\sigma_1^2}{\tilde{\sigma}_1^2} + \frac{1}{2}\ln\frac{\sigma_2^2}{\tilde{\sigma}_2^2} = K$. This ensures that the two models are otherwise identical except for the ability to optimally reallocate information processing resources. Note that Figure 1.2 illustrates a symmetric equilibrium whereby the factors are ex-ante equally risky and the exogenous signals are equally informative, hence the symmetry around and intersection of the two lines at the $\alpha = 0.5$ exposure midpoint.

In terms of levels, note that relative to the equal capacity, exogenous information benchmark, uncertainty is lower when the firm can optimally allocate information resources across risk factor exposures. This is because learning effectively reduces the uncertainty about individual factors. The reduction in uncertainty achieved through learning operates through what will be further referred to as the *information channel.* Recall that the endogenous learning firm optimally learns more about the risk factor fundamentals are relatively more exposed to. The regions at the left and right of the $\alpha = 0.5$ midpoint depict situations of relatively higher exposure to a factor whose effective or learning-adjusted uncertainty is lower under the endogenous learning model than under the benchmark. Consequently, for any given level of exposure in these regions, overall uncertainty will be higher under the exogenous information model than under the endogenous information model. Appendix **A.1** provides an analytical proof of this intuition.

In terms of dynamics, note that both models illustrates a non-monotonic relationship between overall uncertainty and the degree of exposure to a risk factor. However, whereas in the benchmark model uncertainty is convex in exposure, in the endogenous learning model uncertainty is concave in exposure when the firm optimally learns about both risk factors i.e. at the interior optimum. Alternatively stated, uncertainty increases with exposure to relatively important risk factors when information is exogenously given, but it decreases with exposure when information about the two risks is endogenously chosen. Thus, one implication of the endogenous learning model is that in the cross-section of entities exposed to a relatively important risk factor, an entity that is more exposed to that risk will have a better understanding of it and will thus face a lower uncertainty than an entity which is less exposed to it.

Proposition 1. *Provided that a risk is relatively important, uncertainty decreases with exposure when the firm optimally learns i.e.* $\frac{\partial \hat{\sigma}^2}{\partial \alpha} < 0$ *if* $\alpha > 0.5$ *and* $k \in (0, K)$.

Proof. See Appendix **A.1**.

Proposition 1 sheds light on why the shock transmission patterns observed during the financial crisis of 2007-2008 were different relative to those observed during past contagious crises such as the Mexican crisis of 1994 and the Asian crisis of 1997.[19] It is because in the last crisis the shock originated in a country that plays a central role in the global economy: the United States is both a country that is relatively important for most other countries ($\alpha > 0.5$), as well as a country that foreign decision-makers are likely to learn about ($k \in (0, K)$). Thus, variation in the extent to which other countries were exposed to the United States, translates into variation in uncertainty which follows the dynamics illustrated at the right of the $\alpha = 0.5$ exposure midpoint in Figure 1.2. In other words, my model suggests that countries which were relatively more exposed to the United States shock faced a relatively lower level of uncertainty regarding the implications of the shock.

[19] During these past crises the epicentre country in which the shock originated was more severely affected than the other countries that were subsequently affected by the shock. More generally, the impact of the original shock increased with exposure to it during the Mexican and Asian crises, but it decreased with exposure during the last financial crisis of 2007-2008.

This provides a potential explanation for why during the financial crisis of 2007-2008, unlike during previous crises, the transmission of the shock decreased with exposure to the shock.

The benchmark model illustrates a diversification effect whereby high exposure to a risk factor implies high overall uncertainty, and the lowest level of uncertainty is achieved when exposure to the two factors is equal. To understand the forces that are at play let us focus on the right part of Figure 1.2, where exposure to factor 1 is relatively higher i.e. $\alpha > 0.5$. Given two equally risky factors, which is the case in light of the symmetric equilibrium considered, overall uncertainty is driven by the factor that is important in terms of exposure. Consequently, increasing exposure to factor 1 beyond the 0.5 midpoint results in an increase in overall uncertainty. This works through what will be further referred to as the *exposure channel*. The idea behind it is that the overall risk of a bundle of factors which are equally risky will be driven by those that are important in terms of exposure; in other words, for fixed risk overall dynamics are dictated by exposure.

The endogenous learning model, on the other hand, predicts that uncertainty is convex in exposure when the firm optimally learns about one factor only, and it is concave in exposure when the firm learns about both factors.[20] In the corners, when the firm devotes all capacity to learning about one factor, information is essentially exogenous, so uncertainty dynamics will be the same as in the benchmark and will operate through the exposure channel. At the interior optimum though, uncertainty initially increases with exposure, but once the factor becomes important in terms of exposure, uncertainty will start to decrease. To understand the mechanism behind this, let us start from the situation in which exposure to the two factors is equal i.e. $\alpha = 0.5$. As exposure to factor 1 is increased beyond this 0.5 midpoint, the uncertainty about factor 1 is effectively reduced through learning. The reduction in uncertainty entailed by learning is stronger than the increase in exposure, so increasing exposure to factor 1 results in a decrease in overall uncertainty. In other words, the information channel dominates the exposure channel. However, the benefits of learning are limited and increasing exposure beyond a certain point will undo the reduction in uncertainty achieved through learning, resulting in an increase in overall uncertainty. In other words, the exposure channel will overturn the information channel if exposure is high enough and the firm only learns about one factor. This is reflected in the turning point in the uncertainty, which occurs when the learning-adjusted risk exposures of the two factors are equal.[21]

[20] Note that uncertainty in the endogenous information model is convex on the same interval of exposure parameters over which the factor-specific capacity allocation illustrated in Figure 1.1 is extreme i.e. either zero or maximum.

[21] At the first corner (when learning about factor 2) uncertainty decreases if $\alpha\sigma_1^2 < (1-\alpha)\sigma_2^2 e^{-2K}$, which essentially reads that increasing exposure to a factor whose learning-adjusted risk is relatively lower decreases overall uncertainty. At the second corner (when learning about factor 1), uncertainty starts to increase when $\alpha\sigma_1^2 e^{-2K} > (1-\alpha)\sigma_2^2$ i.e. when learning-adjusted risk is higher.

In sum, the benchmark illustrates a classic diversification effect which operates through the exposure channel. In the endogenous learning model, learning substitutes for diversification in reducing risk, resulting in a specialization effect. In a portfolio allocation context, Van Nieuwerburgh and Veldkamp (2010) also make the point that diversification is not optimal if portfolio choice is preceded by information choice. While they focus on the implications of information choice in terms of portfolio holdings when decision-makers choose both their information about and exposure to risks, I focus on the role of information in terms of the impact of shocks when decision-makers choose their information about risks they are exogenously exposed to. Section 1.4.4 relaxes the assumption of exogenous exposures to risk factors and provides a characterization of the optimal exposure points.

1.3.3 Implications for investment

Reducing uncertainty will enable the firm to take an investment decision that is more aligned with underlying fundamentals and thus reduce the loss due to suboptimal investment.[22] Insofar as the firm's investment deviates from the first-best optimum obtained under perfect information, it contributes to the transmission of shocks. The basic result is that the transmission of shocks decreases with exposure when the firm optimally learns about the shocks affecting fundamentals. Consequently, relative to the exogenous learning benchmark, the impact of shocks that fundamentals are relatively less exposed to is amplified, while the impact of shocks that fundamentals are relatively more exposed to is attenuated.

Recall that the firm optimally chooses a level of investment that is equal to the expected level of fundamentals conditional on the information available at the intermediate date 2 i.e. $\lambda = E_2[\theta] = \hat{\theta}$. Thus, the investment decision is essentially a response to information about the realization of shocks affecting fundamentals. Figure 1.3 depicts the relationship between the degree of exposure to a shock and the response or sensitivity of investment to that shock that is implied by the exogenous information benchmark (red line), the endogenous information model (blue line) as well as a full information model (black line).

Note that the shock sensitivity increases with exposure under all the three models considered but the rate of increase is different. Under the full information model shocks can be perfectly observed so there is a one to one mapping between exposure to the shock and the sensitivity to it. This is the optimal or first-best response; to the extent that responses deviate or are not

[22]Mathematically a lower level of uncertainty means higher response to signals that are informative about the underlying shocks affecting fundamentals.

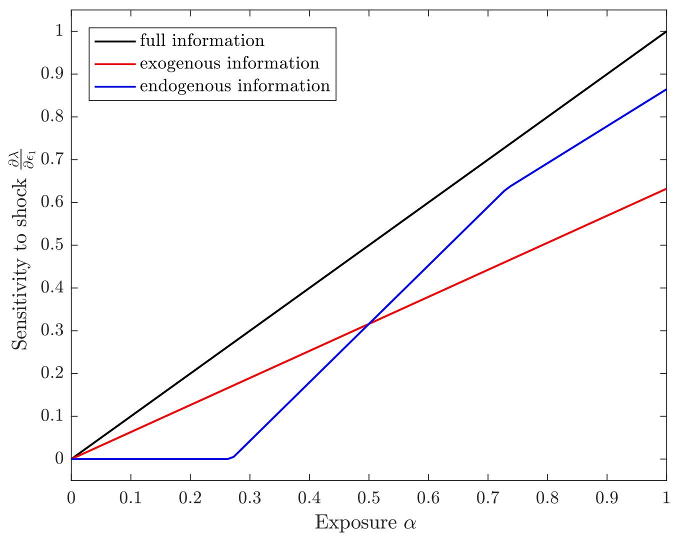


Figure 1.3 Sensitivity of Investment to Shocks.
The figure illustrates the relationship between the degree of exposure to factor 1, α, and the sensitivity of investment to a shock to factor 1 that is implied by the exogenous information model (red line), the endogenous information model (blue line) and a perfect information model (black line). The parameter values are $\sigma_1 = \sigma_2 = 1$, $\tilde{\sigma}_{s_1} = \tilde{\sigma}_{s_2} = 0.75$ and $K = 1$.

aligned with it they are said to be suboptimal and to act as a shock transmission mechanism. Under exogenous information, the response departs from the optimal one as exposure increases because shocks are observed with a precision that is fixed so the exposure effect dominates. Under the endogenous info model three situations can be observed: for sufficiently low exposure, the firm does not acquire any information and there is no response to shocks; shock sensitivity is zero because as the firm is essentially unaware of the underlying shock realization. Second, as exposure increases and the firm starts learning about factor 1, its response to shocks will become increasingly more aligned with the optimal one because the precision of information about shocks affecting fundamentals optimally increases with exposure. Finally, when exposure is sufficiently high that the firm only learns about factor 1, the response starts to decrease again with exposure because information precision, albeit set at the maximum, is essentially exogenous.

Thus, information frictions introduce a loss or inefficiency relative to the full information benchmark. When shocks can be perfectly observed they affect investment in proportion to exposure. However, when shocks cannot be perfectly observed investment decisions are relatively less responsive to shocks; the firm fails to fully incorporate underlying shocks into decision-making. In case of positive shocks, the deviation from the full information optimum is negative and represents a situation of underinvestment. The firm increases investment by less than it optimally should, resulting in missed or lost business opportunities. In case of negative shocks the deviation is positive and decreases with exposure. This is a situation of overinvestment, whereby the firm reduces investment by less than it optimally should, resulting in wasted resources through excess capacity. In order to ease analysis and abstract from the nature of shocks, I define the loss due to suboptimal action as $L \equiv (\hat{\theta} - \theta)^2$. This symmetric loss function captures a more general situation in which shocks are changes in circumstances that the firm needs to adapt to rather than good or bad events, such as the introduction of standards, technological disruption, terms of trade changes.

Figure 1.4 plots the loss due to suboptimal investment against exposure to factor 1. The example considers a one standard deviation shock to factor 1, abstracts from factor 2 shocks as well as from information shocks. Under the exogenous information benchmark, loss increases monotonically with exposure to the shock (red line). Given that signal noise is fixed, this result works through the exposure channel and is due to increasing exposure to a risk that is constant. Under the endogenous learning model, three scenarios can be observed as exposure increases (blue line). First, when exposure is sufficiently low such that the firm does not learn about factor 1, the loss due to suboptimal investment is increasing in exposure. The rate of increase is higher relative to the exogenous information benchmark because the firm chooses to observe no information, as opposed to fixed information, about factor 1. Second, as exposure increases and the firm starts learning about factor 1, the loss due to suboptimal investment decreases with exposure because the firm is able to incorporate shocks more accurately into decision-making. Third, when the firm learns only about factor 1, the loss due to suboptimal investment starts to increase again because information precision is essentially exogenous and as a consequence dynamics resemble the benchmark.

Worth emphasizing is the fact that the predictions of the two models are starkly different when an interior solution is obtained for endogenous information choice i.e. when the firm learns about both risk factors. Whereas the loss due to suboptimal action is increasing with exposure when the firm is exogenously endowed with information about the risks, it is decreasing with exposure when the firm optimally chooses what risks to learn about. Consequently, relative to the exogenous learning benchmark, shocks that fundamentals are relatively more exposed to are attenuated (lower loss) while shocks that fundamentals are relatively less exposed to are amplified (higher loss). The magnitude of amplification decreases with the firm's capacity to learn. In other words, amplification is more severe for more capacity constrained firms that are only able to learn about one risk factor.

If the shock to factor 1 is interpreted as a shock to United States, Figure 1.4 implies that a country which is less exposed to the United States (which is situated towards the left end of

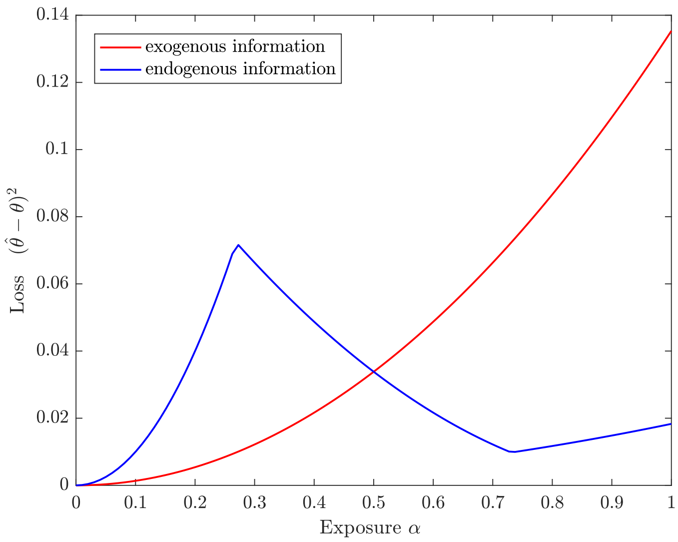


Figure 1.4 Loss Due to Suboptimal Investment

The figure illustrates the relationship between the degree of exposure to factor 1, α, and the loss due to suboptimal action that is induced by a shock to factor 1. This example considers a one standard deviation shock to factor 1 i.e. $\epsilon_1 = \sigma_1$, abstracts from factor 2 shocks i.e. $\epsilon_2 = 0$ and information shocks i.e. $\epsilon_{s_1} = \epsilon_{s_2} = 0$. The parameter values are $\mu_1 = \mu_2 = 1$, $\sigma_1 = \sigma_2 = 1$ and $\tilde{\sigma}_{s_1} = \tilde{\sigma}_{s_2} = 0.75$ and the total capacity implied by these parameters is $K = \frac{1}{2} \ln \sigma_1^2 \left(\sigma_1^{-2} + \tilde{\sigma}_{s_1}^{-2}\right) + \frac{1}{2} \ln \sigma_1^2 \left(\sigma_2^{-2} + \tilde{\sigma}_{s_2}^{-2}\right) = 1$.

the x-axis) will incur a higher loss due to suboptimal action compared to a country which is more exposed to the United States (which is situated towards the right end of the x-axis). It also implies that a country which is less exposed to the United States will incur a higher loss due to suboptimal action compared to the United States itself (which is likely to be situated towards the right end of the x-axis). These predictions are in line with evidence from the GFC that countries other than the epicentre have been more severely affected than United States itself, and more generally countries that were less exposed to the United States were more affected.

The results depicted in Figure 1.4 can be understood by examining how investment decisions act as a shock transmission mechanism analytically. To that end, I define the shock transmission mechanism as the change in the loss function due to a shock i.e. $\frac{\partial L}{\partial \epsilon_i}$. The shock transmission mechanism essentially measures the extent to which shocks translate into losses. The interaction between the magnitude of the shock and the strength of the shock transmission mechanism determines the impact of a shock. Consequently, statements about the impact of a shock amount to statements about the negative consequences of a shock or the losses induced by it. The stronger the shock transmission mechanism, the higher the loss due to suboptimal investment and the higher the impact of the shock is said to be.

Proposition 2. *The transmission of shocks decreases with exposure when the firm optimally learns i.e.* $\frac{\partial^2 L}{\partial \epsilon_1 \partial \alpha} < 0$ *if* $k \in (0, K)$.

Proof. See Appendix **A.2**.

Proposition 2 pins down the mechanism through which the impact of a shock can decrease with exposure to it. When the firm optimally learns about the risk factors affecting fundamentals, the loss due to suboptimal investment that is induced by a shock decreases with exposure. In other words, the contribution of actions to the transmission of shocks decreases with exposure. This is because the reduction in uncertainty that is achieved through learning increases with exposure, and as a consequence the deviation of the firm's investment from the perfect information optimum decreases. Thus, the informational benefit mitigates the direct impact of shocks to risk factors that fundamentals are relatively more exposed to as it enables the firm to take better informed decisions and minimize the loss due to suboptimal investment. By the same token, the impact of shocks to risk factors that fundamentals are relatively less exposed to is amplified through the firm's poorly informed investment decision as it implies a higher loss due to suboptimal investment.

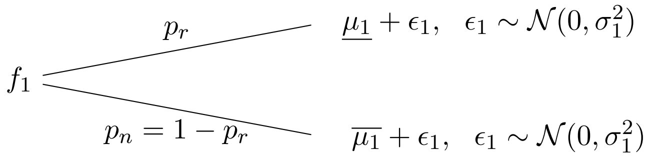
1.4 Extensions

In this section, I consider the following extensions to the baseline model studied so far. First, I account for the degree of anticipation of shocks. Second, I allow for strategic complementarity in investment. Third, I relax the assumption that the risk factors affecting fundamentals are independent. Finally, I relax the assumption that exposures to the risk factors are exogenous.

1.4.1 Extension: shock anticipation

The baseline model considers the case in which the firm allocates an exogenously given information processing capacity K across risk factor exposures. This section endogenizes the capacity available to the firm in a certain state of nature, by linking it to the degree of anticipation of that state. The basic result is that the impact of unanticipated shocks is amplified because the firm optimally allocates less information processing capacity to learning about low probability states of nature.

One of the risk factors affecting fundamentals is assumed to be in one of two possible states of nature: a low-probability, low-mean so-called bad state of nature - interpreted as rare times, and a high-probability, high-mean so-called good state of nature - interpreted as normal times. Let $p_r > 0$ denote the probability that factor 1 is in the bad state of nature. If it is ex-ante unlikely for this state to occur, then p_r will be small, and a rare event or crisis is said to occur if the state is revealed to be bad. This setup can thus be graphed as



where $\underline{\mu_1} < \overline{\mu_1}$ and $p_r < p_n$. To isolate the effect of the degree of anticipation alone, I assume that only the mean level of the risk factors is different in the two states while the priors associated with the shocks affecting the risk factors in the two states are the same.[23]

In this setup, there are two dimensions of information choice: how much information to acquire about a state of the world, and how to allocate that information across risk factors in each state of the world.[24] Let $\mathcal{K}$ denote the *total capacity* or total amount of information resources available to the firm , K_s the *state capacity* or amount of information resources dedicated to state of nature $s \in \{n, r\}$, and k_{is} denote the amount of information processing capacity dedicated to factor i in state $s \in \{n, r\}$. The capacity constraint (1.9) governing the allocation of capacity across risk factors in any state of nature $s \in \{n, r\}$ can now be formulated as

$$k_{1s} + k_{2s} \leq K_s, \quad s \in \{n, r\},$$

and the capacity constraint governing the allocation of total capacity across states of nature is given by

$$K_n + K_r \leq \mathcal{K}.$$

The model is solved following the same steps as in the baseline model in Section **??**, except that now the date-1 problem consists of two steps. As before, the firm first allocates information resources across risk factor exposures given the optimal investment level and an arbitrary state capacity. The second step is to allocate total information resources across states of nature given the optimal investment and optimal capacity allocation across exposures in any state. Let $U_1(K_s)$ denote the date-1 utility of investing optimally at the second date and optimally allocating the available state capacity K_s across risk factors at the first date

[23]In any state of nature, the capacity allocation across risk factors is mean independent but increases with the prior uncertainty surrounding that factor i.e. volatility. Accounting for the intuition that crises episodes are characterized by high volatility would weaken the ensuing result but the main result still holds. This case is dealt with in Appendix **A.3**.

[24]This can be thought of as capturing situations in which decision-makers prepare for different contingencies.

$$U_1(K_s) = \begin{cases} -\alpha^2\sigma_1^2 - (1-\alpha)^2\sigma_2^2 e^{-2K_s} & \text{if } k_1 = 0 \\ -2\alpha(1-\alpha)\sigma_1\sigma_2 e^{-K_s} & \text{if } k_1 = \frac{1}{2}\left(K_s + \ln\frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right) \\ -\alpha^2\sigma_1^2 e^{-2K_s} - (1-\alpha)^2\sigma_2^2 & \text{if } k_1 = K_s \end{cases} \tag{1.18}$$

The date-1 problem for the allocation of total capacity across states is

$$\max_{K_n, K_r} \; p_n U_1(K_n) + p_r U_1(K_r) \tag{1.19}$$

$$\text{s.t. } K_n + K_r \leq \mathcal{K} \text{ and } K_s \geq 0, \; s \in \{n, r\} \tag{1.20}$$

The basic result is that the optimal level of information processing capacity dedicated to a state of nature increases with the probability of occurrence of the state. Appendix **A.3** provides a full characterization of the equilibria. At the interior optimum, when the firm learns about both factors, the optimal level of capacity dedicated to the rare state is

$$K_r = \begin{cases} 0 & \text{if } \frac{p_r}{p_n} < e^{-\mathcal{K}} \\ \frac{1}{2}\left(\mathcal{K} + \ln\frac{p_r}{p_n}\right) & \text{if } \frac{p_r}{p_n} \geq e^{-\mathcal{K}} \end{cases} \tag{1.21}$$

Note that a corner solution is possible. More specifically, if the probability of the state of nature is low enough, no capacity is allocated to the state. Otherwise, the information-processing capacity allocated to a state of nature increases with its degree of anticipation, as well as with the total capacity $\mathcal{K}$ available.

The implication of this capacity allocation in terms of state contingent investment schedules is that the impact of shocks decreases with their degree of anticipation because the loss due to suboptimal investment is higher the more unanticipated the shock. The firm optimally devotes little attention to low-probability events. Thus, the contribution of investment decisions to the transmission of shocks is higher the lower their ex-ante probability of occurrence. In other words, the deviation of the firm's investment from the perfect information optimum will be higher the lower the probability of occurrence of a shock, which essentially intensifies its transmission. The amplification of unanticipated shocks will thus be higher.

Proposition 3. *The transmission of shocks decreases with their degree of anticipation when the firm optimally learns i.e.* $\frac{\partial^2 L}{\partial \epsilon_1 \partial p_s} < 0$ *if* $k_{1s} \in (0, K_s)$, $s \in \{n, r\}$.
Proof. See Appendix **A.3**.

Proposition 3 essentially says that the contagious transmission of crises is higher the lower their probability of occurrence. This prediction is in line with evidence that documents a negative relation between the degree of anticipation of crises and the occurrence of contagion (Kaminsky et al., 2003; Rigobon and Wei, 2003; Didier, Mauro, and Schmukler, 2008; Mondria and Quintana-Domeque, 2013). Additionally, it is related to the debates on whether the highly unexpected nature of the Lehman shock might have amplified its transmission. My model predicts that contagion is more likely to occur following unexpected crises because decision-makers optimally prepare less for unexpected states of the nature.

Figure 1.5 plots the loss due to suboptimal investment against exposure to a one standard deviation shock to factor 1, for varying degrees of anticipation of the shock. The loss due to suboptimal investment is larger when the shock occurs with a small probability (solid lines), relative to the case in which the shock occurs with a higher probability (dashed lines). The loss due to suboptimal investment that is induced by a shock decreases with the degree of anticipation of that shock because the firm optimally learns less about less anticipated states of nature. Thus, endogenizing the information processing capacity available in a state of nature has the effect of amplifying the impact of shocks occurring in unexpected states; decision-makers are unprepared for unexpected shocks and this amplifies their consequences.

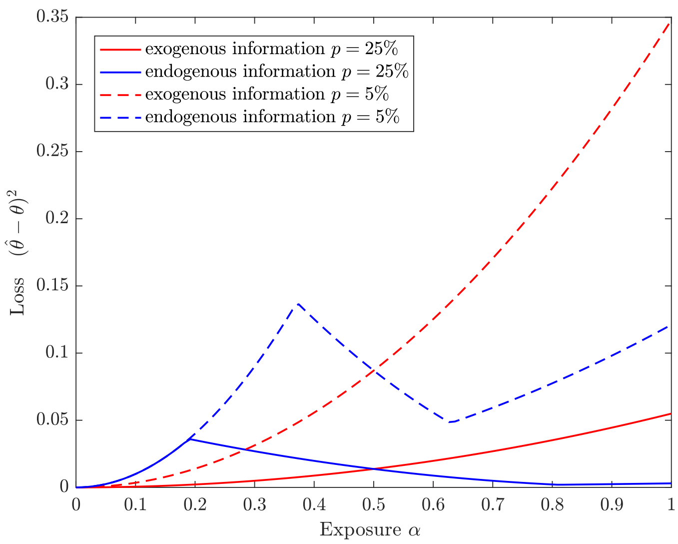


Figure 1.5 Loss Due to Suboptimal Investment for Varying Degrees of Shock Anticipation.

The figure illustrates the relationship between the degree of exposure to factor 1, α, and the loss due to suboptimal action that is induced by a shock to factor 1, for varying degrees of anticipation of the shock. This example considers a one standard deviation shock to factor 1 i.e. $\epsilon_1 = \sigma_1$, abstracts from factor 2 shocks i.e. $\epsilon_2 = 0$ and information shocks i.e. $\epsilon_{s_1} = \epsilon_{s_2} = 0$. The parameter values are $\sigma_1 = \sigma_2 = 1$; when $p = 25\%$ $\tilde{\sigma}_{s_1} = \tilde{\sigma}_{s_2} = 0.55$ and $K = 1.45$, and when $p = 5\%$ $\tilde{\sigma}_{s_1} = \tilde{\sigma}_{s_2} = 1.20$ and $K = 0.53$.

1.4.2 Extension: investment complementarities

This subsection extends the baseline model to account for strategic interactions and their implications for equilibrium information choice and the loss due to suboptimal investment. Relative to the baseline model, strategic complementarity in investment reduces the incentive to learn about risks fundamentals have relatively low exposure to. Thus, the transmission of shocks that fundamentals are relatively less exposed to is intensified when firms seek to coordinate their investment decisions, which implies that the loss due to suboptimal investment is higher relative to the baseline.

There is a continuum of firms indexed by j. Each firm chooses level of investment λ_j to maximize expected profits. The profit function for firm j is given by

$$\pi_j = R\lambda_j - \frac{1}{2}\lambda_j^2 \tag{1.22}$$

The return on investment R is a function of the unknown fundamentals in the economy θ and the average investment in the population $\bar{\lambda} = \int_j \lambda_j$

$$R = (1-r)\theta + r\bar{\lambda}$$

where r is a constant governing the type of strategic interactions between firms. Real investment environments have typically been treated as being characterized by strategic complementarities. In such environments decision-makers want to do what others do. This is modelled by imposing that $r > 0$, which implies that optimal individual responses λ_j increase in the actions of others $\bar{\lambda}$. If $r = 0$ individual actions are independent of the average action in the population and the baseline model is obtained.

As in the baseline model, the solution strategy is to work backwards. At date-2, each firm undertakes a level of investment to maximize the expected profit π_j in (1.22), while taking information choice as given. The objective function is formulated as

$$\max_{\lambda_j} U_{2j} \equiv E_{2j}[R\lambda_j - \frac{1}{2}\lambda_j^2].$$

The first-order condition yields

$$\lambda_j = E_{2j}[R] = (1-r)E_{2j}[\theta] + rE_{2j}[\bar{\lambda}]. \tag{1.23}$$

I consider equilibria in which the mean investment in the population is a linear function of the shocks affecting fundamentals

$$\bar{\lambda} = \psi + \phi_1 \epsilon_1 + \phi_2 \epsilon_2 \tag{1.24}$$

where ψ, ϕ_1 and ϕ_2 are constants determined in equilibrium. Recalling that $E_{2j}[\theta] = \alpha(\mu_1 + E_{2j}[\epsilon_1]) + (1-\alpha)(\mu_2 + E_{2j}[\epsilon_2])$ and substituting conjecture (1.24) into the first order condition (1.23) yields

$$\lambda_j = r\psi + (1-r)(\alpha\mu_1 + (1-\alpha)\mu_2) + [\alpha(1-r) + r\phi_1]E_2[\epsilon_1] + [(1-\alpha)(1-r) + r\phi_2]E_2[\epsilon_2].$$

Calculating first the conditional expectation of the shocks $E_{ij}[\epsilon_i] = (1-\gamma_i)s_{ij}$, and then the mean action in the population yields

$$\bar{\lambda} = r\psi + (1-r)(\alpha\mu_1 + (1-\alpha)\mu_2) + [\alpha(1-r) + r\phi_1](1-\gamma_1)\epsilon_1 + [(1-\alpha)(1-r) + r\phi_2](1-\gamma_2)\epsilon_2.$$

Matching coefficients verifies the conjecture (1.24) that the average investment level is linear in the shocks. The linear weights are given by $\psi = r\psi + (1-r)(\alpha\mu_1 + (1-\alpha)\mu_2)$, $\phi_1 = [\alpha(1-r) + r\phi_1](1-\gamma_1)$ and $\phi_2 = [(1-\alpha)(1-r) + r\phi_2](1-\gamma_2)$. Collecting the unknown coefficients yields

$$\psi = \alpha\mu_1 + (1-\alpha)\mu_2 \tag{1.25}$$

$$\phi_1 = \frac{\alpha(1-r)(1-\gamma_1)}{1-r(1-\gamma_1)} \tag{1.26}$$

$$\phi_2 = \frac{(1-\alpha)(1-r)(1-\gamma_2)}{1-r(1-\gamma_2)}. \tag{1.27}$$

The date-1 problem consists of choosing the optimal level of capacity devoted to learning about each risk factor to maximize expected utility implied by the investment rule (1.24) and the equilibrium coefficients (1.25)-(1.27), subject to the capacity constraint (1.9) and the no-forgetting constraint (1.11)

$$\max_{k_1,k_2} \; U_{1j} = E_{1j}[U_{2j}] = \frac{1}{2}\Big[\psi^2 + \frac{\alpha^2(1-r)^2\sigma_1^2(1-\gamma_1)}{(1-r(1-\gamma_1))^2} + \frac{(1-\alpha)^2(1-r)^2\sigma_2^2(1-\gamma_2)}{(1-r(1-\gamma_2))^2}\Big]$$

$$\text{s.t.} \quad \gamma_i = e^{-2k_i}, \quad \sum k_i \leq K, \quad 0 \leq k_i, \quad i = 1, 2.$$

Numerical results indicate that relative to the baseline model ($r = 0$), the firm is more likely to learn about a single risk rather than both risks when investment actions are strategic complements ($r > 0$). In other words, as the degree of strategic complementarities increases, corner solutions occur more easily. In fact, if the degree of strategic complementarity r is high enough the parameter region for which an interior solution is obtained collapses to a single point.[25] The implication is that for high levels of strategic complementarity, a small change in the exposure parameter can have a large effect on the equilibrium allocation of attention.

At the interior optimum, when the firm optimally learns about both risks, the optimum level of capacity allocated to factor 1 decreases with the degree of complementarity r if exposure to factor 1 is relatively low ($\alpha < 0.5$) but it increases if exposure to factor 1 is relatively high ($\alpha > 0.5$). This is because the firm anticipates that learning optimally increases with exposure in the population, and as a consequence it also chooses to learn less about low-exposure risk factor and more about high-exposure risk factors. In other words, strategic complementarity in investment reduces the incentive to learn about risk factors that fundamentals have relatively low exposure to, which implies that the transmission of shocks to these risk factors will be amplified compared to the baseline model. Consequently, the contagious transmission of shocks is intensified as the degree of strategic complementarity increases.

The implications in terms of shock impact are illustrated in Figure 1.6, which plots the loss due to suboptimal investment against exposure to a one standard deviation shock to factor 1, for varying levels of strategic complementarity. The loss due to suboptimal investment is larger in environments characterized by a higher level of strategic complementarity (dashed lines), relative to the case in which there are no strategic interactions (solid lines). This is due to the fact that the firm's incentive to hedge against shocks through learning is weakened by its desire to coordinate its investment decison with the average investment in the population. Since the firm anticipates other firms will optimally choose to learn less about the risk factors that fundamentals have relatively little exposure to, its incentive to learn about these low-exposure risk factors decreases as the degree of strategic complementarity increases. As a consequence, the loss due to suboptimal action that is induced by a shock is amplified in the presence of strategic complementarities.

[25] For strategic complementary parameters beyond this point multiple equilibria exist.

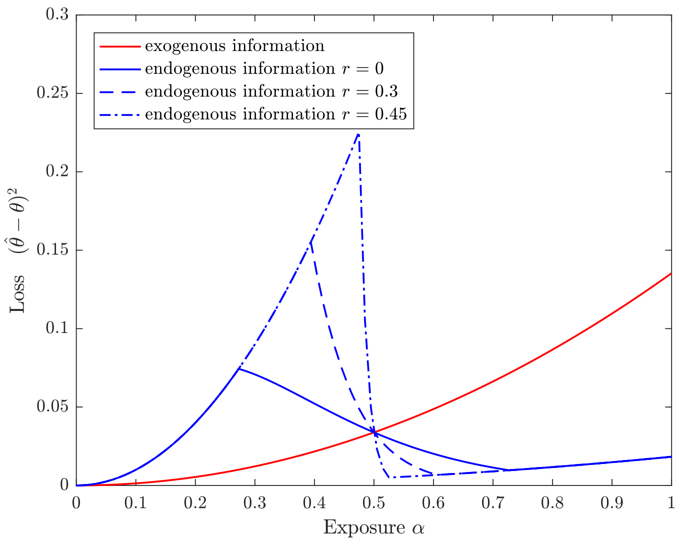


Figure 1.6 Loss Due to Suboptimal Investment for Varying Degrees of Strategic Complementarity.

The figure illustrates the relationship between the degree of exposure to factor 1, α, and the loss due to suboptimal action that is induced by a shock to factor 1, for varying degrees of strategic complementary between the firms populating the economy. This example considers a one standard deviation shock to factor 1 i.e. $\epsilon_1 = \sigma_1$, abstracts from factor 2 shocks i.e. $\epsilon_2 = 0$ and information shocks i.e. $\epsilon_{s_1} = \epsilon_{s_2} = 0$. The other parameter values are $\sigma_1 = \sigma_2 = 1$, $K = 1$ and $\tilde{\sigma}_{s_1} = \tilde{\sigma}_{s_2} = 0.75$.

1.4.3 Extension: correlated risks

The baseline model considers the case in which the risk factors affecting fundamentals are independent. In this section I allow for the risk factors to be correlated. I find that the baseline main result remains unchanged and the loss due to suboptimal investment increases with the degree of correlation between the two risk factors.

To deal with the case of correlated risks it is useful to use matrix notation. As in the baseline model, fundamentals are a sum of risk factors. Let f be a $N \times 1$ vector of risk factors and A be a $N \times 1$ vector of exposures to these factors. Fundamentals can be expressed as

$$\theta = A'f$$

where the risk factors f are ex-ante known to be correlated i.e. the prior variance-covariance matrix of the risk factors f is non-diagonal. Assuming that the prior variance-covariance matrix of the factors f is non-diagonal is equivalent to assuming the following linear structure for the risk factors

$$f = \mu + \Gamma\epsilon, \quad \epsilon \sim \mathcal{N}(0, \Sigma) \tag{1.28}$$

where μ is a $N \times 1$ vector of constants measuring the mean level of each risk factor, ϵ is a $N \times 1$ vector of independent shocks with diagonal variance-covariance matrix Σ, and Γ is an $N \times N$ matrix of loadings which measures the extent to which the independent shocks affect the risk factors.[26] The i^{th} row of the matrix Γ, denoted Γ_i, gives the loadings of the i^{th} risk factor f_i on the independent shocks in the vector ϵ. Thus, each risk factor f_i is expressed as the sum of a factor-specific mean μ_i and the independent random variables or shocks contained in the vector ϵ, which affect it in proportion to the loadings Γ_i.

This factor structure essentially allows for correlations between the risk factors through shared exposure to underlying independent shocks. It accounts for the existence of underlying forces that might be driving more than one of the risk factors affecting fundamentals. I conceptualize these independent shocks as factor-specific shocks. For instance, I interpret shock 1, ϵ_1, as being specific to factor 1, f_1, and shock 2, ϵ_2, as being specific to factor 2, f_2. Correlation is introduced by allowing factor 1 to load on the shock specific to factor 2, and vice versa.

Interpreted in the context of the motivating example, this setup accounts for the reality that domestic and foreign risks are likely to be correlated. The linear factor modelling approach adopted above is essentially equivalent to principal component analysis, which provides a way to decompose correlated risks into independent risks. In the portfolio literature it is common to use principal components analysis to decompose sets of correlated asset returns into independent underlying risk factors such as business-cycle risk, industry-specific risk, and firm-specific risk (Ross, 1976). Similarly, correlated domestic and foreign risks can be decomposed into an exhaustive set of independent underlying risk factors, which can be interpreted as pure country-specific risks.

The firm aims to reduce uncertainty about these underlying shocks through learning. Thus, signals will be about the independent shocks contained in the vector ϵ. I assume that learning about independent shocks is done independently. In other words, the firm acquires independent noisy signals about each of the independent shocks contained in the vector ϵ, and thus receives a $N \times 1$ vector of independent signals

$$s = \epsilon + \epsilon_s, \quad \epsilon_s \sim \mathcal{N}(0, \Sigma_s)$$

where the Σ_s variance-covariance matrix of the ϵ_s signal noise vector is diagonal, thus capturing

[26] The variance-covariance matrix of the risk factors that is implied by (1.28) is $\Gamma\Sigma\Gamma'$. Note that an alternative solution method is to assume that the prior variance-covariance matrix of the shocks is non-diagonal, say Ω, and then use eigen-decomposition to re-write it as $\Omega = \Gamma\Sigma\Gamma'$; learning would then be about the principal components with diagonal variance-covariance matrix Σ.

the assumption that signal about independent risks are independent.

Applying Bayes's rule on the transformed variable $\Gamma^{-1}f$, and then pre-multiplying this solution by Γ, I obtain that posterior beliefs about the correlated risk factors have mean $E[f|s] = \mu + \Gamma(I - \hat{\Sigma}\Sigma^{-1})s$ and variance $V[f|s] = \Gamma\hat{\Sigma}\Gamma'$, where $\hat{\Sigma} \equiv (\Sigma^{-1} + \Sigma_s^{-1})^{-1}$ denotes the posterior variance-covariance matrix of the independent shocks ϵ.[27] Consequently, the posterior mean and variance of fundamentals, respectively, conditional on the information available at time 2 are

$$\begin{aligned} \hat{\theta} &\equiv E[\theta|s] = A'E[f|s] = A'(\mu + \Gamma(I - \hat{\Sigma}\Sigma^{-1})s) \\ \hat{\sigma}^2 &\equiv V[\theta|s] = A'V[f|s]A = A'\Gamma\hat{\Sigma}\Gamma'A \end{aligned}$$

The solution strategy follows the same steps as in the baseline model. The date-2 problem is unchanged and yields solution $\lambda = E[\theta|s] = \hat{\theta}$. Similarly, the date-1 utility decreases with uncertainty regarding fundamentals $V[\theta|s]$, so the date-1 problem is to minimize

$$V[\theta|s] = A'\Gamma\hat{\Sigma}\Gamma'A$$

subject to the information processing constraint

$$\frac{1}{2}\ln\frac{|\Sigma|}{|\hat{\Sigma}|} \leq K \tag{1.29}$$

and the the no-forgetting constraint, which is essentially a restriction that the matrix Σ_s is positive semi-definite. Note that since the variance-covariance matrices that enter the determinants in (1.29) are diagonal, the information processing constraint can be re-written as a sum. Furthermore, define the information-processing capacity devoted to learning about each of the underlying independent shocks as $k_i \equiv \frac{1}{2}\ln\frac{\Sigma_{ii}}{\hat{\Sigma}_{ii}}$. The information-processing constraint (1.29) thus reduces to $\sum_i k_i \leq K$.

The matrix of loadings Γ is essentially a measure of the correlation structure between the risk factors. The rows of the loadings matrix give the loadings of each factor on all the shocks and the columns give the loadings of all the factors on each shock . The i^{th} row of the matrix Γ, denoted Γ_i, gives the loadings of the i^{th} risk factor on the independent shocks in the vector of shocks ϵ. Denote by Γ_j the j^{th} column of the matrix Γ, which gives the loadings of all the risk factors on the j^{th} shock. Define exposure to shock j as

[27]Transforming the variable $f^* = \Gamma^{-1}f = \Gamma^{-1}\mu + \epsilon$, allows applying standard Bayesian rules for updating normally distributed variables and yields posterior mean $E[f^*|s] = \Gamma^{-1}\mu + V[f^*|s]\Sigma_s^{-1}s$ and posterior variance $V[f^*|s] = V[\epsilon|s] = (\Sigma^{-1} + \Sigma_s^{-1})^{-1} \equiv \hat{\Sigma}$.

$$E_j \equiv A'\Gamma_j = \sum_{i=1}^{n} \alpha_i \Gamma_{ij}.$$

This measure of exposure captures the intuition that when the risk factors are correlated, the degree to which an underlying shock affects fundamentals will depend on the interaction between the observable exposure to the risk factors (captured by A) as well as on the loading of the risk factor on the underlying common shocks (captured by Γ). In other words, the effective exposure to a shock depends on the interaction between the observed exposure and the underlying correlation structure.

For ease of exposition, in what follows I consider and solve for the case in which $n = 2$. In this case the relevant effective exposure parameters are $E_j = \alpha_1\Gamma_{1j} + \alpha_2\Gamma_{2j},\ j = 1, 2$. The date-1 problem can thus be expressed as

$$\min_{k_1,k_2}\ V[\theta|s] = A'\Gamma\hat{\Sigma}\Gamma' A = \hat{\Sigma}_{11}E_1^2 + \hat{\Sigma}_{22}E_2^2$$

$$\text{s.t.}\quad \hat{\Sigma}_{ii} = \Sigma_{ii}e^{-2k_i},\quad \sum k_i \leq K,\quad 0 \leq k_i,\quad i = 1, 2.$$

The optimal capacity allocated to the factor-specific shocks is

$$k_1 = \begin{cases} 0 & \text{if } \sqrt{\frac{\Sigma_{11}E_1^2}{\Sigma_{22}E_2^2}} < e^{-K} \\ \frac{1}{2}\left(K + \ln\sqrt{\frac{\Sigma_{11}E_1^2}{\Sigma_{22}E_2^2}}\right) & \text{if } e^{-K} \leq \sqrt{\frac{\Sigma_{11}E_1^2}{\Sigma_{22}E_2^2}} \leq e^{K} \\ K & \text{if } \sqrt{\frac{\Sigma_{11}E_1^2}{\Sigma_{22}E_2^2}} > e^{K} \end{cases} \tag{1.30}$$

and $k_2 = K - k_1$. Note that this solution for the attention allocated to the underlying shocks is similar in spirit to the attention allocated the independent risk factors that was discussed in Section 1.3.1. However, the implications in terms of magnitude of losses are different relative to baseline model.

Figure 1.7 plots the loss due to suboptimal action against exposure to factor 1, for varying degrees of correlation between the two risk factors. The example considers a one standard deviation shock to factor 1 ϵ_1, abstracts from factor 2 shocks ϵ_2 as well as from information shocks ϵ_s. The main result on the non-monotonic relationship between exposure and the loss due to suboptimal investment remains unchanged. However, the figure shows that relative to the zero correlation baseline, the loss due to suboptimal investment increases with the degree of correlation between the two factors (Appendix **A.4** provides an algebraic derivation of this result).

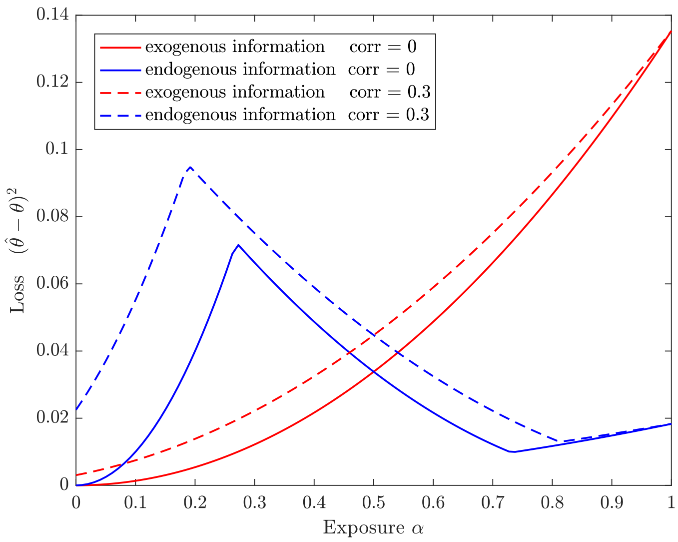


Figure 1.7 Loss Due to Suboptimal Investment for Correlated Risks. The figure illustrates the relationship between the degree of exposure to factor 1, α, and the loss due to suboptimal action that is induced by a shock to factor 1, for varying degrees of correlation between the risk factors. This example considers a one standard deviation factor 1 specific shock i.e. $\epsilon_1 = \Sigma_{11}$, abstracts from factor 2 shocks i.e. $\epsilon_2 = 0$ and information shocks i.e. $\epsilon_{s_1} = \epsilon_{s_2} = 0$. The parameter values are $K = 1$, $\tilde{\sigma}_{s_1} = \tilde{\sigma}_{s_2} = 0.75$, $\Sigma_{11} = \Sigma_{22} = 1$. Both risk factors are assumed to load equally on the underlying independent shocks: $\Gamma_{11} = \Gamma_{22} = 1$ and $\Gamma_{12} = \Gamma_{21} = 0.15$. The implied correlation between the two risk factors is 0.3.

A higher degree of correlation between the two risk factors has two implications. On the one hand, correlation introduces learning complementarity benefits, as the firm can use information about one factor-specific shock to reduce uncertainty about both risk factors. On the other hand correlation also increases the effective exposure to shocks because now a shock specific to factor 1 will affects fundamentals not only through exposure to factor 1, but also through exposure to factor 2. The effect of complementary in learning is to shift the loss turning point along the x-axis. In the specific case illustrated in Figure 1.7, the loss function is shifted to the left relative to the zero-correlations baseline because the firm starts learning about the shock specific to factor 1 at a lower level of observable exposure (the observable exposure plotted on the x-axis is lower than the effective exposure which drives learning choices in (1.30)). The effect of increased effective exposure to shocks is to shift the loss function upwards along the y-axis. The loss is higher relative to the zero-correlations baseline because effective exposure is higher than the observable exposure that is plotted on the x-axis. This result highlights that the apparently unexplained transmission of shocks i.e. transmission of shocks that is not explained by observable measure of exposure to shock can also occur because of underlying correlations between the risk factors driving fundamentals.

1.4.4 Extension: endogenous exposure

The baseline model considered the case in which exposures to risk factors are exogenous. In this section, I allow for exposure to be endogenous quantities determined in equilibrium. The basic result is that it is optimal for the firm to specialize in learning about one risk factor and to be more relatively exposed to that factor.

The solution strategy follows the same steps as in the baseline model in Section 1.3, except that at the first date, in addition to choosing the amount of information processing resources devoted to each risk factor, the firm also chooses the exposure to these risk factors. In particular, given the optimal action and the optimal factor-specific information processing capacity for any given exposure, the firm then chooses the optimal level of exposure by solving

$$\max_{\alpha} U_1 = \begin{cases} -\alpha^2\sigma_1^2 - (1-\alpha)^2\sigma_2^2 e^{-2K} & \text{if } \ k_1 = 0 \\ -2\alpha(1-\alpha)\sigma_1\sigma_2 e^{-K} & \text{if } \ k_1 = \frac{1}{2}\left(K + \ln\frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right) \\ -\alpha^2\sigma_1^2 e^{-2K} - (1-\alpha)^2\sigma_2^2 & \text{if } \ k_1 = K \end{cases} \tag{1.31}$$

The following proposition summarizes the optimal level of exposure result.

Proposition 4. *It is optimal to be relatively more exposed to the risk factor that the firm learns about*

$$\alpha^* = \begin{cases} \frac{\sigma_2^2}{\sigma_2^2+\sigma_1^2 e^{2K}} & \textit{if} \quad k_1 = 0 \\ \frac{\sigma_2^2}{\sigma_2^2+\sigma_1^2 e^{-2K}} & \textit{if} \quad k_1 = K \end{cases} \tag{1.32}$$

If the risk factors are ex-ante equally volatile, then the firm is indifferent between the two exposure allocations. If the risk factors are not ex-ante equally volatile, then it is optimal to be relatively more exposed to the factor that is ex-ante less volatile.

Proof. See Appendix **A.5** for the full characterization of the equilibria.

The analytical expression (1.32) reveals that the optimal level of exposure to factor 1 decreases with factor 1 uncertainty and increases with factor 2 uncertainty. Furthermore, optimal factor 1 exposure increases with total capacity K if the firm chooses to learn about factor 1 and it decreases with capacity K if the firm chooses to learn about factor 2.[28] Note that for any limited precision and ex-ante uncertain risk factor, unit exposure to one factor is not optimal i.e. for any $K < \infty$ and $\sigma_i^2 > 0,\ i = 1,2 \Rightarrow 0 < \alpha^* < 1$.

[28]This is because less capacity constrained agents are able to learn more and the levels of exposure for which specialized learning occurs are more extreme

Therefore, the firm prefers to be relatively more exposed to one factor, the factor that it learns about. Indifference between exposure allocations arises if the risk factors are ex-ante equally volatile. Otherwise, it is optimal to learn about and to be relatively more exposed to the risk factor that is ex-ante less volatile. Thus, it is ex-ante optimal for the firm to specialize in learning about one risk factor and to be relatively more exposed to it. Note that for any finite level of capacity full exposure to a risk factor is never optimal and the optimal exposure is an interior solution. These ex-ante optimal exposure allocations will, however, expose the firm to the risk of incurring a higher loss due to suboptimal investment in the event that the risk factor that it is relatively less exposed to (and about which it does not learn) is hit by a shock.

1.5 Concluding remarks

The financial crisis of 2007-2008 has highlighted the existence of a remarkable and poorly understood type of contagion whereby countries that were relatively less exposed to the crisis epicentre, the United States, were among the most severely affected. In other words, this crisis has shown that the impact of a shock can decrease with exposure to it. In this chapter, I study how endogenous information choice affects decision-makers' reactions to shocks and as a consequence the impact of those shocks. By linking information choice and learning behavior with exposure, the model I propose in this chapter explains the puzzling observation that the impact of a shocks can decrease with exposure to it. The key mechanism in my model is that learning increases with exposure, such that the cost of being highly exposed to a shock is mitigated by the benefit of having a better understanding of it. My model contributes to understanding observed cross-sectional and time-series patterns of contagion. In particular, my model explains how countries that are more exposed to a crisis can be less affected and why contagion is more likely to occur following unexpected crises.

Appendix

A.1 Proof of Proposition 1:

Start by deriving the factor-specific posterior uncertainty. Recalling the relation between the information processing capacity allocated to a risk factor and reduction in uncertainty achieved by it implied by the entropy constraint in (1.10), factor-specific posterior uncertainty can be determined to be

$$\hat{\sigma}_1^2 = \begin{cases} \sigma_1^2 & \text{if } \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} < e^{-K} \\ \frac{1-\alpha}{\alpha}\sigma_1\sigma_2 e^{-K} & \text{if } e^{-K} \leq \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} \leq e^{K} \\ \sigma_1^2 e^{-2K} & \text{if } \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} > e^{K} \end{cases} \tag{A.1}$$

and $\hat{\sigma}_2^2 = \sigma_1^2\sigma_2^2\hat{\sigma}_1^{-2}e^{-2K}$. Note that factor-specific risk is effectively reduced through learning. Relative to a full information benchmark (i.e. $\hat{\sigma}_1^2 = \sigma_1^2$), risk is lower if exposure is relatively high and/or prior uncertainty is relatively high (i.e. $\hat{\sigma}_1^2 < \sigma_1^2$ if $\alpha \geq 0.5$ and/or $\sigma_1 \geq \sigma_1$); this can also be interpreted as risks being under-estimated.

Uncertainty under the exogenous information benchmark is given by

$$\tilde{\sigma}_B^2 = \alpha^2\tilde{\sigma}_1^2 + (1-\alpha)^2\tilde{\sigma}_2^2 \tag{A.2}$$

Under the endogenous information model uncertainty $\hat{\sigma}^2 = \alpha^2\hat{\sigma}_1^2 + (1-\alpha)^2\hat{\sigma}_2^2$ is given by

$$\hat{\sigma}^2 = \begin{cases} \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2 e^{-2K} & \text{if } \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} < e^{-K} \\ 2\alpha(1-\alpha)\sigma_1\sigma_2 e^{-K} & \text{if } e^{-K} \leq \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} \leq e^{K} \\ \alpha^2\sigma_1^2 e^{-2K} + (1-\alpha)^2\sigma_2^2 & \text{if } \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} > e^{K} \end{cases} \tag{A.3}$$

Contrast comparative statics with respect to exposure under the two models (at the interior optimum for endogenous information choice)

$$\frac{\partial\tilde{\sigma}_B^2}{\partial\alpha} = 2\alpha\tilde{\sigma}_1^2 - 2(1-\alpha)\tilde{\sigma}_2^2 > 0 \text{ if } \alpha > 0.5 \tag{A.4}$$

$$\frac{\partial\hat{\sigma}^2}{\partial\alpha} = 2(1-2\alpha)\sigma_1\sigma_2 e^{-K} < 0 \text{ if } \alpha > 0.5 \tag{A.5}$$

Analytically, we can contrast the results under the learning model with the exogenous information benchmark results by setting equal the information processing capacity in the two models (plug K when the capacity constraint is binding i.e. $\frac{1}{2}\ln\frac{\sigma_1^2\sigma_2^2}{\tilde{\sigma}_1^2\tilde{\sigma}_2^2} = K \;\Rightarrow\; e^{-K} = \frac{\tilde{\sigma}_1\tilde{\sigma}_2}{\sigma_1\sigma_2}$ into result (A.3)). This is informative of the factor-specific risk reduction entailed by learning, or the learning-adjusted risk of a factor and its contribution to overall uncertainty.

$$\hat{\sigma}^2 = \begin{cases} \alpha^2\sigma_1^2 + (1-\alpha)^2\tilde{\sigma}_2^2\frac{\tilde{\sigma}_1^2}{\sigma_1^2} & \text{if} \quad \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} < \frac{\tilde{\sigma}_1\tilde{\sigma}_2}{\sigma_1\sigma_2} \\ \alpha^2\tilde{\sigma}_1^2\frac{1-\alpha}{\alpha}\frac{\tilde{\sigma}_2}{\tilde{\sigma}_1} + (1-\alpha)^2\tilde{\sigma}_2^2\frac{\alpha}{1-\alpha}\frac{\tilde{\sigma}_1}{\tilde{\sigma}_2} & \text{if} \quad \frac{\tilde{\sigma}_1\tilde{\sigma}_2}{\sigma_1\sigma_2} \le \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} \le \frac{\sigma_1\sigma_2}{\tilde{\sigma}_1\tilde{\sigma}_2} \\ \alpha^2\tilde{\sigma}_1^2\frac{\tilde{\sigma}_2^2}{\sigma_2^2} + (1-\alpha)^2\sigma_2^2 & \text{if} \quad \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} > \frac{\sigma_1\sigma_2}{\tilde{\sigma}_1\tilde{\sigma}_2} \end{cases} \tag{A.6}$$

These analytical results confirm the message conveyed in Figure 1.2 that relative to the equal capacity, exogenous information benchmark, uncertainty is lower when decision-makers are allowed to optimally allocate their information resources across risk factor exposures. It can be shown that the corner solutions are always smaller than the benchmark solutions. For the first benchmark solution: $\alpha^2\tilde{\sigma}_1^2 + (1-\alpha)^2\tilde{\sigma}_2^2 > \alpha^2\sigma_1^2 + (1-\alpha)^2\tilde{\sigma}_2^2\frac{\tilde{\sigma}_1^2}{\sigma_1^2} \Leftrightarrow \tilde{\sigma}_2^2 > \frac{\alpha^2}{(1-\alpha)^2}\sigma_1^2$ which is true in light of the condition for obtaining the corner solution, which can be written as $\frac{\tilde{\sigma}_1}{\sigma_1}\tilde{\sigma}_2 > \frac{\alpha}{1-\alpha}\sigma_1$. Since $\tilde{\sigma}_2 \ge \frac{\tilde{\sigma}_1}{\sigma_1}\tilde{\sigma}_2 \left(> \frac{\alpha}{1-\alpha}\sigma_1\right) \Rightarrow \tilde{\sigma}_2^2 > \frac{\alpha^2}{(1-\alpha)^2}\sigma_1^2$. Similarly, for the second corner solution we have that $\alpha_1^2\tilde{\sigma}_1^2 + \alpha_2^2\tilde{\sigma}_2^2 > \alpha^2\tilde{\sigma}_1^2\frac{\tilde{\sigma}_2^2}{\sigma_2^2} + \alpha_2^2\sigma_2^2 \Leftrightarrow \frac{\alpha}{1-\alpha}\frac{1}{\sigma_2} > \frac{1}{\tilde{\sigma}_1}$. This follows from the corner solution condition which can be re-written as $\frac{\alpha}{1-\alpha}\frac{1}{\sigma_2} \ge \frac{\sigma_2}{\tilde{\sigma}_1\tilde{\sigma}_2}$ and the fact that $\frac{1}{\tilde{\sigma}_1}\frac{\sigma_2}{\tilde{\sigma}_2} > \frac{1}{\tilde{\sigma}_1}$. For the interior solution we have that $\alpha^2\tilde{\sigma}_1^2 + (1-\alpha)^2\tilde{\sigma}_2^2 \ge 2\alpha_1\alpha_2\tilde{\sigma}_1\tilde{\sigma}_2$, which holds because $(\alpha_1\tilde{\sigma}_1 - \alpha_2\tilde{\sigma}_2)^2 \ge 0$.

Importantly, they shed further light into the mechanism behind the observed dynamics. Relative to the benchmark, when exposure to a factor is very low and there is no learning about it (corner solution), the effective or learning-adjusted risk of that factor is higher ($\sigma_1^2 \ge \tilde{\sigma}_1^2$), while the effective risk of the other factor is lower ($\tilde{\sigma}_2^2\frac{\tilde{\sigma}_1^2}{\sigma_1^2} \le \tilde{\sigma}_1^2$). At the interior optimum, factor-specific effective risk is higher if factor-specific exposure is relatively low (all else equal $\tilde{\sigma}_1^2\frac{1-\alpha}{\alpha} \ge \tilde{\sigma}_1^2$ if $\alpha < 1-\alpha$) and it is lower if factor-specific exposure is relatively high (all else equal $\tilde{\sigma}_1^2\frac{1-\alpha}{\alpha} \le \tilde{\sigma}_1^2$ if $\alpha > 1-\alpha$).

A.2 Proof of Proposition 2:

Define $\gamma_i = \frac{\hat{\sigma}_i^2}{\sigma_i^2}$ so that we can re-write the conditional mean value of fundamentals as

$$\hat{\theta} \equiv E[\theta|s_1, s_2] \quad = \quad \alpha\left[\mu_1 + (1-\gamma_1)s_1\right] + (1-\alpha)\left[\mu_2 + (1-\gamma_2)s_2\right] \tag{A.7}$$

where

$$\gamma_1 = \begin{cases} 1 & \text{if } \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} < e^{-K} \\ \frac{1-\alpha}{\alpha}\frac{\sigma_2}{\sigma_1}e^{-K} & \text{if } e^{-K} \le \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} \le e^{K} \\ e^{-2K} & \text{if } \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} > e^{K} \end{cases} \tag{A.8}$$

$$\gamma_2 = \gamma_1^{-1}e^{-2K} \tag{A.9}$$

Recalling that $\theta = \alpha(\mu_1 + \epsilon_1) + (1-\alpha)(\mu_2 + \epsilon_2)$ and $s_i = \epsilon_i + \epsilon_{s_i}$, the loss due to suboptimal action is given by

$$L \equiv (\hat{\theta} - \theta)^2 = \Big(\alpha\left[(1-\gamma_1)\epsilon_{s_1} - \gamma_1\epsilon_1\right] + (1-\alpha)\left[(1-\gamma_2)\epsilon_{s_2} - \gamma_2\epsilon_2\right]\Big)^2 \tag{A.10}$$

such that the interpretation of the parameters $\gamma_i,\ i = 1,2$ is that of the weight assigned to the shocks affecting fundamentals. The loss function under the exogenous information benchmark takes the same form except that the weight coefficients are different $\tilde{\gamma}_i \equiv \frac{\tilde{\sigma}_i^2}{\sigma_i^2},\ i = 1,2$

Introduce the exogenous information, equal capacity benchmark

$$L_B = (\hat{\theta}_B - \theta)^2 = (\alpha\left[(1-\tilde{\gamma}_1)\epsilon_{s_1} - \tilde{\gamma}_1\epsilon_1\right] + (1-\alpha)\left[(1-\tilde{\gamma}_2)\epsilon_{s_2} - \tilde{\gamma}_2\epsilon_2\right])^2 \tag{A.11}$$

where $\tilde{\gamma}_1 \equiv \frac{\tilde{\sigma}_1^2}{\sigma_1^2}$, $\tilde{\gamma}_2 \equiv \frac{\tilde{\sigma}_2^2}{\sigma_2^2}$, $\gamma_1 \equiv \frac{\hat{\sigma}_1^2}{\sigma_1^2} = \frac{1-\alpha}{\alpha}\sigma_1\sigma_2 e^{-K}$, $\gamma_2 \equiv \frac{\hat{\sigma}_2^2}{\sigma_2^2} = \frac{\alpha}{1-\alpha}\sigma_1\sigma_2 e^{-K}$.

Define the shock transmission mechanism as the loss induced by a shock

$$\frac{\partial L}{\partial \epsilon_1} = -2\alpha\gamma_1\Big(\alpha\left[(1-\gamma_1)\epsilon_{s_1} - \gamma_1\epsilon_1\right] + (1-\alpha)\left[(1-\gamma_2)\epsilon_{s_2} - \gamma_2\epsilon_2\right]\Big) \tag{A.12}$$

$$\frac{\partial L_B}{\partial \epsilon_1} = -2\alpha\tilde{\gamma}_1\Big(\alpha\left[(1-\tilde{\gamma}_1)\epsilon_{s_1} - \tilde{\gamma}_1\epsilon_1\right] + (1-\alpha)\left[(1-\tilde{\gamma}_2)\epsilon_{s_2} - \tilde{\gamma}_2\epsilon_2\right]\Big) \tag{A.13}$$

We are interested in how the transmission mechanism varies with exposure α. Let us recall that the weight coefficients in the endogenous learning model are functions of exposure $\gamma_1 \equiv \frac{\hat{\sigma}_1^2}{\sigma_1^2} = \frac{1-\alpha}{\alpha}\sigma_1\sigma_2 e^{-K}$, $\gamma_2 \equiv \frac{\hat{\sigma}_2^2}{\sigma_2^2} = \frac{\alpha}{1-\alpha}\sigma_1\sigma_2 e^{-K}$. Abstracting from factor 2 effects by setting $\epsilon_2 = \epsilon_{s_2} = 0$, and using the result that $\frac{\partial \gamma_1}{\partial \alpha} = -\frac{1}{\alpha(1-\alpha)}\gamma_1$, we have that

$$\frac{\partial^2 L}{\partial \epsilon_1 \partial \alpha} = 2\alpha\gamma_1\left[\left(\frac{2\alpha - 1}{1-\alpha} - \frac{2\alpha}{1-\alpha}\gamma_1\right)\epsilon_{s_1} - \frac{2\alpha}{1-\alpha}\gamma_1\epsilon_1\right] \tag{A.14}$$

$$\frac{\partial^2 L_B}{\partial \epsilon_1 \partial \alpha} = -4\alpha\tilde{\gamma}_1\left[(1-\tilde{\gamma}_1)\epsilon_{s_1} - \tilde{\gamma}_1\epsilon_1\right] \tag{A.15}$$

In the limiting case in which the signal noise is zero $\epsilon_{s_1} = 0$ (the signal is perfectly informative), it is clear that $\frac{\partial^2 L}{\partial \epsilon_1 \partial \alpha} < 0$ (the transmission mechanism decreases with exposure) and $\frac{\partial^2 L_B}{\partial \epsilon_1 \partial \alpha} > 0$ (the transmission mechanism increases with exposure). These results hold more generally if the signal is informative i.e. the signal noise is sufficiently small $\epsilon_{s_1} < \frac{\gamma_1}{1-\gamma_1}\epsilon_1$.

A.3 Proof of Proposition 3:

The date-1 utility function (1.18) is a continuous piecewise function that is increasing in state capacity. It is useful to distinguish between three types of equilibria: (i) the equilibrium capacity allocation across factors has the property $k_1 = 0$, (ii) the equilibrium capacity allocation across factors has the property $k_1 = K$, and (iii) the equilibrium capacity allocation across factors has the property $k_1 = \frac{1}{2}\left(K + \ln \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right)$. Substituting the constraint (1.20) into the objective function (1.19) and differentiating with respect to K_r yields first-order condition in the three set of equilibria

$$\begin{cases} 2p_r(1-\alpha)^2\sigma_2^2 e^{-2K_r} - 2p_n(1-\alpha)^2\sigma_2^2 e^{-2(\mathcal{K}-K_r)} & \text{if } k_1 = 0 \\ 2p_r\alpha(1-\alpha)\sigma_1\sigma_2 e^{-K_r} - 2p_n\alpha(1-\alpha)\sigma_1\sigma_2 e^{-(\mathcal{K}-K_r)} & \text{if } k_1 = \frac{1}{2}\left(K + \ln \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right) \\ 2p_r\alpha^2\sigma_1^2 e^{-2K_r} - 2p_n\alpha^2\sigma_1^2 e^{-2(\mathcal{K}-K_r)} & \text{if } k_1 = K \end{cases} \tag{A.16}$$

Solving for K_r yields

$$K_r = \begin{cases} \frac{1}{4}\left(2\mathcal{K} + \ln \frac{p_r}{p_n}\right) & \text{if } k_1 = 0 \text{ or } k_1 = K \\ \frac{1}{2}\left(\mathcal{K} + \ln \frac{p_r}{p_n}\right) & \text{if } k_1 = \frac{1}{2}\left(K + \ln \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right) \end{cases} \tag{A.17}$$

Imposing the no-forgetting constraint $0 \leq K_r$ and noting that $K_r \leq \mathcal{K}$ is always satisfied because $\ln \frac{p_r}{p_n} < 0$ when $p_r < p_n$, the optimal capacity allocation across states when the firm can only learn about one state (when a corner solution is obtained for capacity allocation across factors) is given by

$$K_r = \begin{cases} 0 & \text{if } \frac{p_r}{p_n} < e^{-2\mathcal{K}} \\ \frac{1}{4}\left(2\mathcal{K} + \ln \frac{p_r}{p_n}\right) & \text{if } \frac{p_r}{p_n} \geq e^{-2\mathcal{K}} \end{cases} \tag{A.18}$$

$$K_n = \mathcal{K} - K_r \tag{A.19}$$

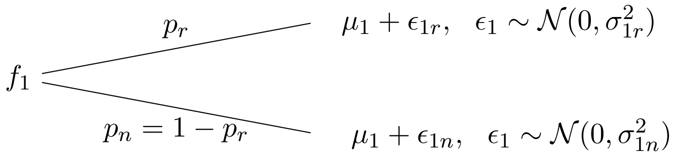
and the optimal information processing capacity allocated to the rare state of nature when the firm can only learn about one state (when an interior solution is obtained for capacity allocation across factors) is given by

$$K_r = \begin{cases} 0 & \text{if } \frac{p_r}{p_n} < e^{-\mathcal{K}} \\ \frac{1}{2}\left(\mathcal{K} + \ln\frac{p_r}{p_n}\right) & \text{if } \frac{p_r}{p_n} \geq e^{-\mathcal{K}} \end{cases} \tag{A.20}$$

$$K_n = \mathcal{K} - K_r \tag{A.21}$$

Note that if the firm only learns about one as opposed to both factors, it is more likely that it will dedicate information processing resources to the rare state (since $e^{-2\mathcal{K}} < e^{-\mathcal{K}}$) but the overall capacity allocated to the state is smaller (since $\frac{1}{4}\left(2\mathcal{K} + \ln\frac{p_r}{p_n}\right) < \frac{1}{2}\left(\mathcal{K} + \ln\frac{p_r}{p_n}\right)$).

If the shock variance is different in the two states



then the capacity allocation is

$$K_r = \begin{cases} 0 & \text{if } \frac{p_r\sigma_{1r}}{p_n\sigma_{1n}} < e^{-\mathcal{K}} \\ \frac{1}{2}\left(\mathcal{K} + \ln\frac{p_r\sigma_{1r}}{p_n\sigma_{1n}}\right) & \text{if } e^{-\mathcal{K}} \leq \frac{p_r\sigma_{1r}}{p_n\sigma_{1n}} \leq e^{\mathcal{K}} \\ \mathcal{K} & \text{if } \frac{p_r\sigma_{1r}}{p_n\sigma_{1n}} > e^{\mathcal{K}} \end{cases}$$

A.4 Derivation of correlated risks result

The shock-specific posterior uncertainty implied by the capacity allocation (1.30) is given by

$$\hat{\Sigma}_{11} = \begin{cases} \Sigma_{11} & \text{if } \sqrt{\frac{\Sigma_{11}E_1^2}{\Sigma_{22}E_2^2}} < e^{-K} \\ \sqrt{\frac{\Sigma_{11}\Sigma_{22}E_2^2}{E_1^2}}e^{-K} & \text{if } e^{-K} \leq \sqrt{\frac{\Sigma_{11}E_1^2}{\Sigma_{22}E_2^2}} \leq e^{K} \\ \Sigma_{11}e^{-2K} & \text{if } \sqrt{\frac{\Sigma_{11}E_1^2}{\Sigma_{22}E_2^2}} > e^{K} \end{cases} \tag{A.22}$$

and $\hat{\Sigma}_{22} = \Sigma_{11}\Sigma_{22}e^{-2K}\hat{\Sigma}_{11}^{-1}$

Consequently, the uncertainty about fundamentals is given by

$$V[\theta|s] = \hat{\Sigma}_{11}E_1^2 + \hat{\Sigma}_{22}E_2^2 = \begin{cases} \Sigma_{11}E_1^2 + \Sigma_{22}E_2^2e^{-2K} & \text{if } \sqrt{\frac{\Sigma_{11}E_1^2}{\Sigma_{22}E_2^2}} < e^{-K} \\ 2\sqrt{\Sigma_{11}\Sigma_{22}E_1^2E_2^2}e^{-K} & \text{if } e^{-K} \leq \sqrt{\frac{\Sigma_{11}E_1^2}{\Sigma_{22}E_2^2}} \leq e^{K} \\ \Sigma_{11}E_1^2e^{-2K} + \Sigma_{22}E_2^2 & \text{if } \sqrt{\frac{\Sigma_{11}E_1^2}{\Sigma_{22}E_2^2}} > e^{K} \end{cases}$$

The loss due to suboptimal investment is

$$\begin{aligned} L \equiv (E[\theta|s] - \theta)^2 &= (A'E[f|s] - A'f)^2 = [A'(\mu + \Gamma(I - \hat{\Sigma}\Sigma^{-1})s - A'(\mu + \Gamma\epsilon)]^2 \\ &= [A'(\mu + \Gamma(I - \hat{\Sigma}\Sigma^{-1})(\epsilon + \epsilon_s) - A'(\mu + \Gamma\epsilon)]^2 \\ &= [A'\Gamma(I - \hat{\Sigma}\Sigma^{-1})\epsilon_s + A'\Gamma(I - \hat{\Sigma}\Sigma^{-1})\epsilon - A'\Gamma\epsilon]^2 \\ &= [A'\Gamma(I - \hat{\Sigma}\Sigma^{-1})\epsilon_s - A'\Gamma\hat{\Sigma}\Sigma^{-1}\epsilon]^2 \end{aligned}$$

Abstracting from information shocks i.e. $\epsilon_s = 0$ we have that

$$L = (A'\Gamma\hat{\Sigma}\Sigma^{-1}\epsilon)^2$$

A.5 Proof of Proposition 4:

The objective function (1.31) is a continuous piecewise function, which is concave in exposure when a corner solution is obtained for capacity allocation, and convex in exposure when an interior solution is obtained. Hence, an interior solution is obtained for optimal exposure if a corner solution is obtained for information choice, and a corner solution is obtained for optimal exposure if an interior solution is obtained for information choice. The first-order condition is

$$\frac{\partial U_1}{\partial \alpha} = \begin{cases} -2\alpha\sigma_1^2 + 2(1-\alpha)\sigma_2^2e^{-2K} & \text{if } k_1 = 0 \\ -2(1-2\alpha)\sigma_1\sigma_2e^{-K} & \text{if } k_1 = \frac{1}{2}\left(K + \ln\frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right) \\ -2\alpha\sigma_1^2e^{-2K} + 2(1-\alpha)\sigma_2^2 & \text{if } k_1 = K \end{cases} \tag{A.23}$$

It is useful to distinguish between three types of equilibria: (i) the equilibrium capacity allocation has the property $k_1 = 0$, (ii) the equilibrium capacity allocation has the property $k_1 = K$, and (iii) the equilibrium capacity allocation has the property $k_1 = \frac{1}{2}\left(K + \ln\frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right)$.

Equilibria (i) and (ii) represents situations in which the firm is only able to learn about one factor. In these situations, it is optimal to be relatively more exposed to the factor the firm learns about, and the optimal point of exposure is the one at which the learning adjusted risk exposures are equal. More specifically, if $k_1 = 0$ optimal exposure is implied by $\alpha\sigma_1^2 = (1-\alpha)\sigma_2^2 e^{-2K}$. If $k_1 = K$ optimal exposure is implied by $\alpha\sigma_1^2 e^{-2K} = (1-\alpha)\sigma_2^2$.

$$\alpha^* = \begin{cases} \frac{\sigma_2^2}{\sigma_2^2+\sigma_1^2 e^{2K}} & \text{if} \quad k_1 = 0 \\ \frac{\sigma_2^2}{\sigma_2^2+\sigma_1^2 e^{-2K}} & \text{if} \quad k_1 = K \end{cases} \tag{A.24}$$

Equilibrium (iii) represents a situation in which the parameter values are such that the firm can learn about both factors. In this case, it is optimal to be as exposed as possible to one factor, where the maximum level of exposure to a factor is implied by the condition for learning about both risks $e^{-K} \le \frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} \le e^K$. To determine which risk factor it is optimal to relatively more exposed to, I compare the expected utility of each corner solution for exposure. The maximum exposure or loading to factor 1 is obtained when $\frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} = e^K$, which implies $\overline{\alpha} = \frac{\sigma_2}{\sigma_2+\sigma_1 e^{-K}}$ and the utility of being relatively more exposed to factor 1 is $U_1(\overline{\alpha}) = -2\left(\frac{\sigma_1\sigma_2}{\sigma_1+\sigma_2 e^K}\right)^2$. The maximum exposure or loading to factor 2 is obtained when $\frac{\alpha\sigma_1}{(1-\alpha)\sigma_2} = e^{-K}$, which implies minimum factor 1 exposure $\underline{\alpha} = \frac{\sigma_2}{\sigma_2+\sigma_1 e^K}$ and the utility of being relatively more exposed to factor 2 is $U_1(\underline{\alpha}) = -2\left(\frac{\sigma_1\sigma_2}{\sigma_1 e^K+\sigma_2}\right)^2$. In the case of symmetric equilibria whereby the two factors are ex-ante equally volatile, the firm will be indifferent between the two exposure allocations i.e. $U_1(\overline{\alpha}) = U_1(\underline{\alpha})$ if $\sigma_1 = \sigma_2$. However, in the case of non-symmetric equilibria, it is optimal to be more exposed to the less volatile risk factor i.e. $U_1(\overline{\alpha}) > U_1(\underline{\alpha})$ if $\sigma_1 < \sigma_2$ hence $\alpha^* = \overline{\alpha}$, and $U_1(\overline{\alpha}) < U_1(\underline{\alpha})$ if $\sigma_1 > \sigma_2$ hence $\alpha^* = \underline{\alpha}$.

$$\alpha^* = \begin{cases} \frac{\sigma_2}{\sigma_2+\sigma_1 e^K} & \text{if} \quad k_1 = \frac{1}{2}\left(K + \ln\frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right) \text{ and } \sigma_1 > \sigma_2 \\ \frac{\sigma_2}{\sigma_2+\sigma_1 e^{-K}} & \text{if} \quad k_1 = \frac{1}{2}\left(K + \ln\frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right) \text{ and } \sigma_1 < \sigma_2 \\ \frac{\sigma_2}{\sigma_2+\sigma_1 e^K} \text{ or } \frac{\sigma_2}{\sigma_2+\sigma_1 e^{-K}} & \text{if} \quad k_1 = \frac{1}{2}\left(K + \ln\frac{\alpha\sigma_1}{(1-\alpha)\sigma_2}\right) \text{ and } \sigma_1 = \sigma_2 \end{cases} \tag{A.25}$$

However, since the firm is not constrained to learn about both risk factors (i.e. to be in equilibria of the type (iii)), it will optimally choose to learn about one factor only and to be relatively more exposed to the factor it learns about. This follows from the fact that the expected utility associated with the optimal levels of exposure (A.25) obtained in equilibrium (iii) is lower than the utility associated with the optimal levels of exposure (A.24) that are obtained in equilibria (i) and (ii). Indifference between these latter exposure allocations (A.24) arises if the risk factors are ex-ante equally volatile, but it is otherwise optimal to be relatively more exposed to the factor that is ex-ante less volatile.

Chapter 2

Endogenously Ambiguous Information and Cautious Behaviour

2.1 Introduction

Information is important in shaping economic outcomes but carries a substantial risk of being misinterpreted. There is widespread evidence that the confidence that decision-makers have regarding the interpretation of information varies across time and markets. Specifically, there is evidence that certain states of the world, such as economic downturns and financial crises in particular, are characterized by stronger reactions to negative news by decision-makers, i.e. asymmetric reaction to news, as well as stronger correlations between markets i.e. asymmetric contagion.[1] This state-dependent cautious behavior, whereby negative news affect conditional actions more than good news, has been rationalized by assuming ambiguity-aversion. In this chapter, I endogenize ambiguity-averse behavior using costly information acquisition.

The theory I propose exploits an information acquisition mechanism whereby uncertainty in the interpretation of information increases endogenously in highly unanticipated states of the world, which causes ambiguity-averse decision-makers to behave cautiously by reacting more strongly to negative relative to positive news. More specifically, decision-makers in the model can invest in information about future states of the world such that upon the occurrence of any state, they receive signals about payoff-relevant variables. The degree of ambiguity of these signals is determined by their informational investment decision. The benefit of investing in information varies positively

[1]See Andersen, Bollerslev, Diebold, and Vega (2007); Bollerslev and Todorov (2011); Ben-David, Graham, and Harvey (2013); Kuhnen (2015); Zhou (2015); Williams (2014) for evidence on asymmetric reactions to news, and Ang and Chen (2002); Connolly and Wang (2003); Yuan (2005); Boyer, Kumagai, and Yuan (2006); Ozsoy (2013) for evidence on contagion asymmetry.

with the ex-ante likelihood of the state, but the cost does not. As a consequence, the ambiguity of information in a state of the world varies inversely with the ex-ante degree of anticipation of that state.

The model features uncertainty in the form of information that is difficult to interpret. It captures the intuition that the occurrence of highly unanticipated events creates a lot of uncertainty about their implications and about how things are going to be played out on account of a lack of precedent. An example of such a highly unanticipated event is the bankruptcy of Lehman Brothers, the emblematic event that triggered the collapse of the global economy. While Lehman was undoubtedly large and interconnected, it has been argued that it was not the event itself that played a crucial role in the unfolding of the crisis but it was the information content of the event.[2] Prior to Lehman's failure, an expectation of government assistance prevailed, which was strengthened by Bear-Stern's bailout earlier that year, as well as by the implicit government guarantees behind Fannie Mae and Freddie Mac's liabilities. The fact that Lehman was allowed to fail shocked the global financial community as it was widely assumed that no major financial institution would be allowed to go bankrupt. The subsequent bailout of AIG, along with a series of other seemingly impossible events helped form a strong feeling of uncertainty as economic decision-makers really had no idea what might happen.[3] The model I propose captures this intuition by building on the distinction between risk and Knightian uncertainty or ambiguity.[4]

Much of the subsequent policy measures taken by governments and regulators around the world have contained an element of uncertainty. From unconventional monetary policy, such as large-scale asset purchase programmes and zero or negative interest rates, to stringent banking regulations and a wide range of macroprudential tools, the lack of precedent characterising all these measures created an overall sense of uncertainty. The concerns regarding the limited knowledge on macroprudential tools have been numerous and the implications of unconventional monetary policy continue to be a matter of debate among experts.[5] It is not hard to imagine that all these unprecedented events, untested financial innovations and policy measures, have lead market participants to question their understanding of the environment and have made learning difficult. Decision-makers could no longer draw on their experience to interpret information and guide choice,

[2]As noted by the Committee on Capital Markets Regulation (2014) "the anti-bailout signal transmitted by the failure of Lehman, not the failure itself, triggered the spread of contagion effects".

[3]See The Federal Reserve of St. Louis for a timeline of events and policy actions during the financial crisis.

[4]Whereas risk refers to situations in which probabilities can be assigned to all possible outcomes of a situation, Knightian uncertainty refers to situations in which outcomes cannot be associated with a uniquely determined probability i.e. neither the outcomes nor the probabilities associated with the outcomes are known.

[5]See Claessens (2014); Sixteenth Jacques Polak Annual Research Conference (2015).

news and their implications were not always easy to understand, giving way to speculation and interpretation instead. Forward guidance, one of the new unconventional policy tools that the Federal Reserve has relied on to fulfil its dual mandate since 2008, makes the most compelling case for the fact that information is both important and carries a substantial risk of being misinterpreted.

Against this background, this chapter studies the implications of ambiguous information, which is modelled using the framework of learning under ambiguity of Epstein and Schneider (2007), for investment decisions by firms. The contagion theory explored in this chapter operates through the real channel of corporate investment insofar as evidence suggests that private investment has been a major contributor to the output dynamics observed in the aftermath of the Lehman shock. The model features a representative firm that undertakes investment to maximize profits. The return on investment is affected by random exogenous shocks that cannot be perfectly observed. Before investing, the firm chooses how much information to observe about the shocks affecting investment returns. Importantly, I assume that the information that the firm observes about the shocks is ambiguous: the precision of signals is not exactly known but is only known to lie in a range of possible signal precisions. The firm can reduce, at a cost, the range of precisions associated with a signal, i.e. its degree of ambiguity. My model departs from the framework proposed by Epstein and Schneider (2007) in that the degree of ambiguity of a signal is the endogenous outcome of an information acquisition problem rather than being exogenosuly given.

The representative firm in the model is assumed to be averse to ambiguity, which I model using the Minimum Expected Utility (MEU) framework of Gilboa and Schmeidler (1989). Specifically, the firm lacks the confidence to assign unique probabilities to all relevant events and instead evaluates actions using a worst-case probability chosen from a set of multiple probabilities. A key implication of combining worst-case evaluation as in Gilboa and Schmeidler (1989) with learning under ambiguity as in Epstein and Schneider (2008) is an asymmetric response to news whereby the worst-case likelihood used to interpret a signal depends on the nature of the signal itself. This implies that good news is interpreted as very imprecise and is given less weight, while bad news is interpreted as very precise and given more weight. Beliefs are endogenous and depend on the nature of signals, and as a consequence the firm behaves cautiously by reacting more strongly to bad relative to good news.

The model predicts that the ambiguity of the information that the firm chooses to receive in a state of the world optimally decreases with the ex-ante probability of the state. Furthermore, information ambiguity is zero in highly anticipated states of the world. In other words, the firm optimally choose to learn more about states that are deemed to be more likely. This follows from the fact that benefit of acquiring better information increases with the ex-ante likelihood of a state, but the cost does not. Consequently, uncertainty regarding the interpretation of information increases endogenously in highly unexpected states of nature, which causes the ambiguity-averse firm to behave cautiously by reacting more strongly to bad news than to good news. However, in highly anticipated states of the world there is no uncertainty regarding the interpretation of signals. The firm's behaviour no longer exhibits ambiguity-aversion, and good and bad news affect conditional actions in a symmetric fashion. Thus, the model predicts that ambiguity-averse behavior is state-dependent: the firm does not behave cautiously in all states of the world, but only in rare states that are expected to occur with low probability.

The model captures the intuition that decision-makers behave cautiously following exceptional events and its prediction is in line with growing empirical evidence documenting stronger reactions to bad news relative to good news (Conrad, Cornell, and Landsman, 2002; Andersen et al., 2007; Bollerslev and Todorov, 2011; Ben-David et al., 2013; Williams, 2014; Kearney and Liu, 2014; Zhou, 2015; Kuhnen, 2015; Li, Tiwari, and Tong, 2016). This asymmetric response to news is a prediction that is unique to a model incorporating ambiguity-aversion and would not obtain in a model that incorporates risk-aversion only as in that case good and bad news would be given equal weights.

The model provides a micro-foundation for crises-contingent theories of contagion, which are based on the idea that contagion occurs when shocks are transmitted through channels that change or are only active during crises; in contrast, the transmission of shocks that occurs through stable channels that exist at all times does not constitute contagion, but only interdependence (Forbes and Rigobon, 2001; Pericoli and Sbracia, 2003; Forbes, 2012). The notion of contagion adopted in this chapter is that of a crisis-contingent change in the transmission of a shock to the real economy, where a crisis is defined as a state occurring with low probability. In my model, contagion is the outcome of learning about an underlying state when there is uncertainty about the interpretation of information, and the channel of contagion is beliefs. The model predicts that decision-makers behave cautiously during crises because the ambiguity of the information they act on increases endogenously. This is because when information is costly it is not optimal to prepare, through learning, for states of the world that are expected to occur with low probability, such as crises. Relative to a no-ambiguity benchmark, the endogenous ambiguity model I propose predicts that the transmission of negative shocks is amplified, while the transmission of positive shocks is attenuated as the degree of anticipation of these shocks decreases.

I also relate my model to the extensive literature that defines contagion as comovement in excess of some benchmark (i.e. normal or tranquil times, absence of frictions) and tests for contagion manifested as excess correlation (King and Wadhwani, 1990; Forbes and Rigobon, 2002; Karolyi, 2003; Pericoli and Sbracia, 2003; Jotikasthira et al., 2012; Bekaert et al., 2014). I do so by extending the baseline model to a two-country setup, which allows examining cross-country conditional investment correlations. I find that relative to the no-ambiguity benchmark negative shocks induce excess comovement when the information regarding these shocks is ambiguous. On the other hand, positive shocks induce less comovement relative to the no-ambiguity benchmark when the information regarding them is ambiguous. These effects are stronger the more unanticipated the shocks. Furthermore, the model predicts that correlations are asymmetric in good and bad states of the world, which is in line with market evidence (Ang and Chen, 2002; Connolly and Wang, 2003; Yuan, 2005; Boyer et al., 2006; Ozsoy, 2013).

2.1.1 Related literature

This chapter is mainly related and contributes to the literature on contagion, and the literature on ambiguity.

The literature on contagion has yet to reach a consensus on how exactly to define contagion and what are the channels through which shocks are transmitted. The definitions of contagion vary across studies and even aspects over which there was initial agreement have evolved over time, highlighting our limited understanding of the contagion phenomenon. In the early literature, global shocks such as changes in global or US interest rates, risk or liquidity were not classified as contagion. However, academic papers analyzing the spread of the 2007-2008 Global Financial Crisis (GFC) have found that global shocks have played a key role in the transmission of this crisis (Calomiris, Love, and Peria, 2010; Chudik and Fratzscher, 2011a; Forbes and Warnock, 2012; Fratzscher, 2012a; Eichengreen, Mody, Nedeljkovic, and Sarno, 2012; Forbes, 2012). In fact most of these papers avoid using the term contagion due to the difficulty in reconciling this global transmission mechanism with the bilateral linkages that were previously the focus of research.

Notwithstanding the disagreement, a widely accepted view regarding what constitutes contagion is the one popularized by Forbes and Rigobon (2002). According to this view contagion occurs if shocks are transmitted through mechanisms or linkages that are active only during crises, while the transmission of shocks that occurs through stable mechanisms that exist at all times constitutes interdependence. Worth noting is the fact that the debate on how exactly to define contagion is not just academic but has important implications for measuring contagion and evaluating policy responses: while contagion warrants policy intervention, interdependence does not.

The model I propose in this chapter can be framed in the context of the literature on crisis-contingent theories of contagion.[6] The idea underlying this literature is that the mechanism through which shocks are transmitted change during crises. The model essentially provides an information based micro-foundation for crisis-contingent contagion theories, as it explains why and how the behavior of decision-makers, and as a consequence the transmission of shocks, changes during crisis episodes. The model shows that when information is costly it is not optimal to prepare for highly unlikely events. Consequently, decision-makers behave differently during crises because the ambiguity of the information they act on increases endogenously in these low probability state.

The GFC has highlighted that the risks of contagion have not declined, as was argued in the mid 2000s, but instead they have just changed (Forbes, 2012). In addition to emphasizing the importance of global shocks, it has brought to light the existence of a poorly understood type of contagion, one that is triggered by unanticipated events and fuelled by uncertainty about how events will play out on account of a lack of precedent. A model which highlights the importance of unexpected events in generating contagion is proposed by Oh (2013), who studies a model in which contagion of a liquidity crisis between two unrelated firms occurs because of learning activity within a common creditor pool, and shows that contagion is more likely if the triggering event occurs with low probability. Another paper that is related to mine is Kannan and Köhler-Geib (2009), who propose a model of international contagion in which the degree of anticipation of crises, through its impact on investor uncertainty, determines the occurrence of contagion. More specifically, the incidence of surprise crises in other countries leads investors to doubt the accuracy of their information gathering technology and this increases the probability of a crisis in the home country. However, uncertainty in this model only refers to the variance of signal noise, i.e. the signal precision, and the model does not incorporate elements of Knightian uncertainty or ambiguity, which represents the central part of my model.

The study of ambiguity was motivated by the classic Ellsberg (1961) experiments, which have shown that the difference between risky and ambiguous situations is behaviourally meaningful. More specifically, given two urns, one with a known composition of 50 red and 50 black balls (risky urn) and one with an unknown composition of 100 balls in total (ambiguous urn), agents prefer to bet on a red draw from the risky urn as opposed to the ambiguous urn, and they also prefer to bet on a black draw from the risky urn as opposed to the ambiguous one. Such preference is inconsistent with the existence of a single prior on the composition of the ambiguous urn and cannot be rationalized using the standard economic model of decision making under risk, the Subjective

[6] See Forbes and Rigobon (2001) for an overview of this literature

Expected Utility (SEU) axiomized by Savage (1972).[7] Ellsberg's conjecture has been repeatedly confirmed in a large number of subsequent experimental studies, thus lending support to the study of ambiguity-averse preferences.[8]

The substantial body of experimental evidence has motivated an extensive theoretical literature that aims to explain empirically observed phenomena using ambiguity-aversion. A large number of applications of models of ambiguity to finance focus on the implications for asset pricing and portfolio choice.[9] Epstein and Schneider (2008), the paper that I build on, is part of this literature. They propose the model of learning from ambiguous signals and show how such signals induce an asymmetric response to news. The focus is on assessing the impact of ambiguous information on stock prices, in terms of contribution to risk premia as well as negative skewness in returns. I depart from their model by endogenising the ambiguity of information using costly information acquisition i.e. the degree of ambiguity of a signal is the endogenous outcome of an information acquisition problem rather than being exogenosuly given. Additionally, I focus on the implications of endogenously ambiguous information for corporate investment and the transmission of shocks.

This chapter is part of the growing body of literature that uses ambiguity to explain events from the GFC (Caballero and Krishnamurthy, 2008; Routledge and Zin, 2009; Uhlig, 2010; Easley and O'Hara, 2010; Guidolin and Rinaldi, 2010; Caballero and Simsek, 2013; Pritsker, 2013; Cukierman and Izhakian, 2015; Dicks and Fulghieri, 2019). The paper most related to mine is Dicks and Fulghieri (2019), who also show that ambiguity in itself, through its impact on beliefs, is a new source of contagion. In their model the channel of contagion is a change in beliefs which endogenously depends on the composition of portfolios such that bad news on one asset class induces investors to hold worse expectations on other asset classes as well. In contrast, in my model beliefs endogenously depend on the nature of the signals such that bad news are interpreted as being very precise and incorporated more in conditinal actions whereas good news are interpreted as very imprecise and given less weight. Furthermore, whereas Dicks and Fulghieri (2019) incorporate uncertainty in a Diamond and Dybvig (1983) model and focus on systemic bank runs, the focus in my model is on real investment as I incorporate uncertainty in a simple model of corporate

[7]Essentially, the idea at the core of the Ellsberg pardox is that when an agent has insufficient or inadequate information to form a unique prior, then the agent considers a set of plausible probability distributions and not a single one.

[8] Trautmann and Van De Kuilen (2015) provide an extensive review of the experimental work on ambiguity attitudes and conclude that notwithstanding the caveats on the role of the elicitation methods and the robustness of ambiguity aversion, "there is clear evidence that on average, and across various elicitation methods, ambiguity aversion is the typical qualitative finding".

[9]Epstein and Schneider (2010) and Guidolin and Rinaldi (2013) provide an excellent review of this literature, whose common emerging theme is that ambiguity-averse agents command a discount for holding ambiguous assets and take more conservative positions.

investment.

Despite the extensive theoretical literature on ambiguity in finance, there is relatively little empirical evidence of Ellsberg-type ambiguity attitude outside the lab. However, initial evidence suggests that experimental measures of ambiguity correlate with behaviour outside the lab, supporting the external validity of the ambiguity-aversion concept (Anderson, Ghysels, and Juergens, 2009; Antoniou, Harris, and Zhang, 2015). Particularly relevant are those empirical studies that document an asymmetric response to news using aggregate stock market data (Ozsoy, 2013; Zhou, 2015), data on mutual fund flows (Li et al., 2016), and earnings announcements (Williams, 2014).

The rest of the chapter proceeds as follows. Section 2.2 outlines the model, and the main results are discussed in Section 2.3. Section 2.4 extends the baseline model to a two-country setup and relates its predictions to the empirical literature on contagion manifested as excess correlation. Section 2.5 concludes and the Appendix contains all proofs and derivations.

2.2 Model

This section formally introduces the mechanism through which information becomes endogenously ambiguous. It outlines a baseline model of learning under ambiguity in which decision-makers can reduce, at a cost, the degree of ambiguity of information to be received in a given state of the world. The basic results is that the optimal degree of information ambiguity decreases with the degree of anticipation of the state of the world.

2.2.1 Structure of the economy

There is a risk-neutral, ambiguity-averse representative firm in the economy. The firm undertakes investment with an aim to maximize expected profits. Realized profits are

$$\pi = \lambda f - C(\lambda) \tag{2.1}$$

where f is the return on investment, λ is the chosen level of investment, and $C(\lambda)$ is the cost of investment. The cost function $C(\lambda)$ is increasing and convex in the scale of investment, λ, i.e. $C'(\lambda) > 0$ and $C''(\lambda) > 0$. It is assumed to take the form $\frac{\lambda^2}{2}$.

The return on investment, f, is given by

$$f = \mu + \epsilon \tag{2.2}$$

where μ is a constant, interpreted as the mean return on investment, and ϵ is a normally distributed random variable, interpreted as a random exogenous shock affecting investment returns and further referred to simply as a shock.

The shock affecting investment returns is realised but cannot be perfectly observed when the firm chooses its investment. Instead, the firm chooses how much information to observe about the shock, before investing. The sequence of events is illustrated below.

information chosen	information observed | take action	payoff realized
$t = 1$	$t = 2$	$t = 3$

The model can thus be broken down into three periods 1, 2 and 3. In the first period the representative firm chooses its information. In the second period, the firm observes the chosen information and optimally decides on a level of investment. In the third period the payoff of the investment is realized and utility is consumed. The firm's objective function is to maximize date-1 expected utility given by

$$U_1 \equiv \; E_1[u_1(E_2[u_2(\pi)])] \tag{2.3}$$

where $E_i[\cdot]$ and $U_i[\cdot]$ denote the expected value and expected utility, respectively, conditional on the information available at time i.

I assume that the representative firm in the economy is averse to ambiguity. Preferences are described using the Minimum Expected Utility (MEU) axiomized Gilboa and Schmeidler (1989), which represents the standard model of decision-making in the presence of ambiguity. Firms operating under the MEU framework lack the confidence to assign unique probabilities to all relevant events and instead they evaluate actions using a worst-case probability chosen from a set of multiple probabilities

$$\min_{p \in \mathcal{P}} E^p[\cdot]. \tag{2.4}$$

In contrast, under the canonical model of decision making under risk, the Subjective Expected Utility (SEU) axiomized by Savage (1972), firms are able to assign unique probabilities to all relevant events, and evaluate actions using a unique probability measure, interpreted as their subjective probability or belief

$$E^p[\cdot].$$

2.2.2 Information structure

The firm acquires ambiguous signals about the shock affecting investment returns in any given state of the world. It can reduce the degree of ambiguity of the signals at a cost.

The signal conveying information about the exogenous shock affecting returns is assumed to be ambiguous

$$s = \epsilon + \epsilon_s, \quad \epsilon_s \sim \mathcal{N}(0, \sigma_s^2), \quad \sigma_s^2 \in [\underline{\sigma_s^2}, \overline{\sigma_s^2}]. \tag{2.5}$$

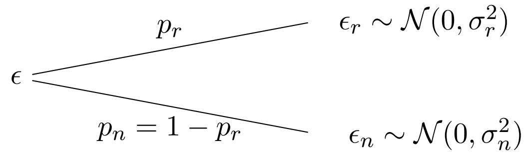
The fact that the variance of the signal noise is not precisely known, but it is known only to lie in an interval makes the signal ambiguous and captures the idea that the interpretation of information can sometimes be difficult. In modelling terms, this means that the signal is related to the parameter of interest by a family or set of likelihoods traced by the variance of the signal noise, σ_s^2. The assumption of a set of likelihoods accommodates the possibility that the precision of the signal is not exactly known to the firm, but it is only known to lie in an interval or set of possible signal precisions $\left[1/\overline{\sigma_s^2}, 1/\underline{\sigma_s^2}\right]$. In the benchmark case of a firm that does not perceive signals as ambiguous, the set of possible signal noise variances collapses to a singleton, σ_s^2. I assume that the interval of signal noise variances contemplated by the firm in my model is centred around this value, σ_s^2, and the size of this belief set depends on the value of an underlying parameter $\zeta \in [0, 1]$, which measures the ambiguity of information. More specifically, the bounds of the set of signal noise variances are defined as

$$\underline{\sigma_s^2} = (1 - \zeta)\sigma_s^2 \tag{2.6}$$

$$\overline{\sigma_s^2} = (1 + \zeta)\sigma_s^2. \tag{2.7}$$

Introducing the parameter ζ represents a departure from the original Epstein and Schneider (2008) model of learning under ambiguity and is endogenously determined in my model as the outcome of an information acquisition problem. The information ambiguity parameter ζ essentially governs the range of possible interpretations associated with the signal and it captures the intuition that the firm might be uncertain regarding the exact interpretation or information content of a signal. I assume that the firm can reduce the range of possible signal interpretations, i.e. the degree of ambiguity of the signal, at a cost K. The information acquisition cost $K(\zeta)$ is decreasing and convex in information ambiguity, ζ, i.e. $K'(\zeta) < 0$ and $K''(\zeta) > 0$. This can be thought of as the firm investing in resources that would facilitate the interpretation and understanding of information.

The firm is endowed with prior beliefs that the shock ϵ affecting investment returns can be in two possible states of nature: a low-probability, high-volatility state interpreted as rare times or a crisis, and a high-probability, low-volatility state of nature, interpreted as normal times. The prior over the shock in the rare state is $\epsilon_r \sim \mathcal{N}(0, \sigma_r^2)$, while the prior associated with the normal state is $\epsilon_n \sim \mathcal{N}(0, \sigma_n^2)$, where $\sigma_n^2 < \sigma_r^2$. This setup can thus be graphed as:



where $p_r \in (0, 0.5)$ is the probability of occurrence of the rare state of nature, $p_n = 1 - p_r$ is the probability that the normal state would occur.

Upon the occurrence of any state, the firm receives an ambiguous signal about the shock affecting the return on investment in that state

$$s_i = \epsilon_i + \epsilon_{s_i}, \quad \epsilon_{s_i} \sim \mathcal{N}\left(0, \sigma_{s_i}^2\right), \quad \sigma_{s_i}^2 \in [(1-\zeta_i)\sigma_s^2, (1+\zeta_i)\sigma_s^2], \quad i \in \{r, n\}.$$

The degree of ambiguity of the signal received in each state of nature is determined by the firm's informational investment decision. Specifically, the firm can choose the degree of information ambiguity $\zeta_i,\ i \in \{r, n\}$, at a cost $K(\zeta_i)$ that does not vary with the ex-ante likelihood of the state.

Learning follows standard Bayesian rules for updating normal random variables. However, since the signal is described by a family of probability distributions, a set of posterior mean beliefs is obtained and it is given by

$$E[f|s_i] \quad = \quad \mu + \frac{\sigma_i^2}{\sigma_i^2 + \sigma_{s_i}^2} s_i, \quad \sigma_{s_i}^2 \in \left[(1-\zeta_i)\sigma_s^2,\ (1+\zeta_i)\sigma_s^2\right], \quad i \in \{r, n\}. \tag{2.8}$$

For any given state of nature, even though there is a unique prior over the parameter of interest, updating leads to a non-degenerate set of posteriors, due to the fact that the signal is related to the parameter by a family of likelihoods. This can be thought of as a situation in which there is confidence in the initial information about the environment but firms entertain multiple theories about how signals were generated and what their information content is. Hence, although there is no ambiguity ex-ante, the ambiguous signal introduces ambiguity into posterior beliefs about fundamentals.

Define the information content of a signal as

$$\gamma_i(\sigma_{s_i}^2) \equiv \frac{\sigma_i^2}{\sigma_i^2 + \sigma_{s_i}^2}, \quad \sigma_{s_i}^2 \in \left[(1-\zeta_i)\sigma_s^2,\ (1+\zeta_i)\sigma_s^2\right], \quad i \in \{r, n\}. \tag{2.9}$$

This ratio measures the fraction of prior variance that is resolved by the signal.[10] In any given state $i \in \{r, n\}$, the information content of the signal decreases as the signal noise variance, $\sigma_{s_i}^2$, ranges over the interval created by the ambiguity parameter, ζ_i, i.e. $\left[(1-\zeta_i)\sigma_s^2,\ (1+\zeta_i)\sigma_s^2\right]$. The signal is most informative when signal noise variance is lowest, i.e. $\sigma_{s_i}^2 = (1-\zeta_i)\sigma_s^2$, and this represents the upper bound on information content, denoted $\overline{\gamma_i} \equiv \gamma_i\left((1-\zeta_i)\sigma_s^2\right)$. On the other hand, the least informative signal is obtained when signal noise variance is highest, i.e. $\sigma_{s_i}^2 = (1+\zeta_i)\sigma_s^2$, and this represents the lower bound on information content, denoted $\underline{\gamma_i} \equiv \gamma_i\left((1+\zeta_i)\sigma_s^2\right)$. If there is no information ambiguity, the signal noise variance take the value around which the interval is centred, i.e. $\sigma_{s_i}^2 = \sigma_s^2$, and the information content of the signal is denoted $\gamma_i \equiv \gamma_i\left(\sigma_s^2\right)$. The signal information content in these three cases is thus given by

$$\gamma_i = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_s^2}, \quad \overline{\gamma_i} = \frac{\sigma_i^2}{\sigma_i^2 + (1-\zeta_i)\sigma_s^2}, \quad \underline{\gamma_i} = \frac{\sigma_i^2}{\sigma_i^2 + (1+\zeta_i)\sigma_s^2}, \quad i \in \{r, n\}. \tag{2.10}$$

The relation between these variables is $\underline{\gamma_i} < \gamma_i < \overline{\gamma_i}$. The difference $(\overline{\gamma_i} - \underline{\gamma_i})$ measures the confidence in the information content of a signal; the greater the difference, the lower the confidence regarding the interpretation of information. An increase in information ambiguity maps into a loss of confidence as it shifts the lower bound on information content lower and the upper bound higher i.e. $\frac{\partial \underline{\gamma_i}}{\partial \zeta_i} < 0$ and $\frac{\partial \overline{\gamma_i}}{\partial \zeta_i} > 0$. In the benchmark case in which there is no information ambiguity, and the upper and lower bounds coincide $(\overline{\gamma_i} = \underline{\gamma_i} = \gamma_i)$ and the firm knows exactly how much information the signal contains.

2.2.3 Solving the model

A solution to the model is: a choice of information ambiguity ζ_i to maximize date-1 expected utility (2.3), given ambiguity-averse preferences described by (2.4), and subject to the constraint that the degree of information ambiguity satisfies $\zeta_i \in [0, 1]$, and rational expectations about the date-2 (conditional) investment; posterior beliefs which are formed according to Bayes' law (2.9) given a signal about the shock; a choice of investment that maximizes expected utility, given the signal realization.

[10]It is essentially the regression coefficient obtained by regressing the signal on the shock.

The model is solved using backward induction. First, given an arbitrary level of information ambiguity, the firm decides the optimal level of investment. Second, given the optimal level of investment for each information ambiguity level, the firm chooses the optimal degree of information ambiguity.

2.3 Results

In this section, I derive the optimal choice of information ambiguity, and discuss the implications of ambiguous information for investment and the transmission of shocks.

2.3.1 Optimal investment

At date 2 the firm chooses the optimal level of investment taking information ambiguity as given. Since the following analysis holds for any arbitrary level of information ambiguity, I abstract from differentiating between the rare and normal states and suppress the subscript i for notational convenience. The firm's date-2 problem is

$$\max_{\lambda} U_2 \equiv E_2[u_2(\pi)]$$

where $E_2[\cdot]$ denotes the expected value conditional on the information available at date 2. Recalling the firm's profit function (2.1) and the ambiguity-averse preferences considered (2.4), date-2 utility is given by

$$U_2 = E_2[u_2(\pi)] = \min_{\sigma_s^2} E\left[\pi|s\right] = \lambda \min_{\sigma_s^2} E\left[f|s\right] - \frac{\lambda^2}{2}. \tag{2.11}$$

The first order conditional yields the following optimal level of investment when information ambiguity ζ is taken as given

$$\lambda(\zeta) = \min_{\sigma_s^2} E\left[f|s\right]. \tag{2.12}$$

Thus, for any given information choice, the optimal investment level is proportional to the expected return on investment. Using the Bayesian updating result in (2.9) and the signal information content notation in (2.10), the conditional investment level is given by

$$\lambda(\zeta) = \min_{\sigma_s^2} E\left[f|s\right] \quad = \quad \begin{cases} \mu + \underline{\gamma} s & \text{if} \quad s \geq 0 \\ \mu + \overline{\gamma} s & \text{if} \quad s < 0 \end{cases} \tag{2.13}$$

Ambiguous information introduces cautious behavior, which is defined as asymmetry in conditional actions. More specifically, news are incorporated into beliefs, and as a consequence investment decisions, in an asymmetric fashion: good news ($s \geq 0$) is interpreted as very imprecise and given less weight, while bad news ($s < 0$) is interpreted as very precise and given more weight. This is because ambiguity-averse firms evaluate actions using a worst-case belief that is chosen from a set of conditional probabilities. If an ambiguous signal conveys good news then the worst case is that the signal is unreliable and viceversa. Hence, the worst-case likelihood used to interpret a signal depends on the nature of the signal itself. In other words, beliefs are endogenous and depend on the nature of signals, with the implication that bad news affect conditional actions more than good news.

In order to assess the implications of information ambiguity for belief formation and conditional investment a suitable benchmark is needed for comparison. I consider as benchmark a model in which there is no information ambiguity, i.e. $\zeta = 0$. Under this no-ambiguity benchmark news are given the same weights, i.e. γ, regardless of their nature, and there is no asymmetric response to news

$$\lambda(0) \quad = \quad \mu + \gamma s. \tag{2.14}$$

Figure 2.1 illustrates the optimal level of investment obtained under the ambiguity model (red line) and the no-ambiguity benchmark (blue line) conditional on the signal, s, which takes both positive and negative values.

The first thing to note is that the conditional investment level is lower when information is ambiguous relative to the no-ambiguity benchmark. This follows from the relationship $\underline{\gamma} < \gamma < \overline{\gamma}$. When information is ambiguous, negative news induce a decrease in investment that is excessive compared to the no-ambiguity benchmark, i.e. $\gamma < \overline{\gamma}$. On the other hand, positive news induce an increase in investment that is lower when information is ambiguous than under the benchmark, i.e. $\underline{\gamma} < \gamma$. These results can also be interpreted as amplification of negative shocks and attenuation of positive shocks, respectively. As a consequence, ambiguous information introduces a loss in investment relative to the no-ambiguity benchmark.

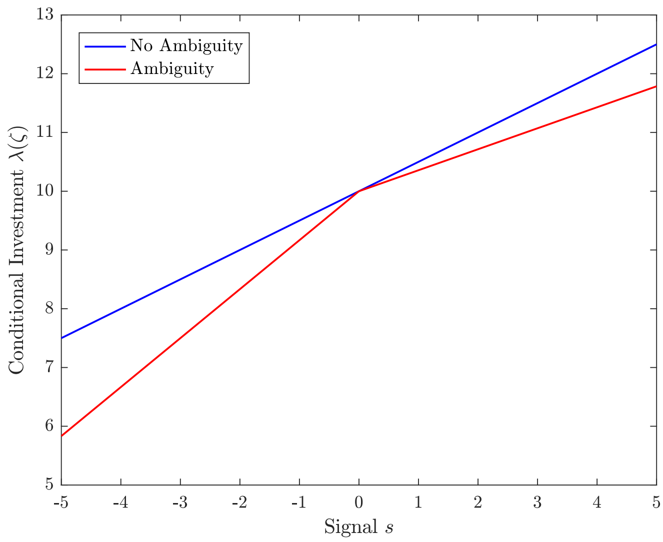


Figure 2.1 Optimal Investment.
The figure illustrates the relationship between the signal s and the conditional level of investment implied by the ambiguity model (red line) and the no-ambiguity benchmark (red line). The parameter values are $\mu = 10$, $\sigma = 1$, $\sigma_s = 1$, $\zeta = 0.8$

Second, note that investment decisions respond more strongly to bad news relative good news. This follows from the relation $\underline{\gamma} < \overline{\gamma}$, which implies that the decrease in investment in response negative news is stronger than the increase in response to positive news. This asymmetry in conditional investment levels is also reflected in asymmetry in conditional investment correlations. Section 2.4 extends the model to a two-country setup and discusses the implications of ambiguous information in terms of cross-country conditional investment correlations. Worth noting is that the asymmetric response to news is unique to a model incorporating ambiguity-aversion. It would not obtain in a model that incorporates risk-aversion only as in that case good and bad news would be given equal weights.[11] However, when ambiguity-averse firms process news of uncertain quality, investment decisions respond more strongly to bad news than good news.

Substituting the optimal investment choice (2.12) into the objective function (2.15) delivers the indirect date-2 utility of having any degree of information ambiguity ζ and investing optimally

$$U_2(\zeta) = \left(\min_{\sigma_s^2} E\left[f|s\right]\right)^2 - \frac{1}{2}\left(\min_{\sigma_s^2} E\left[f|s\right]\right)^2 = \frac{1}{2}\left(\min_{\sigma_s^2} E\left[f|s\right]\right)^2. \tag{2.15}$$

[11]To see this, consider the date 1 valuation of a risk-averse SEU firm. Risk preferences are represented by a von Neumann-Morgenstern utility function with constant absolute risk aversion (CARA) i.e. $u(\pi) = -e^{-\phi\pi}$, where $\phi > 0$ is the degree of absolute risk aversion, and the shock affecting returns follows $\epsilon \sim \mathcal{N}(0, \sigma^2)$

$$\lambda = E[f|s] - \frac{1}{2}\phi V[f|s] = \mu + \gamma s - \frac{1}{2}\phi\left(1-\gamma\right)\sigma^2.$$

2.3.2 Optimal information choice

At date 0 the firm chooses the optimal degree of information ambiguity taking as given the optimal level of investment. Specifically, the firm maximizes the probability weighted sum of the expected payoff in each state, net the cost of acquiring information relevant for that state. The firm's date-1 problem is

$$\max_{\zeta_r,\zeta_n} \; p_r U_1(\zeta_r) + p_n U_1(\zeta_n) - K(\zeta_r) - K(\zeta_n) \tag{2.16}$$

$$\text{s.t. } \; 0 \leq \zeta_i \leq 1, \quad i \in \{r, n\} \tag{2.17}$$

where $U_1(\zeta_i),\ i \in \{r, n\}$ denotes the date-1 expected utility in the rare and normal states, respectively, and $K(\zeta_i),\ i \in \{r, n\}$ denotes the cost of acquiring information for the rare and normal states. Recalling the date-2 utility of having any posterior beliefs and investing optimally in (2.15), and the ambiguity-averse preferences considered (2.4), date-1 expected utility is given by

$$U_1(\zeta_i) \equiv E_1[u_1(U_2(\zeta_i))] = \min_{\sigma^2_{s_i}} E[U_2(\zeta_i)], \quad i \in \{r, n\}.$$

The solution for the optimal degree of information ambiguity is summarised in the following proposition, and represents the key mechanism in the model.

Proposition 1. *(State-dependent ambiguous information)*
Information ambiguity decreases with the ex-ante probability of the state i.e. $\frac{\partial \zeta_i}{\partial p_i} < 0,\ i \in \{r, n\}$.
Information ambiguity is zero in highly anticipated states of the world i.e. $\zeta_i = 0$ *if* $p_i > \overline{p_i}$ *where the threshold probability* $\overline{p_i}$ *is such that* $\frac{\partial K(\zeta_i)}{\partial \zeta_i}\big|_{\zeta_i=0} = \overline{p_i}\frac{\partial U_1(\zeta_i)}{\partial \zeta_i}\big|_{\zeta_i=0},\ i \in \{r, n\}$.
Information ambiguity is maximum in highly unanticipated states of the world i.e. $\zeta_i = 1$ *if* $p_i < \underline{p_i}$ *where the threshold probability* $\underline{p_i}$ *is such that* $\frac{\partial K(\zeta_i)}{\partial \zeta_i}\big|_{\zeta_i=1} = \underline{p_i}\frac{\partial U_1(\zeta_i)}{\partial \zeta_i}\big|_{\zeta_i=1},\ i \in \{r, n\}$.

Proof. See Appendix **A.1**.

The optimal degree of information ambiguity in a state of the world decreases with the ex-ante probability of occurrence of that state. This is due to the fact that the benefit of improving the quality of information to be received in a state of the world increases with the probability of the state but the cost does not. As a consequence, the firm is less likely to prepare, by reducing information ambiguity, for rare states that are expected to occur with low probability. In fact, for sufficiently unlikely states, the firm will optimally choose not to reduce information ambiguity at all, thus setting $\zeta = 1$. On the other hand, the firm will completely reduce the ambiguity of information to be received in highly anticipated states of the world by setting $\zeta = 0$. The

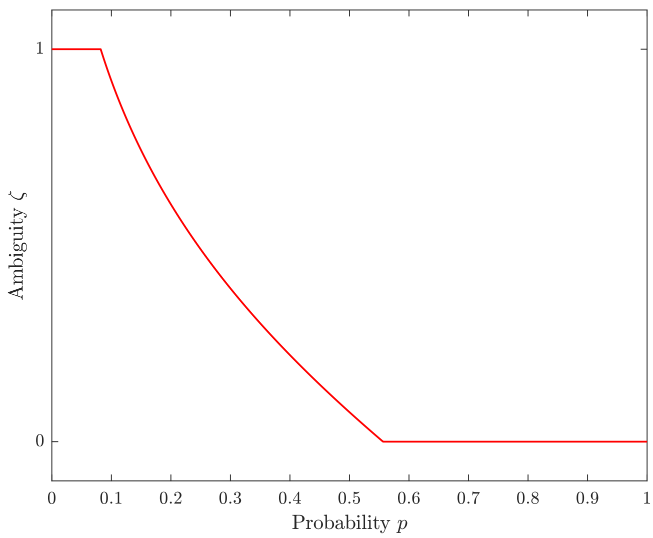


Figure 2.2 Optimal Information Ambiguity.
The figure illustrates the relationship between the optimal level of information ambiguity in a state of nature and the ex-ante probability of occurrence of that state. The cost of acquiring information is assumed to take the form $K(\zeta) = k\exp(-\zeta)$. The parameter values are $\mu = 10$, $\sigma = 1$, $\sigma_s = 1$, $k = 1.5$

implication is that information ambiguity endogenously increases in rare times, while in normal times there is no uncertainty regarding the interpretation of information.

Figure 2.2 provides a graphical illustration of this result. It plots the optimal level of information ambiguity in a state of nature against the probability of occurrence of that state. It illustrates that for very low ex-ante probabilities of state occurrence information ambiguity is set at its maximum as the firm optimally chooses not prepare for rare events. At the interior optimum, the degree of information ambiguity acquired for a state of nature decreases with the probability of occurrence of the state. For states expected to occur with sufficiently high probability the ambiguity of information is minimized and set at zero.

An important implication of this result is that ambiguity-averse behavior is state-dependent: the firm does not behave cautiously in all states of the world, but only in rare states that are expected to occur with sufficiently low probability. This result can be understood by recalling that a non-zero degree of information ambiguity entails an asymmetric response to news, something that I term cautious or ambiguity-averse behavior. Proposition 1 shows that in states of the world that occur with sufficiently low probability information ambiguity is non-zero, which commands a stronger response to bad relative to good news. On the other hand, in states that are expected to occur with high probability, there is no information ambiguity and as a consequence behavior no longer exhibits ambiguity-aversion, as good and bad news affect conditional actions symmetrically.

Corollary 1. *(State-dependent cautious behavior) The firm does not behave cautiously in all states of the world, but only in sufficiently unanticipated states of the world.*

Interpreted in terms of crisis and normal times i.e. highly unanticipated and anticipated states of the world, respectively, this corollary says that during crises the firm behaves cautiously by reacting more strongly to negative than positive news. During normal times, however, the firm's behavior no longer exhibits cautiousness, such that negative and positive news affect conditional actions in a symmetric fashion.

2.3.3 Implications for the transmission of shocks

This section explores the relationship between the degree to which shocks affect investment decisions and the degree of anticipation of these shocks. To that end, I define the shock transmission mechanism as the change in investment due to a shock i.e. $\frac{\partial \lambda}{\partial \epsilon} = \gamma^*$, where γ^* denotes the random variable that is equal to γ when information is not ambiguous, $\underline{\gamma}$ when information is ambiguous and positive, and $\overline{\gamma}$ when information is ambiguous and negative. Thus, the shock transmission mechanism is essentially a measure of the degree to which shocks are incorporated into investment decisions.

The model predicts that the shock transmission mechanism is state-dependent. In low probability states of the world the shock transmission mechanism is asymmetric: the transmission of unanticipated negative shocks is amplified, while the transmission of unanticipated positive shocks is attenuated. However, in high probability states of the world the shock transmission mechanism is the same in case of both negative and positive shocks.

Proposition 2. *(Crisis-contingent contagion) The shock transmission mechanism changes during crises. Specifically, the transmission of negative shocks is amplified, and the transmission of positive shocks is attenuated as the degree of anticipation of these shocks decrease i.e. if $p < \overline{p}$ then*

$$\frac{\partial \gamma^*}{\partial p} = \begin{cases} > 0 & if \;\; s \geq 0 \\ < 0 & if \;\; s < 0 \end{cases} .$$

Proof. See Appendix **A.2**.

Figure 2.3 provides a graphical illustration of how the transmission of shocks depends on their degree of anticipation. If the degree of anticipation of a shock is sufficiently low so that the ambiguity of information regarding that shock is non-zero, i.e. $p < \overline{p}$, then the shock transmission

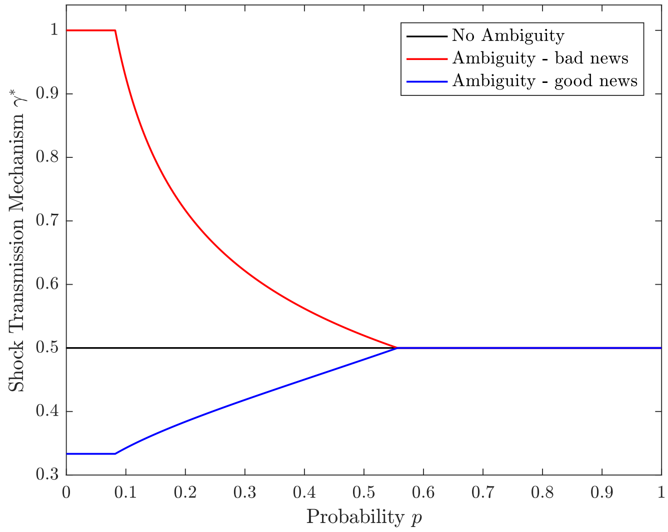


Figure 2.3 Shock Transmission Mechanism.
The figure illustrates the relationship between the ex-ante probability of occurrence of a state and shock transmission mechanism in that state conditional on ambiguous good news (blue line), ambiguous bad news (red line), and news that are not ambiguous (black line). The cost of acquiring information is assumed to take the form $K(\zeta) = k \exp(-\zeta)$. The parameter values are $\mu = 10$, $\sigma = 1$, $\sigma_s = 1$, $k = 1.5$.

is asymmetric: negative shocks are amplified and positive shocks are attenuated. This shock transmission asymmetry decreases with the degree of anticipation of the shock, and it vanishes for sufficiently high probabilities. Consequently, in high probability states of the world the shock transmission mechanism is no longer asymmetric.

Worth noting is that, the amplification of unanticipated shocks could be also obtained with risk-averse decision-makers and noisy information. Presumably one could think that foreign information becomes more noisy and/or investors become more risk averse after negative unexpected shocks. However, information processing would be symmetric. These represent two alternative and non mutually exclusive channels whose validity remains an empirical question. A testable implication is that following unexpected events bad news affect conditional actions more than good news. The effect is stronger the more unanticipated the shock.

2.4 Two-country model

The synchronized collapse of financial and macroeconomic aggregates lies at the center of a large literature on contagion. This extensive literature defines contagion as comovement in excess of some benchmark (i.e. normal or tranquil times, absence of frictions) and tests for contagion manifested as excess correlation. This section extends the model to a two-country setup in order to examine cross-country conditional investment correlations and relate its predictions to the literature on contagion manifested as excess comovement.

I focus on cross-sectional differences in information ambiguity across firms operating in two different countries and so I abstract from differentiating between the rare and the normal states of nature, and instead use the subscript j to differentiate between the two countries; the analysis holds regardless of the state of nature. The interest lies in exploring heterogeneity and deriving the conditions under which a synchronized collapse in investment occurs, which in the context of this model manifests as an increase in correlations between the levels of investment.

I consider the case of two representative firms which operate in different countries, and which have profit functions that follow the same functional form as in the baseline model (2.1). Each firm's return on investment now depends on the fundamentals of the economy in which the firm operates. Fundamentals, in turn, are affected both by idiosyncratic shocks, which are specific to each country, as well as an aggregate shock, which affects fundamentals in multiple countries

$$f_j = \mu_j + \epsilon_{d_j} + \alpha_j \epsilon_g, \quad j \in \{1, 2\} \tag{2.18}$$

where μ_j is the mean level of fundamentals in the two countries $j \in \{1, 2\}$, ϵ_{d_j} are country-specific, idiosyncratic shocks - interpreted as domestic shocks, ϵ_g is a common, aggregate shock - interpreted as a global shock, and $\alpha_j \in [0, 1]$ are country-specific exposures or sensitivities to the global shock. The domestic and global shocks are mutually independent and normally distributed with zero mean.

Each firm knows the structure of the economy in which it operates but cannot perfectly observe the shocks affecting fundamentals. Instead, each firm learns about the realization of the domestic and global shocks by acquiring independent signals. The signal conveying information about each domestic shock ϵ_{d_j}, henceforth referred to as the domestic signal, is assumed to be noisy

$$x_j = \epsilon_{d_j} + \epsilon_{x_j}, \quad \epsilon_{x_j} \sim \mathcal{N}(0, \sigma_{x_j}^2), \quad j \in \{1, 2\} \tag{2.19}$$

where all the shocks are mutually independently distributed.

On the other hand, the signal conveying information about the global shock ϵ_g, henceforth referred to as the global signal, is assumed to be ambiguous

$$y = \epsilon_g + \epsilon_y, \quad \epsilon_y \sim \mathcal{N}(0, \sigma_y^2), \quad \sigma_y^2 \in [(1-\zeta)\sigma^2, (1+\zeta)\sigma^2]. \tag{2.20}$$

The assumption that the precision of the domestic signal is objectively known, but the precision of the global signal is only known to lie in a range of signal precisions, captures firm's uncertainty regarding the interpretation of foreign information.[12] The notion that agents have an advantage in interpreting domestic news but are unsure as to how accurate their foreign information-processing technology lies at the heart of an extensive literature on international finance and home bias.

Note that I assume that both firms observe the same public signal about the shock. However, I allow for the possibility that they might interpret it differently. Mounting evidence based on survey data on expectations has shown that there is substantial heterogeneity in how decision-makers perceive the current and form inferences about the future economic conditions. As noted by Acemoglu, Chernozhukov, and Yildiz (2016), in most cases the source of disagreement among decision-makers does not seem to be differences in observations or experiences but differences in interpreting the available data. Consequently, I consider the case of differentially interpreted public information by assuming that firms observe the same public global signal but may not all interpret it identically. Technically, this refers to a setup in which the distribution of error in public signal is heterogeneous. Following the literature on differential interpretation of information (Kim and Verrecchia, 1994; Kandel and Pearson, 1995), I model this by endowing firms with different likelihoods about the error in the public signal. More specifically, I assume that the representative firm in country j believes that signal noise follows $\epsilon_{y_j} \sim \mathcal{N}(0, \sigma^2_{y_j})$. Consequently, firm j's estimate of the global shock conditional on the public signal can be backed out as

$$\epsilon_{g_j} = y - \epsilon_{y_j}, \quad \epsilon_{y_j} \sim \mathcal{N}(0, \sigma^2_{y_j}), \quad \sigma^2_{y_j} \in \left[(1-\zeta_j)\sigma^2_j, (1+\zeta_j)\sigma^2_j\right] \tag{2.21}$$

and the resulting distribution associated with the global shock is $\epsilon_{g_j} \sim \mathcal{N}(0, \sigma^2_{g_j})$, $j \in \{1, 2\}$. On the other hand, the firm is endowed with the prior that the distribution associated with the domestic shock is $\epsilon_{d_j} \sim \mathcal{N}(0, \sigma^2_{d_j})$, $j \in \{1, 2\}$.

This can also be thought of a situation in which firms form estimates based on the public signal and an error but may not disagree about the error. Note that I combine differential interpretation of public information, which translates into heterogeneous signal noise variances, σ^2_j, with ambiguous information, which translates into heterogeneous degrees of information ambiguity, ζ_j, thus yielding the interval $\sigma^2_{y_j} \in \left[(1-\zeta_j)\sigma^2_j, (1+\zeta_j)\sigma^2_j\right]$, $j \in \{1, 2\}$. Given a signal, this information structure

[12] Worth noting is that the chapter's focus on learning about domestic versus global shocks is motivated by the international contagion framework considered, and by evidence that global shocks have played an important role during the global financial crisis (Forbes and Warnock, 2012; Fratzscher, 2012b; Chudik and Fratzscher, 2011b; Calomiris, Love, and Pería, 2012; Eichengreen et al., 2012), but the model more generally applies to more or less understood shocks.

implies a clear trade-off between uncertainty related to the global shock and uncertainty related to the signal noise. The precision of the signals allows firms some latitude in interpreting public signals. The intuition is that although firms might observe the same public signal, they can interpret it differently and as a consequence entertain different beliefs about the impact of the global shock on fundamentals.

The solution method follows the same steps as in the baseline model. As before, the optimal level of investment for each of the firms is given by the expected level of fundamentals, conditional on the available information

$$\lambda_j = E[f|x_j, y] = \mu_j + \gamma_{x_j} x_j + \alpha_j \gamma_{y_j} y, \quad j \in \{1, 2\} \tag{2.22}$$

where γ_{x_j} and γ_{x_j} denote the information content of the domestic and global signals, respectively

$$\gamma_{x_j} = \frac{\sigma^2_{d_j}}{\sigma^2_{d_j} + \sigma^2_{x_j}} \quad \text{and} \quad \gamma_{y_j} = \frac{\sigma^2_{g_j}}{\sigma^2_{g_j} + \sigma^2_{y_j}}, \quad \sigma^2_{y_j} \in \left[(1-\zeta_j)\sigma^2_j, (1+\zeta_j)\sigma^2_j\right], \quad j \in \{1, 2\}. \tag{2.23}$$

The conditional correlation between the levels of investment undertaken by the representative firms in the two countries, λ_1 and λ_2, is given by

$$\rho \equiv [\lambda_1, \lambda_2 | x_1, x_2, y] = \frac{\alpha_1 \alpha_2 \gamma^*_{y_1} \gamma^*_{y_2} \left(1 - \frac{2}{\pi}\right) V[y]}{\sqrt{\gamma^2_{x_1}\ V[x_1] + \alpha_1^2 \gamma^{*2}_{y_1} \left(1 - \frac{2}{\pi}\right) V[y]} \sqrt{\gamma^2_{x_2}\ V[x_2] + \alpha_2^2 \gamma^{*2}_{y_2} \left(1 - \frac{2}{\pi}\right) V[y]}} \tag{2.24}$$

where $\gamma^*_{y_j}$ denotes the random variable that is equal to γ_{y_j} when information is not ambiguous, $\underline{\gamma_{y_j}}$ when information is ambiguous and positive, and $\overline{\gamma_{y_j}}$ when information is ambiguous and negative. To the extend that the information content of the signal, $\gamma^*_{y_j}$, is driven by the degree of information ambiguity perceived by the two firms, it will be a key driver of the correlation between the levels of conditional investment undertaken by these firms.

Proposition 3. *(Asymmetric Contagion) Conditional correlation is increasing with the information content of the global signal i.e.* $\frac{\partial \rho}{\partial \gamma^*_{y_i}} > 0$.

Proof. See Appendix **A.3**.

The relation $\underline{\gamma_y} < \gamma_y < \overline{\gamma_y}$ implies that relative to the no-ambiguity benchmark, correlations are higher conditional on negative news when information is ambiguous. On the other hand, ambiguous information means that correlations conditional on positive news are lower relative to the no-ambiguity benchmark. This result is also illustrated in Figure 2.4, which illustrates the relationship between information ambiguity and the correlation between the optimal levels of

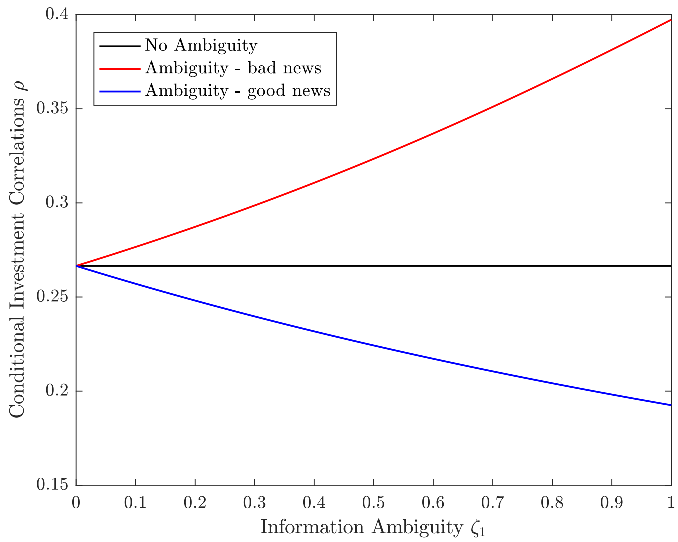


Figure 2.4 Conditional Investment Correlations.
The figure illustrates the relationship between the degree of information ambiguity in country 1, ζ_1, and the correlation between the optimal levels of investment in the two countries conditional on ambiguous good news (blue line), ambiguous bad news (red line), and news that are not ambiguous (black line). The examples abstract from information ambiguity in country 2 and considers a symmetric equilibrium by setting $\zeta_2 = 0$, $\alpha_j = 1$, $\sigma_{d_j} = 1$, $\sigma_{x_j} = 1$, $\sigma_{g_j} = 1$, $\sigma_{y_j} = 1$, $j \in \{1, 2\}$.

investment in the two countries conditional on ambiguous positive news (blue line), ambiguous negative news (red line) and news that are not ambiguous (black line). The example considers the case in which information ambiguity is non-zero in one country, and abstracts from information ambiguity in the other country.

Interpreted in terms of shocks, these results indicate that relative to the no-ambiguity benchmark negative shocks induce excess comovement when the information regarding those shocks is ambiguous. On the other hand, positive shocks induce less comovement relative to the no-ambiguity benchmark when the information regarding them is ambiguous. Also worth noting is the asymmetry in correlations: correlations are higher conditional on negative rather than positive shocks when the information regarding these shocks is ambiguous. This prediction is in line with market evidence (Ang and Chen, 2002; Connolly and Wang, 2003; Yuan, 2005; Boyer et al., 2006; Ozsoy, 2013).

The excess correlation prediction is in line with the empirical literature on contagion, which defines contagion as excess correlation relative to some benchmark which reflects fundamentals or lack of frictions (King and Wadhwani, 1990; Claessens, Dornbusch, and Park, 2001; Forbes and Rigobon, 2002; Karolyi, 2003; Pericoli and Sbracia, 2003; Bekaert, Harvey, and Ng, 2005; Jotikasthira et al., 2012; Bekaert et al., 2014). However, as emphasized by Forbes and Rigobon

(2002) when the increase in correlation is caused by an increase in volatility, the resulting effect should not be classified as contagion but as interdependence. The excess correlation obtained in my model qualifies as contagion because it is a consequence of an (endogenous) increase in information ambiguity and is not due to an increase in the volatility of fundamentals. In other words, it is not induced by an actual change in fundamentals but result solely from a change in the ambiguity of information.

2.5 Concluding remarks

Building on the premise that information is important in shaping economic outcomes but carries a substantial risk of being misinterpreted, I propose a model in which costly information acquisition leads to endogenous variation in information ambiguity and cautious behaviour. In my model uncertainty regarding the interpretation of information increases endogenously in the aftermath of highly unanticipated events. This causes ambiguity-averse decision-makers to behave cautiously by reacting more strongly to bad news than to good news. However, in highly anticipated states of the world there is no uncertainty regarding the interpretation of signals. Decision-makers' behaviour no longer exhibits ambiguity-aversion, and good and bad news affect conditional actions in a symmetric fashion. The model explains why and how the behaviour of decision-makers changes during crises, and delivers predictions that are in line with observed market outcomes.

Appendix

A.1 ***Proof of Proposition 1:***

The Lagrangian associated with problem (2.16) - (2.17) is

$$\mathcal{L} = \sum_{i \in \{r,n\}} p_i U_1(\zeta_i) - K(\zeta_i) + \vartheta_i \zeta_i + \varphi_i (1-\zeta_i)$$

The first order conditions are

$$\begin{aligned} p_i \frac{\partial U_1(\zeta_i)}{\partial \zeta_i} - \frac{\partial K(\zeta_i)}{\partial \zeta_i} + \vartheta_i - \varphi_i &= 0 \\ \vartheta_i \zeta_i &= 0 \\ \varphi_i (1-\zeta_i) &= 0 \end{aligned}$$

Implicit differentiation yields

$$\frac{\partial p_i}{\partial \zeta_i} \frac{\partial U_1(\zeta_i)}{\partial \zeta_i} + p_i \frac{\partial^2 U_1(\zeta_i)}{\partial \zeta_i^2} = \frac{\partial^2 K(\zeta_i)}{\partial \zeta_i^2}$$

$$\Rightarrow \frac{\partial p_i}{\partial \zeta_i} = \frac{\frac{\partial^2 K(\zeta_i)}{\partial \zeta_i^2} - p_i \frac{\partial^2 U_1(\zeta_i)}{\partial \zeta_i^2}}{\frac{\partial U_1(\zeta_i)}{\partial \zeta_i}}$$

I want to prove that information ambiguity decrease with the ex-ante probability of the state i.e. $\frac{\partial \zeta_i}{\partial p_i} < 0$. I assume that marginal cost of improving information quality / reducing information ambiguity is higher than the marginal benefit i.e. $\frac{\partial^2 K(\zeta_i)}{\partial \zeta_i^2} > \frac{\partial^2 U_1(\zeta_i)}{\partial \zeta_i^2}$, which is a technical assumption to guarantee an interior solution. Hence, the proof relies on showing that $\frac{\partial U_1(\zeta_i)}{\partial \zeta_i} < 0$.

To evaluate the date-1 expected utility function, $U_1(\zeta_i) = \min_{\sigma^2_{s_i}} E[U_2(\zeta_i)]$, recall the expression for date-2 utility, $U_2(\zeta_i)$, in (2.15)

$$\begin{aligned} U_1(\zeta_i) &= \min_{\sigma^2_{s_i}} E[U_2(\zeta_i)] = \min_{\sigma^2_s} E\left[\frac{1}{2}\left(\min_{\sigma^2_{s_i}} E\left[f|s_i\right]\right)^2\right] \\ &= \frac{1}{2}\min_{\sigma^2_{s_i}} \left(E\left[\min_{\sigma^2_{s_i}} E\left[f|s_i\right]\right]^2 + V\left[\min_{\sigma^2_{s_i}} E\left[f|s_i\right]\right]\right) \end{aligned}$$

Recalling the expression for the conditional mean, $\min\limits_{\sigma^2_{s_i}} E\left[f|s_i\right]$, in (2.9) yields

$$\begin{aligned} U_1(\zeta_i) &= \frac{1}{2}\min_{\sigma^2_{s_i}} \left[E\left[\mu + \underline{\gamma_i}s_i + \left(\overline{\gamma_i} - \underline{\gamma_i}\right)\min\{s_i, 0\}\right]^2 + V\left[\mu + \underline{\gamma_i}s_i + \left(\overline{\gamma_i} - \underline{\gamma_i}\right)\min\{s_i, 0\}\right]\right] \\ &= \frac{1}{2}\min_{\sigma^2_{s_i}} \left[(\mu + E[\gamma_i s_i])^2 + V[\gamma_i s_i]\right] \end{aligned}$$

To compute the moments of $\gamma_i s_i$, where γ_i is a random variable which takes the value $\underline{\gamma_i}$ if $s_i \geq 0$ and it takes the value $\overline{\gamma_i}$ if $s_i < 0$, I make use of the formulas for the moments of truncated normal distributions. Specifically, for a normally distributed random variable $x \sim \mathcal{N}(0, \sigma_x^2)$ it holds that $E[x|x \geq 0] = \frac{2}{\sqrt{2\pi}}\sigma_x$, $E[x|x < 0] = -\frac{2}{\sqrt{2\pi}}\sigma_x$ and $E[x^2|x \geq 0] = E[x^2|x < 0] = \sigma_x^2$. The expected value of the term $\gamma_i s_i$ is thus given by

$$\begin{aligned} E[\gamma_i s_i] = E[\gamma_i E[s_i|\gamma_i]] &= \frac{1}{2}\underline{\gamma_i}E[s_i|\gamma_i = \underline{\gamma_i}] + \frac{1}{2}\overline{\gamma_i}E[s_i|\gamma_i = \overline{\gamma_i}] \\ &= \frac{1}{2}\underline{\gamma_i}E[s_i|s_i \geq 0] + \frac{1}{2}\overline{\gamma_i}E[s_i|s_i < 0] \\ &= -(\overline{\gamma_i} - \underline{\gamma_i})\frac{1}{\sqrt{2\pi}}\sqrt{V[s_i]} \end{aligned}$$

and the variance is given by

$$\begin{aligned} V[\gamma_i s_i] &= E[\gamma_i^2 s_i^2] - E[\gamma_i s_i]^2 = E[\gamma_i^2 E[s_i^2|\gamma_i]] - E[\gamma_i s_i]^2 \\ &= \frac{1}{2}\underline{\gamma_y}^2 E[s_i^2|\gamma_i = \underline{\gamma_i}] + \frac{1}{2}\overline{\gamma_i}^2 E[s_i^2|\gamma_i = \overline{\gamma_i}] - E[\gamma_i s_i]^2 \\ &= \frac{1}{2}\underline{\gamma_i}^2 E[s_i^2|s_i \geq 0] + \frac{1}{2}\overline{\gamma_i}^2 E[s_i^2|s_i < 0] - E[\gamma_i s_i]^2 \\ &= \frac{1}{2}\underline{\gamma_i}^2 V[s_i] + \frac{1}{2}\overline{\gamma_i}^2 V[s_i] - \frac{1}{2\pi}(\overline{\gamma_i} - \underline{\gamma_i})^2 V[s_i] \\ &= \frac{1}{2}\left(\overline{\gamma_i}^2 + \underline{\gamma_i}^2 - \frac{1}{\pi}(\overline{\gamma_i} - \underline{\gamma_i})^2\right) V[s_i]. \end{aligned}$$

Therefore, the date-1 utility can be re-written as

$$\begin{aligned} U_1(\zeta_i) &= \frac{1}{2}\min_{\sigma^2_{s_i}} \left[(\mu + E[\gamma_i s_i])^2 + V[\gamma_i s_i]\right] \\ &= \frac{1}{2}\min_{\sigma^2_{s_i}} \left[\left(\mu - (\overline{\gamma_i} - \underline{\gamma_i})\frac{1}{\sqrt{2\pi}}\sqrt{V[s_i]}\right)^2 + \frac{1}{2}\left(\overline{\gamma_i}^2 + \underline{\gamma_i}^2 - \frac{1}{\pi}(\overline{\gamma_i} - \underline{\gamma_i})^2\right) V[s_i]\right] \\ &= \frac{1}{2}\min_{\sigma^2_{s_i}} \left[\mu^2 + \frac{1}{2}(\overline{\gamma_i}^2 + \underline{\gamma_i}^2)V[s_i] - 2\mu(\overline{\gamma_i} - \underline{\gamma_i})\frac{1}{\sqrt{2\pi}}\sqrt{V[s_i]}\right] \end{aligned}$$

where the variance of the signal is $V[s_i] = \sigma_i^2 + \sigma_{s_i}^2$, $\sigma_{s_i}^2 \in \left[(1 - \zeta_i)\sigma_s^2,\ (1 + \zeta_i)\sigma_s^2\right]$, $i \in \{r, n\}$.

Minimizing the term in square brackets with respect to the variance of the signal noise $\sigma_{s_i}^2 \in \left[(1-\zeta_i)\sigma_s^2,\ (1+\zeta_i)\sigma_s^2\right]$ is equivalent to minimizing it with respect to the variance of the signal $V[s_i] = \sigma_i^2 + \sigma_{s_i}^2 \in \left[\sigma_i^2 + (1-\zeta_i)\sigma_s^2,\ \sigma_i^2 + \ (1+\zeta_i)\sigma_s^2\right]$. The first derivative with respect to $V[s_i]$

$$\frac{1}{2}(\overline{\gamma_i}^2 + \underline{\gamma_i}^2) - (\overline{\gamma_i} - \underline{\gamma_i})\frac{\mu}{\sqrt{2\pi}}\frac{1}{\sqrt{V[s_i]}} = 0$$

yields solution[13]

$$V[s_i]^* = \frac{2}{\pi}\frac{\mu^2(\overline{\gamma_i} - \underline{\gamma_i})^2}{(\overline{\gamma_i}^2 + \underline{\gamma_i}^2)^2}.$$

Denote the lower bound of the signal variance interval $\underline{V} \equiv \sigma_i^2 + (1-\zeta_i)\sigma_s^2$ and the upper bound $\overline{V} \equiv \sigma_i^2 + (1+\zeta_i)\sigma_s^2$. The signal variance that minimizes the term in square brackets is thus given by

$$V[s_i] = \begin{cases} \underline{V} = \sigma_i^2 \overline{\gamma_i}^{-1} & \text{if } V[s_i]^* < \underline{V} \\ V[s_i]^* = \frac{2}{\pi}\frac{\mu^2(\overline{\gamma_i} - \underline{\gamma_i})^2}{(\overline{\gamma_i}^2 + \underline{\gamma_i}^2)^2} & \text{if } V[s_i]^* \in [\underline{V}, \overline{V}] \\ \overline{V} = \sigma_i^2 \underline{\gamma_i}^{-1} & \text{if } V[s_i]^* > \overline{V} \end{cases}$$

Therefore, the expression for the date-1 expected utility is

$$U_1(\zeta_i) = \begin{cases} \frac{1}{2}\left(\mu^2 + \frac{\sigma_i^2}{2}\frac{(\overline{\gamma_i}^2 + \underline{\gamma_i}^2)}{\overline{\gamma_i}} - \frac{2\mu\sigma_i}{\sqrt{2\pi}}\frac{(\overline{\gamma_i} - \underline{\gamma_i})}{\sqrt{\overline{\gamma_i}}}\right) & \text{if } V[s_i]^* < \underline{V} \\ \frac{1}{2}\left(\mu^2 - \frac{1}{\pi}\frac{\mu^2(\overline{\gamma_i} - \underline{\gamma_i})^2}{\overline{\gamma_i}^2 + \underline{\gamma_i}^2}\right) & \text{if } V[s_i]^* \in [\underline{V}, \overline{V}] \\ \frac{1}{2}\left(\mu^2 + \frac{\sigma_i^2}{2}\frac{(\overline{\gamma_i}^2 + \underline{\gamma_i}^2)}{\underline{\gamma_i}} - \frac{2\mu\sigma_i}{\sqrt{2\pi}}\frac{(\overline{\gamma_i} - \underline{\gamma_i})}{\sqrt{\underline{\gamma_i}}}\right) & \text{if } V[s_i]^* > \overline{V} \end{cases}$$

I now verify that $\frac{\partial U_1(\zeta_i)}{\partial \zeta_i}$ in each of the three cases. Recall

$$\underline{\gamma_i} = \frac{\sigma_i^2}{\sigma_i^2 + (1+\zeta_i)\sigma_s^2} \Rightarrow \frac{\partial \underline{\gamma_i}}{\partial \zeta_i} = -\frac{\sigma_i^2\sigma_s^2}{(\sigma_i^2 + (1+\zeta_i)\sigma_s^2)^2} = -\frac{\sigma_s^2}{\sigma_i^2}\underline{\gamma_i}^2 \quad \text{(A.1)}$$

$$\overline{\gamma_i} = \frac{\sigma_i^2}{\sigma_i^2 + (1-\zeta_i)\sigma_s^2} \Rightarrow \frac{\partial \overline{\gamma_i}}{\partial \zeta_i} = \frac{\sigma_i^2\sigma_s^2}{(\sigma_i^2 + (1-\zeta_i)\sigma_s^2)^2} = \frac{\sigma_s^2}{\sigma_i^2}\overline{\gamma_i}^2 \quad \text{(A.2)}$$

[13]Note that the second derivative is positive, hence the solution is a minimum.

(a) Consider the case when $V[s_i]^* > \overline{V}$.

$$\begin{aligned}
\frac{\partial U(\zeta_i)}{\partial \zeta_i} &= \frac{\sigma_i^2}{4}\frac{\left(2\overline{\gamma_i}\frac{\partial\overline{\gamma_i}}{\partial\zeta_i}+2\underline{\gamma_i}\frac{\partial\underline{\gamma_i}}{\partial\zeta_i}\right)\underline{\gamma_i}-(\overline{\gamma_i}^2+\underline{\gamma_i}^2)\frac{\partial\underline{\gamma_i}}{\partial\zeta_i}}{\underline{\gamma_i}^2} - \frac{\mu\sigma_i}{\sqrt{2\pi}}\frac{\left(\frac{\partial\overline{\gamma_i}}{\partial\zeta_i}-\frac{\partial\underline{\gamma_i}}{\partial\zeta_i}\right)\sqrt{\underline{\gamma_i}}-(\overline{\gamma_i}-\underline{\gamma_i})\frac{1}{2\sqrt{\underline{\gamma_i}}}\frac{\partial\underline{\gamma_i}}{\partial\zeta_i}}{\underline{\gamma_i}} \\
&= \frac{\sigma_i^2}{4}\frac{\sigma_s^2}{\sigma_i^2}\frac{2\left(\overline{\gamma_i}^3-\underline{\gamma_i}^3\right)\underline{\gamma_i}+(\overline{\gamma_i}^2+\underline{\gamma_i}^2)\underline{\gamma_i}^2}{\underline{\gamma_i}^2} - \frac{\mu\sigma_i}{\sqrt{2\pi}}\frac{\sigma_s^2}{\sigma_i^2}\frac{2\left(\overline{\gamma_i}^2+\underline{\gamma_i}^2\right)\underline{\gamma_i}+(\overline{\gamma_i}-\underline{\gamma_i})\underline{\gamma_i}^2}{2\underline{\gamma_i}\sqrt{\underline{\gamma_i}}} \\
&= \frac{\sigma_s^2}{4}\left[\frac{2(\overline{\gamma_i}^3-\underline{\gamma_i}^3)}{\underline{\gamma_i}}+(\overline{\gamma_i}^2+\underline{\gamma_i}^2)-\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}\frac{2\left(\overline{\gamma_i}^2+\underline{\gamma_i}^2\right)}{\sqrt{\underline{\gamma_i}}}-\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}(\overline{\gamma_i}-\underline{\gamma_i})\sqrt{\underline{\gamma_i}}\right] \\
&= \frac{\sigma_s^2}{2\underline{\gamma_i}}\left[(\overline{\gamma_i}^3-\underline{\gamma_i}^3)-\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}\left(\overline{\gamma_i}^2+\underline{\gamma_i}^2\right)\sqrt{\underline{\gamma_i}}\right]+\frac{\sigma_s^2}{4}\left[(\overline{\gamma_i}^2+\underline{\gamma_i}^2)-\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}(\overline{\gamma_i}-\underline{\gamma_i})\sqrt{\underline{\gamma_i}}\right]
\end{aligned}$$

Condition $V[s_i]^* > \overline{V}$ can be re-written as $\frac{2}{\pi}\frac{\mu^2}{\sigma_i^2}(\overline{\gamma_i}-\underline{\gamma_i})^2\underline{\gamma_i} > (\overline{\gamma_i}^2+\underline{\gamma_i}^2)^2$, which implies that $(\overline{\gamma_i}^2+\underline{\gamma_i}^2)-\sqrt{\frac{2}{\pi}\frac{\mu}{\sigma_i}}(\overline{\gamma_i}-\underline{\gamma_i})\sqrt{\underline{\gamma_i}} < 0$. Thus, the second term in square brackets in negative. The first term in square brackets is negative because $\sqrt{\frac{2}{\pi}\frac{\mu}{\sigma_i}}\left(\overline{\gamma_i}^2+\underline{\gamma_i}^2\right)\sqrt{\underline{\gamma_i}} > \sqrt{\frac{2}{\pi}\frac{\mu}{\sigma_i}}(\overline{\gamma_i}-\underline{\gamma_i})\sqrt{\underline{\gamma_i}}$ and $(\overline{\gamma_i}^3-\underline{\gamma_i}^3) < (\overline{\gamma_i}^2+\underline{\gamma_i}^2)$. This follows from the fact that $\underline{\gamma_i}\in[0,1]$, $\overline{\gamma_i}\in[0,1]$, $\underline{\gamma_i}<\overline{\gamma_i}$. Therefore $(\overline{\gamma_i}^3-\underline{\gamma_i}^3) < (\overline{\gamma_i}^2+\underline{\gamma_i}^2) \;\Leftrightarrow\; \overline{\gamma_i}^2(\overline{\gamma_i}-1) < \underline{\gamma_i}^2(\underline{\gamma_i}+1)$ is true since the left hand side term is negative while the right hand side term is positive.

(b) Consider the case when $V[s_i]^* \in [\underline{V},\overline{V}]$. Date-1 expected utility can be re-written as

$$U(\zeta_i) = \frac{\mu^2}{2}-\frac{\mu^2}{2\pi}\frac{\overline{\gamma_i}^2-2\overline{\gamma_i}\underline{\gamma_i}+\underline{\gamma_i}^2}{\overline{\gamma_i}^2+\underline{\gamma_i}^2} = \frac{\mu^2}{2}-\frac{\mu^2}{2\pi}+\frac{\mu^2}{\pi}\frac{\overline{\gamma_i}\underline{\gamma_i}}{\overline{\gamma_i}^2+\underline{\gamma_i}^2}$$

so that the first order derivative with respect to ambiguity is

$$\begin{aligned}
\frac{\partial U_1(\zeta_i)}{\partial\zeta_i} &= \frac{\mu^2}{\pi}\frac{\left(\frac{\partial\overline{\gamma_i}}{\partial\zeta_i}\underline{\gamma_i}+\frac{\partial\underline{\gamma_i}}{\partial\zeta_i}\overline{\gamma_i}\right)(\overline{\gamma_i}^2+\underline{\gamma_i}^2)-\overline{\gamma_i}\underline{\gamma_i}\left(2\overline{\gamma_i}\frac{\partial\overline{\gamma_i}}{\partial\zeta_i}+2\underline{\gamma_i}\frac{\partial\underline{\gamma_i}}{\partial\zeta_i}\right)}{(\overline{\gamma_i}^2+\underline{\gamma_i}^2)^2} \\
&= \frac{\mu^2}{\pi}\frac{\sigma_s^2}{\sigma_i^2}\frac{\left(\overline{\gamma_i}^2\underline{\gamma_i}-\underline{\gamma_i}^2\overline{\gamma_i}\right)(\overline{\gamma_i}^2+\underline{\gamma_i}^2)-2\overline{\gamma_i}\underline{\gamma_i}(\overline{\gamma_i}^3-\underline{\gamma_i}^3)}{(\overline{\gamma_i}^2+\underline{\gamma_i}^2)^2} \\
&= \frac{\mu^2}{\pi}\frac{\sigma_s^2}{\sigma_i^2}\frac{\overline{\gamma_i}\underline{\gamma_i}\left(\overline{\gamma_i}-\underline{\gamma_i}\right)\left[\overline{\gamma_i}^2+\underline{\gamma_i}^2-2\overline{\gamma_i}^2-2\overline{\gamma_i}\underline{\gamma_i}-2\underline{\gamma_i}^2\right]}{(\overline{\gamma_i}^2+\underline{\gamma_i}^2)^2} \\
&= \frac{\mu^2}{\pi}\frac{\sigma_s^2}{\sigma_i^2}\frac{\overline{\gamma_i}\underline{\gamma_i}\left(\overline{\gamma_i}-\underline{\gamma_i}\right)\left[-\overline{\gamma_i}^2-2\overline{\gamma_i}\underline{\gamma_i}-\underline{\gamma_i}^2\right]}{(\overline{\gamma_i}^2+\underline{\gamma_i}^2)^2} \\
&= -\frac{\mu^2}{\pi}\frac{\sigma_s^2}{\sigma_i^2}\frac{\overline{\gamma_i}\underline{\gamma_i}\left(\overline{\gamma_i}-\underline{\gamma_i}\right)\left(\overline{\gamma_i}+\underline{\gamma_i}\right)^2}{(\overline{\gamma_i}^2+\underline{\gamma_i}^2)^2} < 0
\end{aligned}$$

(c) Consider the case when $V[s_i]^* < \underline{V}$

$$
\begin{aligned}
\frac{\partial U_1(\zeta_i)}{\partial \zeta_i} &= \frac{\sigma_i^2}{4}\frac{\left(2\overline{\gamma_i}\frac{\partial\overline{\gamma_i}}{\partial\zeta_i}+2\underline{\gamma_i}\frac{\partial\underline{\gamma_i}}{\partial\zeta_i}\right)\overline{\gamma_i}-(\overline{\gamma_i}^2+\underline{\gamma_i}^2)\frac{\partial\overline{\gamma_i}}{\partial\zeta_i}}{\overline{\gamma_i}^2}-\frac{\mu\sigma_i}{\sqrt{2\pi}}\frac{\left(\frac{\partial\overline{\gamma_i}}{\partial\zeta_i}-\frac{\partial\underline{\gamma_i}}{\partial\zeta_i}\right)\sqrt{\overline{\gamma_i}}-(\overline{\gamma_i}-\underline{\gamma_i})\frac{1}{2\sqrt{\overline{\gamma_i}}}\frac{\partial\overline{\gamma_i}}{\partial\zeta_i}}{\overline{\gamma_i}}\\
&= \frac{\sigma_i^2}{4}\frac{\sigma_s^2}{\sigma_i^2}\frac{2\left(\overline{\gamma_i}^3-\underline{\gamma_i}^3\right)\overline{\gamma_i}-(\overline{\gamma_i}^2+\underline{\gamma_i}^2)\overline{\gamma_i}^2}{\overline{\gamma_i}^2}-\frac{\mu\sigma_i}{\sqrt{2\pi}}\frac{\sigma_s^2}{\sigma_i^2}\frac{2(\overline{\gamma_i}^2+\underline{\gamma_i}^2)\overline{\gamma_i}-(\overline{\gamma_i}-\underline{\gamma_i})\overline{\gamma_i}^2}{2\overline{\gamma_i}\sqrt{\overline{\gamma_i}}}\\
&= \frac{\sigma_s^2}{4}\left[\frac{2(\overline{\gamma_i}^3-\underline{\gamma_i}^3)}{\overline{\gamma_i}}-(\overline{\gamma_i}^2+\underline{\gamma_i}^2)-\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}\frac{2\left(\overline{\gamma_i}^2+\underline{\gamma_i}^2\right)}{\sqrt{\overline{\gamma_i}}}+\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}(\overline{\gamma_i}-\underline{\gamma_i})\sqrt{\overline{\gamma_i}}\right]\\
&= \frac{\sigma_s^2}{2\overline{\gamma_i}}\left[(\overline{\gamma_i}^3-\underline{\gamma_i}^3)-\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}\left(\overline{\gamma_i}^2+\underline{\gamma_i}^2\right)\sqrt{\overline{\gamma_i}}\right]-\frac{\sigma_s^2}{4}\left[(\overline{\gamma_i}^2+\underline{\gamma_i}^2)-\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}(\overline{\gamma_i}-\underline{\gamma_i})\sqrt{\overline{\gamma_i}}\right]
\end{aligned}
$$

Condition $V[s_i]^* < \underline{V}$ implies $\frac{2}{\pi}\frac{\mu^2}{\sigma_i^2}(\overline{\gamma_i}-\underline{\gamma_i})^2\overline{\gamma_i} < (\overline{\gamma_i}^2+\underline{\gamma_i}^2)^2$, which implies that $(\overline{\gamma_i}^2+\underline{\gamma_i}^2) - \sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}(\overline{\gamma_i}-\underline{\gamma_i})\sqrt{\overline{\gamma_i}} > 0$. Thus, the second term in square brackets is positive. The first term in square brackets is negative because because $\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}\left(\overline{\gamma_i}^2+\underline{\gamma_i}^2\right)\sqrt{\underline{\gamma_i}} > \sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_i}(\overline{\gamma_i}-\underline{\gamma_i})\sqrt{\underline{\gamma_i}}$ and $(\overline{\gamma_i}^3-\underline{\gamma_i}^3) < (\overline{\gamma_i}^2+\underline{\gamma_i}^2)$.

Corner solutions

If the non-negativity constraint binds i.e. $\zeta_i = 0$ then the first FOC becomes

$$p_i\frac{\partial U_1(\zeta_i)}{\partial\zeta_i}\bigg|_{\zeta_i=0} - \frac{\partial K(\zeta_i)}{\partial\zeta_i}\bigg|_{\zeta_i=0} + \vartheta_i = 0$$

which is equivalent to

$$\frac{\partial K(\zeta_i)}{\partial\zeta_i}\bigg|_{\zeta_i=0} \geq p_i\frac{\partial U_1(\zeta_i)}{\partial\zeta_i}\bigg|_{\zeta_i=0}$$

because $\vartheta_i \geq 0$. Define a critical point p_i^* such that $\frac{\partial K(\zeta_i)}{\partial\zeta_i}\big|_{\zeta_i=0} = p_i^*\frac{\partial U_1(\zeta_i)}{\partial\zeta_i}\big|_{\zeta_i=0}$. For values of p_i smaller than this threshold probability p_i^* the inequality holds. For values of p_i beyond this threshold probability p_i^*, the ambiguity of information ζ_i is constrained at the minimum i.e. $\zeta_i = 0$. In other words, for sufficiently probable states of the world there is no information ambiguity (information ambiguity is completely reduced).

If the other complementary slackness constraint binds i.e. $\zeta_i = 1$ then the first FOC becomes

$$p_i\frac{\partial U_1(\zeta_i)}{\partial\zeta_i}\bigg|_{\zeta_i=1} - \frac{\partial K(\zeta_i)}{\partial\zeta_i}\bigg|_{\zeta_i=1} - \varphi_i = 0$$

Since it has to be that the Lagrange multiplier is positive i.e. $\varphi_i \geq 0$ we have

$$\left.\frac{\partial K(\zeta_i)}{\partial \zeta_i}\right|_{\zeta_i=1} \leq p_i \left.\frac{\partial U_1(\zeta_i)}{\partial \zeta_i}\right|_{\zeta_i=1}$$

For probability values p_i smaller than a defined threshold p_i^* for which $\frac{\partial K(\zeta_i)}{\partial \zeta_i}\big|_{\zeta_i=1} = p_i^* \frac{\partial U_1(\zeta_i)}{\partial \zeta_i}\big|_{\zeta_i=1}$, the ambiguity of information is constrained at the maximum i.e. $\zeta_i = 1$. For values of p_i larger than this threshold the inequality is satisfied. In other words, for sufficiently improbable states of the world there is no learning (information ambiguity is not reduced at all).

A.2 *Proof of Proposition 2:*

Define the shock transmission mechanism as the change in investment induced by a shock $\frac{\partial \lambda}{\partial \epsilon} = \gamma^*$, where γ^* is a random variable taking the value γ when information is not ambiguous, $\underline{\gamma}$ when information is ambiguous and positive, and $\overline{\gamma}$ when information is ambiguous and negative. The interest lies in examining how the shock transmission mechanism changes with the degree of anticipation of the shock.

$$\frac{\partial \gamma^*}{\partial p} = \frac{\partial \gamma^*}{\partial \zeta}\frac{\partial \zeta}{\partial p} = \begin{cases} > 0 & \text{if} \quad s \geq 0 \\ < 0 & \text{if} \quad s < 0 \end{cases}$$

since $\frac{\partial \zeta}{\partial p} < 0$ as shown in Proposition 1, and $\frac{\partial \underline{\gamma}}{\partial \zeta} < 0$, $\frac{\partial \overline{\gamma}}{\partial \zeta} > 0$ (A.1)-(A.2)

$$\frac{\partial \gamma^*}{\partial \zeta} = \begin{cases} \frac{\partial \underline{\gamma}}{\partial \zeta} = -\frac{\sigma^2 \sigma_s^2}{(\sigma^2 + (1+\zeta)\sigma_s^2)^2} < 0 & \text{if} \quad s \geq 0 \\ \frac{\partial \overline{\gamma}}{\partial \zeta} = \frac{\sigma^2 \sigma_s^2}{(\sigma^2 + (1-\zeta)\sigma_s^2)^2} > 0 & \text{if} \quad s < 0 \end{cases}$$

A.3 *Proof of Proposition 3:*

To derive the conditional correlation between investment levels across countries, first derive the variance and covariance of the conditional country-specific level of investment in (2.22):

$$\begin{aligned} V[\lambda_j | x_j, y] &= V[\mu_j + \gamma_{x_j} x_j + \alpha_j \gamma_{y_j} y] = \gamma_{x_j}^2 V[x_j] + \alpha_j^2 V[\gamma_y y] \\ &= \gamma_{x_j}^2 V[x_j] + \alpha_j^2 \gamma_{y_j}^{*2} V[y | y \lesseqgtr 0] \end{aligned}$$

$$
\begin{aligned}
\text{cov}[\lambda_1, \lambda_2|x_1, x_2, y] &= \text{cov}[\mu_1 + \gamma_{x_1}x_1 + \alpha_1\gamma_{y_1}y,\ \mu_2 + \gamma_{x_2}x_2 + \alpha_2\gamma_{y_2}y|y \lesseqgtr 0] \\
&= \alpha_1\alpha_2\text{cov}[\gamma_{y_1}y, \gamma_{y_2}y|y \lesseqgtr 0] \\
&= \alpha_1\alpha_2 \left(E[\gamma_{y_1}\gamma_{y_2}y^2|y \lesseqgtr 0] - E[\gamma_{y_1}y|y \lesseqgtr 0]E[\gamma_{y_2}y|y \lesseqgtr 0]\right) \\
&= \alpha_1\alpha_2\gamma_{y_1}\gamma_{y_2} \left(E[y^2|y \lesseqgtr 0] - E[y|y \lesseqgtr 0]E[y|y \lesseqgtr 0]\right) \\
&= \alpha_1\alpha_2\gamma^*_{y_1}\gamma^*_{y_2}V[y|y \lesseqgtr 0]
\end{aligned}
$$

The conditional variance of the global signal, $V[y|y \lesseqgtr 0]$, can be derived by making use of formulas for the moments of truncated normal distributions.[14] Specifically, $V[y|y \geq 0] = V[y|y < 0] = \left(1 - \frac{2}{\pi}\right) V[y]$, so that the conditional investment correlation is

$$
\rho \equiv [\lambda_1, \lambda_2|x_1, x_2, y] = \frac{\alpha_1\alpha_2\gamma^*_{y_1}\gamma^*_{y_2}\left(1 - \frac{2}{\pi}\right) V[y]}{\sqrt{\gamma^2_{x_1}\ V[x_1] + \alpha_1^2\gamma^{*2}_{y_1}\left(1 - \frac{2}{\pi}\right) V[y]}\sqrt{\gamma^2_{x_2}\ V[x_2] + \alpha_2^2\gamma^{*2}_{y_2}\left(1 - \frac{2}{\pi}\right) V[y]}}
$$

The derivative of this conditional correlation with respect to the information content extracted by the firm in country 1 is

$$
\frac{\partial\rho}{\partial\gamma^*_{y_1}} = \frac{\alpha_1\alpha_2\gamma^2_{x_1}\gamma^*_{y_2}V[x_1]\left(1 - \frac{2}{\pi}\right) V[y]}{(\gamma^2_{x_1}\ V[x_1] + \alpha_1^2\gamma^{*2}_{y_1}\left(1 - \frac{2}{\pi}\right) V[y])^{3/2}\sqrt{\gamma^2_{x_2}\ V[x_2] + \alpha_2^2\gamma^{*2}_{y_2}\left(1 - \frac{2}{\pi}\right) V[y]}} > 0.
$$

[14]Specifically, for a normally distributed random variable $x \sim \mathcal{N}(0, \sigma_x^2)$ it holds that $E[x|x \geq 0] = \sqrt{\frac{2}{\pi}}\sigma_x$, $E[x|x < 0] = -\sqrt{\frac{2}{\pi}}\sigma_x$ and $E[x^2|x \geq 0] = E[x^2|x < 0] = \sigma_x^2$.

Chapter 3

Testing Inequality Restrictions in Multifactor Asset-Pricing Models[1]

3.1 Introduction

Multifactor asset-pricing models seek to explain cross-sectional differences in expected asset returns in terms of exposures to one or more sources of systematic risk. The capital asset pricing model (CAPM) of Treynor (1961), Sharpe (1964), Lintner (1965), and Mossin (1966) is the cornerstone of modern asset-pricing theory. It posits that the expected return on an asset is proportional to its covariance with the return on aggregate market wealth. The CAPM is a single-period model which, as shown by Fama (1970), can be treated as if it holds intertemporally only if the preferences and future investment opportunities are constant. However, as shown by Merton (1973), the CAPM does not hold in an intertemporal setting when the investor faces a state-dependent investment opportunity set.

The intertemporal CAPM (ICAPM) of Merton (1973) extends the CAPM to a multi-period framework. Unlike the single-period maximizer of the CAPM who does not take into account events beyond the current period, the intertemporal maximizer of the ICAPM also takes into account the relationship between current returns and returns that will be available in the future. This gives rise to additional sources of risk that an investor has to hedge against. According to the ICAPM, the expected return on an asset is not only proportional to the asset's covariance with the market portfolio return, but also to the asset's covariance with changes in the investment opportunity set. Following Cochrane (2005), the cross-sectional equilibrium relation between expected return and

[1] This chapter is joint work with Cesare Robotti and Jay Shanken, and adapted from the working paper "Testing Inequality Restrictions in Multifactor Asset-Pricing Models", 2015.

risk in the context of the ICAPM can be expressed as follows:

$$E_t(R_{i,t+1}) - R_{f,t} = \lambda \text{Cov}_t(R_{i,t+1}, R_{m,t+1}) + \lambda_z \text{Cov}_t(R_{i,t+1}, \Delta z_{t+1}), \tag{3.1}$$

where R_i denotes the expected return on asset i, R_f denotes the risk-free rate, R_m is the return on aggregate market wealth or simply the market portfolio, λ is the market price of covariance risk (which corresponds to the coefficient of relative risk aversion of the representative investor), λ_z is the intertemporal price of covariance risk, and Δz denotes innovations in state variables that capture uncertainty about future investment opportunities.

The second term of equation (3.1) is the expected return component that arises as compensation for unexpected changes in the investment opportunity set. These changes in investment opportunities are captured by the state variables z, which are essentially variables that describe the conditional distribution of returns that will be available in the future. The fact that the ICAPM does not explicitly identify these state variables has prompted Fama (1991) to label it a "fishing license", in the sense that it essentially allows applied researchers to choose from a wide range of potential risk factors and use the ICAPM as a theoretical justification for relatively ad-hoc empirical specifications. Although the ICAPM does not explicitly tell us what the state variables are, there are several restrictions that candidate state variables need to satisfy for their innovations to be considered candidate risk factors in an ICAPM setting (Maio and Santa-Clara, 2012). First, the candidate state variables should predict changes in the investment opportunity set. Second, if a state variable predicts positive (negative) changes in investment opportunities in the time-series, then its innovation should earn a positive (negative) intertemporal price of covariance risk in the cross-sectional relation. Finally, the market price of risk should be an economically plausible estimate of the coefficient of relative risk aversion of the representative investor.

Therefore, it would seem that the ICAPM cannot be used as a theoretical justification for any multifactor model as it imposes several restrictions on the time-series and cross-sectional behavior of the candidate state variables and their innovations. However, the mere existence of these theoretical restrictions would not bear much weight against the claim that the ICAPM is a "fishing license" if these restrictions were indeed rarely violated in practice. Current research provides mixed evidence as to whether the ICAPM restrictions are satisfied in empirical tests of multifactor asset-pricing models. Maio and Santa-Clara (2012) consider eight popular multifactor asset-pricing specifications and find that most of these models are not consistent with an ICAPM interpretation. Lutzenberger (2015) replicates the Maio and Santa-Clara (2012) study for the

European stock market and reaches similar conclusions. On the other hand, Boons (2016) focuses on state variables that forecast macroeconomic activity and the prices of covariance risk that the innovations in these state variables earn in a large cross-section of individual stocks. He finds consistency of the considered models with the ICAPM. Cooper and Maio (2018) study traded factor models and focus in particular on a number of recent prominent models incorporating investment and profitability factors. They find that the models studied are "to a large degree compatible with the ICAPM framework" but none of them satisfies all the restrictions imposed by the ICAPM. Barroso, Boons, and Karehnke (2019) focus on non-traded factor models, and account for time-variation in the risk premia by analyzing the conditional asset-pricing implications of the ICAPM. They find that conditional risk premia in a large cross-section of individual stocks are consistent in sign with how the state variables predict consumption growth in the time-series. Despite the lack of consensus, what all these studies have in common is the fact that the consistency of the models with the restrictions imposed by the ICAPM is assessed through a visual exercise whereby the researcher compares the signs of the slope estimates in the predictive regressions with the signs of the price of covariance risk estimates in the cross-sectional regressions.

We present a rigorous econometric framework to formally evaluate the consistency of a multifactor model with the time-series and cross-sectional restrictions imposed by the ICAPM and provide an in-depth empirical analysis to demonstrate the relevance of our methodological results. We focus on the empirical performance of nine multifactor models using two different sets of test assets and different estimation methods. First, we run multiple predictive ordinary least squares (OLS) time-series regressions to estimate the slope coefficients associated with the state variables. This allows us to obtain an *a priori* knowledge of the sign restrictions that the prices of covariance risk must satisfy for the various multifactor models to receive an ICAPM interpretation. Second, we estimate the prices of covariance risk by running two-pass cross-sectional regressions of average realized excess returns on the estimated covariances between the test asset returns and the innovations in the state variables (see Kan, Robotti, and Shanken (2013)). The estimation of the prices of covariance risk is performed using OLS, generalized least squares (GLS), and weighted least squares (WLS) weighting schemes. Third, we develop and implement a multivariate inequality test, based on Wolak (1987, 1989), to determine whether the signs of the prices of covariance risk are consistent with the signs of the slope coefficients in the predictive regressions. This allows us to go beyond the common practice of informally comparing the signs of the estimated coefficients in the predictive regressions with the signs of the estimated prices of covariance risk in the cross-sectional regressions. Our methods account for the estimation error in the covariances and for the fact that the consistency of a multifactor model with the implications of the ICAPM should be evaluated using tests of joint sign restrictions across factors. Importantly, in the estimation of the prices of covariance risk and in the tests of the sign restrictions, we employ asymptotic standard errors that are robust to potential model misspecification in addition to the traditional standard errors computed under the assumption that the model is correctly specified.

Our testing methodology delivers conclusions that are substantially different from the ones reached by following the common practice of visually comparing sign estimates that is used in the existing literature. If we simply compare the signs of the estimates, we find sign consistency in 20 out of 54 cases, which suggests that most multifactor models do not satisfy the restrictions imposed by the ICAPM. However, if we apply our multivariate inequality test, we find that in 43 out of 54 cases we do not have enough evidence to reject the null hypothesis of sign consistency at the 5% level, which indicates that most models do satisfy the ICAPM restrictions. Another important finding is that accounting for potential model misspecification can make a significant difference in terms of the conclusions reached. When the test statistic is computed using the traditional Fama and MacBeth (1973) standard errors, we obtain sign consistency in 32 out of 54 cases, but when misspecification-robust standard errors are used, the sign restrictions are satisfied in 43 out of 54 cases. Moreover, we find that the use of misspecification-robust standard errors makes a substantial difference when the correlation between the returns on the test assets and the factors is low, as it is the case when using size and momentum sorted portfolios (see Kan et al. (2013) for a discussion of this point). Specifically, when the 25 size and momentum sorted portfolios are used as test assets, the sign consistency hypothesis is rejected in 17 out of 27 cases if the test statistics are computed using Fama and MacBeth (1973) standard errors but only in 8 out of 27 cases if misspecification-robust standard errors are used. On the other hand, when the test assets are the 25 size and book-to-market sorted portfolios, the test statistics based on the Fama and MacBeth (1973) asymptotic variance indicate rejection of the null in only 5 out of 27 cases, whereas the misspecification-robust test statistics indicate rejection in 3 out of 27 cases.

The rest of the chapter is organized as follows. Section 2 presents an asymptotic analysis of the estimates of the prices of covariance risk under potentially misspecified models. In addition, we provide the limiting distribution of the sample cross-sectional R^2. Finally, we develop a multiple sign restriction test and show how this test accounts for estimation and model misspecification uncertainty. Section 3 presents our main empirical findings and Section 4 concludes. The proofs of the propositions are provided in the Appendix.

3.2 Asymptotic analysis under potentially misspecified models

As discussed in the introduction, an asset-pricing model seeks to explain cross-sectional differences in expected asset returns in terms of asset exposures computed relative to the model's systematic economic factors. The two-pass cross-sectional regression (CSR) methodology has become the most popular approach for estimating and testing linear asset-pricing models. Despite the existence of many variations of the CSR methodology, the basic approach always involves two steps or passes. In the first pass, the betas of the test assets are estimated from OLS time-series regressions of returns on some common factors. In the second pass, the returns on the test assets are regressed on the betas estimated from the first pass. The intercept and the slope coeficients from the second-pass CSR are the estimates of the zero-beta rate and factor risk premia.

Let f be a K-vector of factors and R a vector of excess returns (i.e., returns on zero investment portfolios) on N test assets. We define $Y = [f',\ R']'$ and its mean and covariance matrix as

$$\mu = E[Y] \equiv \begin{bmatrix} \mu_f \\ \mu_R \end{bmatrix}, \tag{3.2}$$

$$V = \mathrm{Var}[Y] \equiv \begin{bmatrix} V_f & V_{f,R} \\ V_{R,f} & V_R \end{bmatrix}, \tag{3.3}$$

where V is assumed to be positive definite. The multiple regression betas of the N assets with respect to the K factors are defined as $\beta = V_{R,f}V_f^{-1}$. These are measures of systematic risk or the sensitivity of the asset returns to the factors. In addition, we denote the covariance matrix of the residuals of the N assets by $\Sigma = V_R - V_{R,f}V_f^{-1}V_{f,R}$.

In the following analysis, we focus on an excess returns specification of the CSR methodology. This essentially involves constraining the zero-beta rate to equal the risk-free rate, a practice that is common in other parts of the empirical asset-pricing literature. For example, studies that focus on time-series "alphas" when all factors are traded impose this restriction (see, for example, Gibbons, Ross, and Shanken (1989)). We implement the zero-beta rate restriction in the CSR context by working with test asset returns in excess of the T-bill rate, while excluding the constant from the expected return relations. Thus, the proposed K-factor beta-pricing model specifies that asset expected excess returns are linear in the betas, i.e.,

$$\mu_R = \beta\gamma, \tag{3.4}$$

where β is assumed to be of full column rank and γ is a vector consisting of the risk premia on the K factors. When the model is misspecified, the pricing-error vector, $\mu_R - \beta\gamma$, will be nonzero for all values of γ. In that case, it makes sense to choose γ to minimize some aggregation of pricing errors. Denoting by W an $N \times N$ symmetric positive-definite weighting matrix, we define the (pseudo) risk premia as the choice of γ that minimizes the quadratic form of pricing errors:

$$\gamma_W = \text{argmin}_\gamma (\mu_R - \beta\gamma)' W (\mu_R - \beta\gamma) = (\beta' W \beta)^{-1} \beta' W \mu_R. \tag{3.5}$$

The corresponding pricing errors of the N assets are then given by

$$e_W = \mu_R - \beta\gamma_W = [I_N - \beta(\beta' W \beta)^{-1} \beta' W] \mu_R. \tag{3.6}$$

In addition to the pricing errors, researchers are often interested in a normalized goodness-of-fit measure for a model. A popular measure is the cross-sectional R^2. Following Kandel and Stambaugh (1995), this is defined as

$$\rho_W^2 = 1 - \frac{Q}{Q_0}, \tag{3.7}$$

where

$$\begin{aligned} Q_0 &= \mu_R' W \mu_R, & (3.8) \\ Q &= e_W' W e_W = \mu_R' W \mu_R - \mu_R' W \beta (\beta' W \beta)^{-1} \beta' W \mu_R. & (3.9) \end{aligned}$$

Note that $0 \leq \rho_W^2 \leq 1$ and it is a decreasing function of the aggregate pricing errors $Q = e_W' W e_W$. Thus, ρ_W^2 is a natural measure of goodness of fit.

While the betas are typically used as the regressors in the second-pass CSR, there is a potential issue with the use of multiple regression betas when $K > 1$: in general, the beta of an asset with respect to a particular factor depends on what other factors are included in the first-pass time-series OLS regression. As a consequence, the interpretation of the risk premia γ in the context of model selection becomes problematic. To overcome this problem, in the subsequent analysis we focus on an alternative second-pass CSR that uses the covariances $V_{R,f}$ instead of the betas β as the regressors.[2] Let λ_W be the choice of coefficients that minimizes the quadratic form of pricing

[2] Another solution to this problem is to use simple regression betas as the regressors in the second-pass CSR, as in Chen, Roll, and Ross (1986) and Jagannathan and Wang (1996, 1998). Kan and Robotti (2011) provide asymptotic results for the CSR with simple regression betas under potentially misspecified models.

errors:

$$\lambda_W = \text{argmin}_\lambda (\mu_R - V_{R,f}\lambda)' W (\mu_R - V_{R,f}\lambda) = (V_{f,R} W V_{R,f})^{-1} V_{f,R} W \mu_R. \tag{3.10}$$

Given (3.5) and (3.10), there is a one-to-one correspondence between γ_W and λ_W:

$$\lambda_W = V_f^{-1} \gamma_W. \tag{3.11}$$

It is easy to see that the pricing errors from this alternative second-pass CSR, $e_W = \mu_R - V_{R,f}\lambda_W$, are the same as those in (3.6). It follows that the ρ_W^2 for these two CSRs are also identical. However, it is important to note that unless V_f is a diagonal matrix, $\lambda_{W,i} = 0$ does not imply $\gamma_{W,i} = 0$, and vice versa (see Kan et al. (2013) for a detailed discussion of this point).

It should be emphasized that unless the model is correctly specified, γ_W, λ_W, e_W, and ρ_W^2 depend on the choice of W. Popular choices of W in the literature are $W = I_N$ (OLS CSR), $W = V_R^{-1}$ (GLS CSR), and $W = \Sigma_d^{-1}$ (WLS CSR), where $\Sigma_d = \text{Diag}(\Sigma)$. To simplify the notation, we suppress the subscript W from γ_W, λ_W, e_W, and ρ_W^2 when the choice of W is clear from the context.

We now turn to estimation of the models. Let $Y_t = [f_t', \; R_t']'$, where f_t is the vector of K proposed factors at time t and R_t is the vector of N excess returns on the test assets at time t. We assume the time series Y_t is jointly stationary and ergodic, with finite fourth moment. Suppose we have T observations on Y_t and denote the sample moments of Y_t by

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_f \\ \hat{\mu}_R \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} Y_t, \tag{3.12}$$

$$\hat{V} = \begin{bmatrix} \hat{V}_f & \hat{V}_{f,R} \\ \hat{V}_{R,f} & \hat{V}_R \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} (Y_t - \hat{\mu})(Y_t - \hat{\mu})'. \tag{3.13}$$

When the weighting matrix W is known (say OLS CSR), we can estimate λ_W in (3.10) by

$$\hat{\lambda} = (\hat{V}_{f,R} W \hat{V}_{R,f})^{-1} \hat{V}_{f,R} W \hat{\mu}_R. \tag{3.14}$$

In the GLS and WLS cases, the weighting matrix W involves unknown parameters and, therefore, we need to substitute a consistent estimate of W, say $\hat{W}$, in (3.14). This is typically the corresponding matrix of sample moments, $\hat{W} = \hat{V}_R^{-1}$ for GLS and $\hat{W} = \text{Diag}(\hat{\Sigma})^{-1} = \hat{\Sigma}_d^{-1}$ for WLS, where $\hat{\Sigma} = \hat{V}_R - \hat{V}_{R,f}\hat{V}_f^{-1}\hat{V}_{f,R}$.

The sample measure of ρ^2 is similarly defined as

$$\hat{\rho}^2 = 1 - \frac{\hat{Q}}{\hat{Q}_0}, \tag{3.15}$$

where $\hat{Q}_0$ and $\hat{Q}$ are consistent estimators of Q_0 and Q in (3.8) and (3.9), respectively. When W is known, we estimate Q_0 and Q using

$$\hat{Q}_0 = \hat{\mu}_R' W \hat{\mu}_R, \tag{3.16}$$

$$\hat{Q} = \hat{\mu}_R' W \hat{\mu}_R - \hat{\mu}_R' W \hat{V}_{R,f} (\hat{V}_{f,R} W \hat{V}_{R,f})^{-1} \hat{V}_{f,R} W \hat{\mu}_R. \tag{3.17}$$

When W is not known, we replace W with $\hat{W}$.

3.2.1 Asymptotic distribution of $\hat{\lambda}$ under potentially misspecified models

When computing the standard error of $\hat{\lambda}$, researchers typically rely on the asymptotic distribution of $\hat{\lambda}$ under the assumption that the model is correctly specified. In the following proposition, we relax this assumption and provide general expressions for the asymptotic variance of $\hat{\lambda}$ for the OLS, GLS, and WLS cases under potential model misspecification.

Proposition 1. *Under a potentially misspecified model, the asymptotic distribution of $\hat{\lambda}$ is given by*

$$\sqrt{T}(\hat{\lambda} - \lambda) \overset{A}{\sim} N(0_K, V(\hat{\lambda})), \tag{3.18}$$

where

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}']. \tag{3.19}$$

To simplify the expressions for h_t, we define $G_t = V_{R,f} - (R_t - \mu_R)(f_t - \mu_f)'$, $H = (V_{f,R} W V_{R,f})^{-1}$, $A = HV_{f,R}W$, $\lambda_t = AR_t$, $u_t = e'W(R_t - \mu_R)$, and $\Psi_t = Diag(\epsilon_t \epsilon_t')$, where $\epsilon_t = R_t - \mu_R - \beta(f_t - \mu_f)$.
(a) With a known weighting matrix W, $\hat{\lambda} = (\hat{V}_{f,R} W \hat{V}_{R,f})^{-1} \hat{V}_{f,R} W \hat{\mu}_R$ and

$$h_t = (\lambda_t - \lambda) + AG_t\lambda + H(f_t - \mu_f)u_t. \tag{3.20}$$

(b) For GLS, $\hat{\lambda} = (\hat{V}_{f,R} \hat{V}_R^{-1} \hat{V}_{R,f})^{-1} \hat{V}_{f,R} \hat{V}_R^{-1} \hat{\mu}_R$ and

$$h_t = (\lambda_t - \lambda) + AG_t\lambda + H(f_t - \mu_f)u_t - (\lambda_t - \lambda)u_t. \tag{3.21}$$

(c) For WLS, $\hat{\lambda} = (\hat{V}_{f,R}\hat{\Sigma}_d^{-1}\hat{V}_{R,f})^{-1}\hat{V}_{f,R}\hat{\Sigma}_d^{-1}\hat{\mu}_R$ *and*

$$h_t = (\lambda_t - \lambda) + AG_t\lambda + H(f_t - \mu_f)u_t - A\Psi_t\Sigma_d^{-1}e. \tag{3.22}$$

When the model is correctly specified, we have:

$$h_t = (\lambda_t - \lambda) + AG_t\lambda. \tag{3.23}$$

Proof. See Appendix **A.1**.

To conduct statistical tests, we need a consistent estimator of $V(\hat{\lambda})$. This can be obtained by replacing the h_t's with their sample counterparts $\hat{h}_t$'s. In particular, if h_t is uncorrelated over time, then we have $V(\hat{\lambda}) = E[h_t h_t']$, and its consistent estimator is given by

$$\hat{V}(\hat{\lambda}) = \frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}_t'. \tag{3.24}$$

When h_t is autocorrelated, one can use Newey and West's (1987) method to obtain a consistent estimator of $V(\hat{\lambda})$.

Inspection of (3.20) reveals that there are three sources of asymptotic variance for $\hat{\lambda}$. The first term $\lambda_t - \lambda$ measures the asymptotic variance of $\hat{\lambda}$ when the *true* covariances are used in the CSR. For example, if R_t is i.i.d., then λ_t is also i.i.d. and we can use the time-series variance of λ_t to compute the standard error of $\hat{\lambda}$. This coincides with the popular Fama and MacBeth (1973) method. Since the covariances are estimated with error, an errors-in-variables (EIV) problem is introduced in the second-pass CSR. The second term $AG_t\lambda$ is the EIV adjustment term that accounts for the estimation errors in the estimated covariances. The first two terms together give us the $V(\hat{\lambda})$ under the correctly specified model. When the model is misspecified ($e \neq 0_N$), there is a third term $H(f_t - \mu_f)u_t$, which we call the misspecification adjustment term. Traditionally, this term has been ignored by empirical researchers. Comparing (3.21) and (3.22) with the expression for h_t in (3.20), we see that there is an extra term in h_t associated with the use of $\hat{W}$ instead of W. This fourth term vanishes if the weighting matrix W is known.

3.2.2 Asymptotic distribution of the sample cross-sectional R^2

The sample R^2 ($\hat{\rho}^2$) in the second-pass CSR is a popular measure of goodness of fit for a model. A high $\hat{\rho}^2$ is viewed as evidence that the model under study does a good job of explaining the cross-section of expected returns. Lewellen, Nagel, and Shanken (2010) point out several pitfalls to using this approach and explore simulation techniques to obtain approximate confidence intervals for ρ^2.[3] In this subsection, we provide a formal statistical analysis of $\hat{\rho}^2$.

The asymptotic distribution of $\hat{\rho}^2$ crucially depends on the value of ρ^2. When $\rho^2 = 1$ (that is, a correctly specified model), the asymptotic distribution serves as the basis for a specification test of the asset-pricing model. This is an alternative to the various multivariate asset-pricing tests that have been developed in the literature. Although all of these tests focus on an aggregate pricing-error measure, the R^2-based test examines pricing errors in relation to the cross-sectional variation in expected returns, allowing for a simple and appealing interpretation. At the other extreme, the asymptotic distribution when $\rho^2 = 0$ (a misspecified model that does not explain any of the cross-sectional variation in expected returns) permits a test of whether the model has *any* explanatory power for expected returns.

When $0 < \rho^2 < 1$ (a misspecified model that provides some explanatory power), the case of primary interest, $\hat{\rho}^2$ is asymptotically normally distributed around its true value. It is readily verified that the asymptotic standard error of $\hat{\rho}^2$ approaches zero as $\rho^2 \to 0$ or $\rho^2 \to 1$, and thus it is not monotonic in ρ^2. The asymptotic normal distribution of $\hat{\rho}^2$ breaks down for the two extreme cases ($\rho^2 = 0$ or 1) because, by construction, $\hat{\rho}^2$ will always be above zero (even when $\rho^2 = 0$) and below one (even when $\rho^2 = 1$).

Proposition 2. *In the following, we set W to be V_R^{-1} and Σ_d^{-1} for the GLS and WLS cases, respectively.*
(a) When $\rho^2 = 1$,

$$T(\hat{\rho}^2 - 1) = -\frac{T\hat{Q}}{\hat{Q}_0} \overset{A}{\sim} -\sum_{j=1}^{N-K} \frac{\xi_j}{Q_0} x_j, \tag{3.25}$$

where the x_j's are independent χ_1^2 random variables, and the ξ_j's are the eigenvalues of

$$P' W^{\frac{1}{2}} S W^{\frac{1}{2}} P, \tag{3.26}$$

where P is an $N \times (N - K)$ orthonormal matrix with columns orthogonal to $W^{\frac{1}{2}} V_{R,f}$, S is the asymptotic covariance matrix of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t y_t$, and $y_t = 1 - \lambda'(f_t - \mu_f)$ is the normalized stochastic discount factor (SDF).

[3] Jagannathan, Kubota, and Takehara (1998), Kan and Zhang (1999), and Jagannathan and Wang (2007) use simulations to examine the sampling errors of the cross-sectional R^2 and risk premium estimates under the assumption that one of the factors is "useless," that is, independent of returns.

(b) When $0 < \rho^2 < 1$,

$$\sqrt{T}(\hat{\rho}^2 - \rho^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[n_t n_{t+j}]\right), \tag{3.27}$$

where

$$\begin{aligned} n_t &= 2\left[-u_t y_t + (1-\rho^2)v_t\right]/Q_0 && \text{for known } W, && (3.28)\\ n_t &= \left[u_t^2 - 2u_t y_t + (1-\rho^2)(2v_t - v_t^2)\right]/Q_0 && \text{for } \hat{W} = \hat{V}_R^{-1}, && (3.29)\\ n_t &= \left[-2u_t y_t + e'\Gamma_t e + (1-\rho^2)(2v_t - \mu_R'\Gamma_t\mu_R)\right]/Q_0 && \text{for } \hat{W} = \hat{\Sigma}_d^{-1}, && (3.30) \end{aligned}$$

with $\Gamma_t = \Sigma_d^{-1}\Psi_t\Sigma_d^{-1}$ *and* $v_t = \mu_R' W(R_t - \mu_R)$.
(c) When $\rho^2 = 0$,

$$T\hat{\rho}^2 \overset{A}{\sim} \sum_{j=1}^{K} \frac{\xi_j}{Q_0} x_j, \tag{3.31}$$

where the x_j*'s are independent* χ_1^2 *random variables and the* ξ_j*'s are the eigenvalues of*

$$(V_{f,R} W V_{R,f}) V(\hat{\lambda}), \tag{3.32}$$

where $V(\hat{\lambda})$ *is given in Proposition 1.*

Proof. See Appendix **A.2**.

3.2.3 Multiple sign restriction test

In this section, we develop and implement a formal test of multiple sign restrictions. This is a multivariate inequality test based on results in the statistics literature due to Wolak (1987, 1989).

Suppose that interest lies in testing

$$H_0: \ \mathcal{Q}\lambda \geq 0_p \quad \text{vs.} \quad H_1: \ \lambda \in \Re^K, \tag{3.33}$$

where $\mathcal{Q}$ is a $p \times K$ matrix of linear inequality restrictions with rank p ($p \leq K$) and 0_p is a $(p \times 1)$-vector of zeros. The $\mathcal{Q}$ matrix can be set up to incorporate restrictions that either come from some *a priori* knowledge or from theory.

Given the normality result in Proposition 1, the test statistic is constructed by first solving the quadratic programming problem

$$\min_{\lambda} (\hat{\lambda} - \lambda)'\mathcal{Q}'(\mathcal{Q}\hat{V}(\hat{\lambda})\mathcal{Q}')^{-1}\mathcal{Q}(\hat{\lambda} - \lambda) \qquad \text{s.t.} \quad \mathcal{Q}\lambda \geq 0_p, \tag{3.34}$$

where $\hat{V}(\hat{\lambda})$ is a consistent estimator of $V(\hat{\lambda})$. Let $\tilde{\lambda}$ be the optimal solution of the problem in (3.34). The likelihood ratio test of the null hypothesis is

$$LR = T(\hat{\lambda} - \tilde{\lambda})'\mathcal{Q}'(\mathcal{Q}\hat{V}(\hat{\lambda})\mathcal{Q}')^{-1}\mathcal{Q}(\hat{\lambda} - \tilde{\lambda}). \tag{3.35}$$

For computational purposes, it is more convenient to consider the dual problem

$$\min_{\rho} \rho'\mathcal{Q}\hat{\lambda} + \frac{1}{2}\rho'(\mathcal{Q}\hat{V}(\hat{\lambda})\mathcal{Q}')\rho \qquad \text{s.t. } \rho \geq 0_p. \tag{3.36}$$

Let $\tilde{\rho}$ be the optimal solution of the problem in (3.36). The Kuhn-Tucker test of the null hypothesis is given by

$$KT = T\tilde{\rho}'(\mathcal{Q}\hat{V}(\hat{\lambda})\mathcal{Q}')\tilde{\rho}. \tag{3.37}$$

The objective functions of the primal and dual problems evaluated at the optimum $(\tilde{\lambda}, \tilde{\rho})$ are equal and we have that $LR = KT$.

To conduct statistical inference, we need to derive the asymptotic distribution of LR. Wolak (1989) shows that under $H_0: \ \mathcal{Q}\lambda = 0_p$ (that is, the least favorable value of $\mathcal{Q}\lambda$ under the null hypothesis), LR has a weighted chi-squared distribution

$$LR \overset{A}{\sim} \sum_{i=0}^{p} w_i \left((\mathcal{Q}V(\hat{\lambda})\mathcal{Q}')^{-1}\right) X_i = \sum_{i=0}^{p} w_{p-i} \left(\mathcal{Q}V(\hat{\lambda})\mathcal{Q}'\right) X_i, \tag{3.38}$$

where the X_i's are independent χ^2 random variables with i degrees of freedom, $\chi_0^2 \equiv 0$, and the weights w_i sum up to one. To compute the p-value of LR, we replace $V(\hat{\lambda})$ with $\hat{V}(\hat{\lambda})$ in the weight functions.

3.3 Empirical analysis

In this section we evaluate whether several prominent multifactor models satisfy the sign restrictions imposed by the ICAPM. To obtain an *a priori* knowledge of the expected signs of the λ parameters, we first run multiple predictive time-series regressions of the changes in the investment opportunity set (proxied by the future expected return on the aggregate equity market) on the model-specific state variables. Next, we run cross-sectional regressions of average excess returns on the estimated covariances between the excess returns and the innovations in these state variables (i.e., the factors). A multifactor model is said to satisfy the restrictions imposed by the ICAPM if the signs with which the model's state variables predict changes in the investment opportunity set coincide with the signs of the prices of covariance risk that their innovations earn in the cross-section. In addition, since the covariance price of market risk has a natural interpretation of relative risk aversion coefficient, we incorporate in our set of sign restrictions the constraint that the market premium should be positive.

In addition to the models considered in Maio and Santa-Clara (2012), we analyze the five-factor specification proposed by Fama and French (2015). More specifically, we estimate and test nine multifactor models. Four of these models are theory based and contain innovations in state variables that have often been used in the return predictability literature. The rest are empirically motivated models that have sometimes received an ICAPM interpretation in the asset-pricing literature.

The first of the theory motivated models is the specification of Hahn and Lee (2006), which extends the CAPM by including innovations in a term state variable and a default state variable. The multifactor model proposed by Petkova (2006) contains innovations in the dividend yield and in the risk-free rate in addition to the factors in the Hahn and Lee (2006) model. We also test an unrestricted version of the ICAPM specification of Campbell and Vuolteenaho (2004), which incorporates innovations in a price-to-earnings state variable, a term state variable, and a value spread state variable in addition to the market. The last theory motivated model is the multifactor model proposed by Koijen, Lustig, and Van Nieuwerburgh (2017), which includes, in addition to the market return, innovations in the term state variable and in the return-forecasting factor of Cochrane and Piazzesi (2005).

As for the empirically motivated models, the first model we consider is the Fama and French (1993) three-factor model, which extends the CAPM by including size and value in addition to the market. The Carhart (1997) four-factor model extends the Fama and French (1993) three-factor model by including a momentum factor. The Pástor and Stambaugh (2003) model extends the Fama and French (1993) three-factor model by including a liquidity factor. We also consider the five-factor model used by Fama and French (1993) to explain the expected returns on stocks and bonds. Their augmented model includes a term and a default factor in addition to the market, size, and value factors. Finally, we also estimate and test the five-factor model proposed by Fama and French (2015) which incorporates a profitability and an investment factor in addition to the classical three factors, namely market, size and value.

3.3.1 Predictive regressions for ICAPM state variables

In this section, we examine whether and with what sign the candidates state variables forecast changes in investment opportunities. The proxy for the investment opportunity set is the aggregate equity market and changes in investment opportunities are proxied by the monthly return on the value-weighted stock market index (from Kenneth French's website). The sample period is from July 1963 until December 2018. For each of the previously described models, we assess the joint forecasting power of the state variables by running multiple predictive time-series OLS regressions of the following form:

$$r_{t,t+q} = a_q + b_q z_t + u_{t,t+q}, \tag{3.39}$$

where $r_{t,t+q} = r_{t+1} + .. + r_{t+q}$ is the continuously compounded return over q periods, z_t is the set of candidate state variables corresponding to each model, and $u_{t,t+q}$ is a conditionally zero-mean forecasting error. The forecasting horizons q we consider are one, twelve, and sixty months. Here, our interest lies in the estimates of b_q and their associated t-statistics. This is indicative of whether a state variable forecasts positive or negative changes in future investment opportunities and of whether this effect is statistically significant.

We start by describing the state variables that will be used in the theory-based models. The predictive regressions for the Hahn and Lee (2006) model (HL) are given by

$$r_{t,t+q} = a_q + b_q TERM_t + c_q DEF_t + u_{t,t+q}, \tag{3.40}$$

where $TERM$ is slope of the Treasury yield curve, computed as the difference between the yields on ten-year and one-year Treasury bonds, and DEF is corporate bond default spread, computed as the difference between the yields on BAA- and AAA-rated corporate bonds. The yield data used for computing these factors are from the Federal Reserve Bank of St. Louis database (FRED).

For the Petkova (2006) model (P) we have

$$r_{t,t+q} = a_q + b_q TERM_t + c_q DEF_t + d_q DY_t + e_q RF_t + u_{t,t+q}, \tag{3.41}$$

where DY is the aggregate dividend-to-price ratio of the S&P Composite index, computed as the log ratio of annual dividends to the price level of the index (from Robert Shiller's website), and RF is the one-month Treasury bill rate (from Kenneth French's website).

In case of the Campbell and Vuolteenaho (2004) model (CV) we have

$$r_{t,t+q} = a_q + b_q TERM_t + c_q PE_t + d_q VS_t + u_{t,t+q}, \tag{3.42}$$

where PE is the aggregate price-to-earnings ratio of the S&P Composite index, computed as the log ratio of the price level of the index to a ten-year moving average of earnings (cyclically adjusted price-earnings) using data available on Robert Shiller's website, and VS is the value spread of Campbell and Vuolteenaho (2004), computed as the difference between the monthly log book-to-market ratios of the small high-book-to-market portfolio and the small low-book-to-market portfolio using data on the six portfolios sorted on size and book-to-market from Kenneth French's website.

Finally, for the Koijen et al. (2017) model (KLVN), the predictive regression is formulated as

$$r_{t,t+q} = a_q + b_q TERM_t + c_q CP_t + u_{t,t+q}, \tag{3.43}$$

where CP is the Cochrane and Piazzesi (2005) factor, computed as the fitted value from a regression of the average (across maturities) excess bond return on a linear combination of forward rates using the Fama-Bliss data from CRSP.[4]

For the empirical specifications, the state variables are constructed as in Maio and Santa-Clara (2012). Specifically, in the case of the Fama and French (1993) three-factor model (FF3), the state variables corresponding to the size (SMB) and value (HML) factors are approximated using monthly market-to-book data on the six portfolios sorted on size and book-to-market (BM) from Kenneth French's website:

$$SMB^*_{FF3} = \frac{MB_{SL} + MB_{SM} + MB_{SH}}{3} - \frac{MB_{BL} + MB_{BM} + MB_{BH}}{3}, \tag{3.44}$$

$$HML^*_{FF3} = \frac{MB_{SH} + MB_{BH}}{2} - \frac{MB_{SL} + MB_{BL}}{2}, \tag{3.45}$$

where MB_{SL}, MB_{SM}, MB_{SH}, MB_{BL}, MB_{BM}, and MB_{BH} are the monthly market-to-book ratios of the small-low BM, small-medium BM, small-high BM, big-low BM, big-medium BM, and big-high BM portfolios. This approximation allows us to interpret SMB^*_{FF3} and HML^*_{FF3} as the state variables and the factors themselves as innovations in these state variables, that is, $SMB \simeq \Delta SMB^*_{FF3}$ and $HML \simeq \Delta HML^*_{FF3}$ (see Maio and Santa-Clara (2012)). Hence the

[4]For details on the construction of the CP factor see Cochrane and Piazzesi (2005).

predictive regression for FF3 is

$$r_{t,t+q} = a_q + b_q SMB^*_{FF3,t} + c_q HML^*_{FF3,t} + u_{t,t+q}. \tag{3.46}$$

For the Carhart (1997) model (C), we approximate the state variable associated with the momentum factor using cumulative sums of the factor returns over the previous 60 months:

$$CUMD_t = \sum_{s=t-59}^{t} UMD_s, \tag{3.47}$$

where UMD is the momentum factor (from Kenneth French's website). As pointed out by Maio and Santa-Clara (2012) and Cooper and Maio (2018), we use the 60 months cumulative sum because the total cumulative sum is close to being non-stationary and the momentum factors is approximated by the first difference in this constructed state variable $UMD \simeq \Delta CUMD$. Thus, the predictive regression takes the form

$$r_{t,t+q} = a_q + b_q SMB^*_{FF3,t} + c_q HML^*_{FF3,t} + d_q CUMD_t + u_{t,t+q}. \tag{3.48}$$

We adopt a similar approach for constructing the state variable associated with the liquidity factor in the Pástor and Stambaugh (2003) model (PS):

$$CL_t = \sum_{s=t-59}^{t} L_s, \tag{3.49}$$

where L is the non-traded liquidity factor from Lubos Pastor's website. The first difference in the state variable closely approximates the original factor, that is, $L \simeq \Delta CL$. The predictive regression is formulated as

$$r_{t,t+q} = a_q + b_q SMB^*_{FF3,t} + c_q HML^*_{FF3,t} + d_q CL_t + u_{t,t+q}. \tag{3.50}$$

The predictive regression for the Fama and French (1993) five-factor model (FFTD) that incorporates the bond-market factors $TERM$ and DEF is given by

$$r_{t,t+q} = a_q + b_q SMB^*_{FF3,t} + c_q HML^*_{FF3,t} + d_q TERM_t + e_q DEF_t + u_{t,t+q}, \tag{3.51}$$

where SMB^*_{FF3} and HML^*_{FF3} are the state variables defined in (3.44) and (3.45), respectively.

The state variables corresponding to the Fama and French (2015) five-factor model (FF5) are constructed using a similar approach to the one used for obtaining the FF3 state variables, except that we now use market-to-book data on three sets of portfolios instead of just one, namely the six portfolios sorted on size and book-to-market, the six portfolios sorted on size and operating profitability, as well as the six portfolios sorted on size and investment (from Kenneth French's website). More specifically, the state variables corresponding to the size (SMB), value (HML), profitability (RMW), and investment (CMA) factors are obtained by combining monthly market-to-book ratios across the relevant portfolios as follows:

$$SMB^*_{FF5} = \frac{SMB_{B/M} + SMB_{OP} + SMB_{INV}}{3}, \tag{3.52}$$

$$SMB_{B/M} = \frac{MB_{SL} + MB_{SM} + MB_{SH}}{3} - \frac{MB_{BL} + MB_{BM} + MB_{BH}}{3}, \tag{3.53}$$

$$SMB_{OP} = \frac{MB_{SW} + MB_{SM} + MB_{SR}}{3} - \frac{MB_{BW} + MB_{BM} + MB_{BR}}{3}, \tag{3.54}$$

$$SMB_{INV} = \frac{MB_{SC} + MB_{SM} + MB_{SA}}{3} - \frac{MB_{BC} + MB_{BM} + MB_{BA}}{3}, \tag{3.55}$$

$$HML^*_{FF5} = \frac{MB_{SH} + MB_{BH}}{2} - \frac{MB_{SL} + MB_{BL}}{2}, \tag{3.56}$$

$$RMW^*_{FF5} = \frac{MB_{SR} + MB_{BR}}{2} - \frac{MB_{SW} + MB_{BW}}{2}, \tag{3.57}$$

$$CMA^*_{FF5} = \frac{MB_{SC} + MB_{BC}}{2} - \frac{MB_{SA} + MB_{BA}}{2}, \tag{3.58}$$

where MB_{SW}, MB_{SM}, MB_{SR}, MB_{BW}, MB_{BM}, and MB_{BR} are the monthly market-to-book ratios of the small-weak profitability, small-medium profitability, small-robust profitability, big-weak profitability, big-medium profitability, and big-robust profitability portfolios, and MB_{SC}, MB_{SM}, MB_{SA}, MB_{BC}, MB_{BM}, and MB_{BA} are the monthly market-to-book ratios of the small-conservative investment, small-medium investment, small-aggressive investment, big-conservative investment, big-medium investment, and big-aggressive investment portfolios. As before, this approximation enables us to interpret the original factors as innovations in the state variables, that is, $SMB \simeq \Delta SMB^*_{FF5}$, $HML \simeq \Delta HML^*_{FF5}$, $RMW \simeq \Delta RMW^*_{FF5}$, and $CMA \simeq \Delta CMA^*_{FF5}$. Therefore, the predictive regression for the FF5 model is

$$r_{t,t+q} = a_q + b_q SMB^*_{FF5,t} + c_q HML^*_{FF5,t} + d_q RMW^*_{FF5,t} + e_q CMA^*_{FF5,t} + u_{t,t+q}. \tag{3.59}$$

In Table I we present estimation results for the multiple predictive regressions at horizons q of one, twelve, and sixty months. We report slope parameter estimates and associated t-ratios computed using Newey and West (1987) standard errors with q lags to correct for the serial correlation in the residuals induced by the overlapping cumulative returns.

Table I.2
Multiple Predictive Regressions for Theory Motivated ICAPM State Variables

The table presents the estimation results of the multiple long-horizon predictive regressions corresponding to the models explicitly proposed as ICAPM applications. The forecasted variable is the monthly continuously compounded return on the value-weighted stock market index, at horizons q of 1, 12 and 60 months ahead. The forecasting variables are the current values of the term spread (TERM), default spread (DEF), market dividend yield (DY), one-month Treasury bill rate (RF), market price-earnings ratio (PE), value spread (VS), Cochrane-Piazzesi factor (CP). The original sample is from July 1963 to December 2018 but q observations are lost in each of the q-horizon regressions. We report parameter estimates and corresponding Newey-West t-ratios computed with q lags in parenthesis.

	TERM	DEF	DY	RF	PE	VS	CP
Panel A: $q = 1$							
HL	0.15	0.51					
	(1.01)	(0.97)					
P	0.19	0.15	0.01	-0.64			
	(0.83)	(0.23)	(1.83)	(-0.47)			
CV	0.36				-0.00	-0.03	
	(2.19)				(-0.93)	(-2.40)	
KLVN	0.07						0.19
	(0.42)						(1.77)
Panel B: $q = 12$	TERM	DEF	DY	RF	PE	VS	CP
HL	1.21	6.42					
	(0.85)	(2.11)					
P	1.43	2.39	0.14	-8.60			
	(0.85)	(0.72)	(2.29)	(-0.98)			
CV	2.78				-0.09	-0.14	
	(1.87)				(-1.82)	(-0.89)	
KLVN	1.02						0.95
	(0.61)						(0.94)
Panel C: $q = 60$	TERM	DEF	DY	RF	PE	VS	CP
HL	5.50	26.75					
	(1.14)	(2.71)					
P	12.82	5.36	0.45	16.52			
	(2.68)	(0.55)	(4.96)	(0.69)			
CV	8.85				-0.52	0.22	
	(2.67)				(-7.32)	(1.33)	
KLVN	5.60						2.70
	(1.13)						(1.44)

In Panels A, B, C of Table I.1, we report the estimation results from the multiple predictive regressions corresponding to the theoretical models that have been explicitly proposed as ICAPM applications, at horizons of one, twelve, and sixty months, respectively. Several observations are in order. First, there seems to be stronger evidence of return predictability at longer horizons. For the one-month ahead predictive regressions (Panel A), only two out of eleven estimates are statistically significant at the 5% level, while for the sixty-month ahead predictive regressions five estimates are statistically significant at the 5% level (Panel C). The exceptions are the estimated coefficients on the $TERM$ state variable in the HL and KLVN models, on the DEF and RF state variables in the P model, on the VS state variable in the CV model, and the CP state variable in the KLVN model.

Second, the state variables do not predict future returns with the same signs across the different horizons considered. For instance, the RF state variable in the P model negatively affects future market returns in the one-month and twelve-month ahead predictive regressions (Panels A and B) but positively affects future market returns in the sixty-months ahead predictive regression (Panel C). However, these estimates are not statistically significant. Similarly, the VS state variable in the CV model has a negative and statistically significant estimated slope coefficient at the one-month horizon (Panel A), but this estimate becomes positive and statistically insignificant at the sixty-month horizon (Panel C).

In Panels A, B, and C of Table I.2, we report predictive regressions for the empirical specifications at horizons of one, twelve, and sixty months, respectively. The pattern of stronger return predictability at longer horizons seems to persist. At the one-month horizon (Panel A), none of the estimates is statistically significant at the 5% level. In contrast, at the sixty-month horizon (Panel C), the return predictability hypothesis receives support in seven out of sixteen instances at the 5% level.

Furthermore, we can observe the same issue of changing signs across predictive horizons. For example, the $CUMD$ state variable in the C model has a negative estimate in the one-month and twelve-month ahead predictive regressions (Panels A and B), but the slope estimate associated with it becomes positive in the sixty-month ahead predictive regression (Panel C). The CMA^{*}_{FF5} state variable behaves similarly. These estimates are not statistically significant at the 5% level though.

These results raise a number of questions. On the one hand, it is unclear what is the appropriate horizon over which the ability of the state variables to forecast future investment opportunities should be assessed. The horizon choice is somewhat arbitrary from an economic perspective, but from a statistical perspective the choice will naturally be driven by the availability of evidence in support of the predictability hypothesis. In our analysis of sign restrictions, we will rely on the sixty-month horizon, for which there is the strongest evidence of return predictability.

Table I.2
Multiple Predictive Regressions for Empirically Motivated ICAPM State Variables

The table presents the estimation results of the multiple long-horizon predictive regressions corresponding to the empirical models that have been given a ICAPM interpretation. The forecasted variable is the monthly continuously compounded return on the value-weighted stock market index, at horizons q of 1, 12 and 60 months ahead. The forecasting variables are the current values of the size factor for the FF3 model (SMB^*_{FF3}), value factor for the FF3 model (HML^*_{FF3}), cumulative momentum factor (CUMD), cumulative liquidity factor (CL), term spread (TERM), default spread (DEF), size factor for the FF5 model (SMB^*_{FF5}), value factor for the FF5 model (HML^*_{FF5}), profitability factor for the FF5 model (RMW^*_{FF5}), investment factor for the FF5 model (CMA^*_{FF5}). The original sample is from July 1963 to December 2018 but q observations are lost in each of the q-horizon regressions. We report parameter estimates and corresponding Newey-West t-ratios computed with q lags in parenthesis.

Panel A: $q = 1$

	SMB^*_{FF3}	HML^*_{FF3}	CUMD	CL	TERM	DEF	SMB^*_{FF5}	HML^*_{FF5}	RMW^*_{FF5}	CMA^*_{FF5}
FF3	0.01	0.00								
	(0.99)	(1.57)								
C	0.01	0.00	-0.01							
	(0.53)	(1.79)	(-1.33)							
PS	0.00	0.01		0.01						
	(0.07)	(1.70)		(1.04)						
FFTD	0.01	0.00			0.23	0.22				
	(0.88)	(1.38)			(1.45)	(0.34)				
FF5							0.00	-0.00	-0.00	0.01
							(0.43)	(-0.10)	(-0.45)	(1.67)

Panel B: $q = 12$

	SMB^*_{FF3}	HML^*_{FF3}	CUMD	CL	TERM	DEF	SMB^*_{FF5}	HML^*_{FF5}	RMW^*_{FF5}	CMA^*_{FF5}
FF3	0.17	0.02								
	(1.55)	(1.40)								
C	0.17	0.02	-0.01							
	(1.47)	(1.44)	(-0.15)							
PS	0.13	0.03		0.02						
	(0.97)	(1.24)		(0.48)						
FFTD	0.15	0.02			1.96	3.66				
	(1.44)	(1.25)			(1.44)	(1.10)				
FF5							0.06	-0.02	-0.06	0.07
							(1.04)	(-0.67)	(-0.90)	(0.83)

Panel C: $q = 60$

	SMB^*_{FF3}	HML^*_{FF3}	CUMD	CL	TERM	DEF	SMB^*_{FF5}	HML^*_{FF5}	RMW^*_{FF5}	CMA^*_{FF5}
FF3	0.39	0.14								
	(1.86)	(3.81)								
C	0.47	0.14	0.14							
	(2.41)	(3.48)	(0.77)							
PS	0.36	0.15		0.01						
	(1.07)	(2.31)		(0.09)						
FFTD	0.35	0.14			7.51	11.43				
	(1.73)	(4.76)			(2.01)	(1.00)				
FF5							0.12	-0.04	-0.40	-0.02
							(1.07)	(-0.48)	(-2.82)	(-0.14)

On the other hand, it is not entirely clear how to proceed when an estimate in the predictive regression is statistically insignificant, especially in light of the somewhat limited evidence in support of the predictability hypothesis discussed above. Since a statistically insignificant slope estimate is consistent with the true coefficient being either positive or negative, we believe that we should not impose a sign restriction on the corresponding price of covariance risk in this case. Surprisingly, previous studies have failed to take this aspect into account when evaluating the sign consistency of the considered models. Tests of sign consistency merely relied on an eye-balling exercise whereby the researcher simply compared the signs of the estimates in the time-series regressions with the signs of the estimates in the cross-sectional regressions, regardless of precision. As we will see later on, making inferences in the absence of statistical significance can strongly affect one's conclusions on the consistency of a multifactor model with the restrictions imposed by the ICAPM.

3.3.2 Multifactor models

In this section we examine the performance of several multifactor models in cross-sectional tests of asset-pricing models. Our main interest is in assessing whether and with what signs the innovations in the state variables are priced in the cross-section of equity returns. For each of the multifactor models considered, we estimate the prices of covariance risk by running two-pass cross-sectional regressions of average excess returns on the estimated factor covariances. The cross-sectional specification for a generic multifactor model is

$$\mu_R = V_{R,f}\lambda_f, \tag{3.60}$$

where μ_R are the expected excess returns on the test assets, $V_{R,f}$ are the covariances between the excess returns on the test assets and the innovations in the state variables, and λ_f are the prices of covariance risk.

The test assets returns used in the analysis are the monthly value-weighted returns on the 25 Fama-French size and book-to-market ranked portfolios, as well as the 25 Fama-French size and momentum ranked portfolios (from Kenneth French's website). The sample period runs from July 1963 until December 2018 (666 monthly observations). Following Maio and Santa-Clara (2012), we use first differences as proxies for the innovations in the state variables and use the notation Δ to indicate these first differences or changes.

In each of the nine models considered, the first factor is the excess market return (rm), which is proxied by the monthly return on the value-weighted stock market index in excess of the one-month Treasury bill rate (from Kenneth French's website). Hence, in the case of the Hahn and

Lee (2006) model (HL), the cross-sectional specification is

$$\mu_R = V_{R,rm}\lambda_{rm} + V_{R,\Delta term}\lambda_{\Delta term} + V_{R,\Delta def}\lambda_{\Delta def}, \tag{3.61}$$

where $\Delta term$ denotes the change in the slope of the Treasury yield curve and Δdef denotes the change in the corporate bond default spread.

For the ICAPM proposed by Petkova (2006) (P) we have

$$\mu_R = V_{R,rm}\lambda_{rm} + V_{R,\Delta term}\lambda_{\Delta term} + V_{R,\Delta def}\lambda_{\Delta def} + +V_{R,\Delta dy}\lambda_{\Delta dy} + V_{R,\Delta rf}\lambda_{\Delta rf}, \tag{3.62}$$

where Δdy denotes changes in the aggregate dividend-to-price ratio and Δrf denotes changes in the one-month Treasury bill rate.

The Campbell and Vuolteenaho (2004) model (CV) takes the form

$$\mu_R = V_{R,rm}\lambda_{rm} + V_{R,\Delta term}\lambda_{\Delta term} + V_{R,\Delta pe}\lambda_{\Delta pe} + V_{R,\Delta vs}\lambda_{\Delta vs}, \tag{3.63}$$

where Δpe denotes changes in the aggregate price-to-earnings ratio and Δvs denotes changes in the value spread of Campbell and Vuolteenaho (2004).

The Koijen et al. (2017) model (KLVN) is

$$\mu_R = V_{R,rm}\lambda_{rm} + V_{R,\Delta term}\lambda_{\Delta term} + V_{R,\Delta cp}\lambda_{\Delta cp}, \tag{3.64}$$

where Δcp denotes changes in the return-forecasting factor of Cochrane and Piazzesi (2005).

In the case of the Fama and French (1993) three-factor model (FF3) we have

$$\mu_R = V_{R,rm}\lambda_{rm} + V_{R,smb}\lambda_{smb} + V_{R,hml}\lambda_{hml}, \tag{3.65}$$

where smb is the return difference between portfolios of stocks with small and big market capitalizations, and hml is the return difference between portfolios of stocks with high and low book-to-market ratios (from Kenneth French's website).

The cross-sectional specification for the Carhart (1997) four-factor model (C) is

$$\mu_R = V_{R,rm}\lambda_{rm} + V_{R,smb}\lambda_{smb} + V_{R,hml}\lambda_{hml} + V_{R,umd}\lambda_{umd}, \tag{3.66}$$

where umd is return difference between portfolios of stocks with high and low prior returns (from

Kenneth French's website).

For the Pástor and Stambaugh (2003) model (PS), we have

$$\mu_R = V_{R,rm}\lambda_{rm} + V_{R,smb}\lambda_{smb} + V_{R,hml}\lambda_{hml} + V_{R,l}\lambda_l, \tag{3.67}$$

where l is the non-traded liquidity factor (from Lubos Pastor's website).

The Fama and French (1993) three-factor model augmented with the bond-market factors, the term spread and corporate default spread, (FFTD) is

$$\mu_R = V_{R,rm}\lambda_{rm} + V_{R,smb}\lambda_{smb} + V_{R,hml}\lambda_{hml} + V_{R,\Delta term}\lambda_{\Delta term} + V_{R,\Delta def}\lambda_{\Delta def}. \tag{3.68}$$

Finally, the Fama and French (2015) five-factor model (FF5) is

$$\mu_R = V_{R,rm}\lambda_{rm} + V_{R,smb}\lambda_{smb} + V_{R,hml}\lambda_{hml} + V_{R,rmw}\lambda_{rmw} + V_{R,cma}\lambda_{cma}, \tag{3.69}$$

where rmw is the return difference between portfolios of stocks with robust and weak operating profitability, and cma is the return difference between portfolios of stocks with conservative and aggressive investment (from Kenneth French's website).

3.3.3 Sample cross-sectional R^2s of the models

In Table II, we report the sample cross-sectional R^2 ($\hat{\rho}^2$) for each model and investigate whether the model does a good job of explaining the cross-section of expected returns. We denote the p-value of a specification test of $H_0 : \rho^2 = 1$ by $p(\rho^2 = 1)$, and the p-value of a test of $H_0 : \rho^2 = 0$ by $p(\rho^2 = 0)$. Both tests are based on the asymptotic results in Section 2 for the sample cross-sectional R^2 statistic. We also provide an approximate F-test of model specification for comparison, denoted $\hat{Q}_c$. Next, we report the asymptotic standard error of the sample R^2, se($\hat{\rho}^2$), computed under the assumption of a misspecified model that provides some explanatory power i.e. $0 < \rho^2 < 1$. Finally, No. of para. is the number of parameters in each asset-pricing model.

The F-test is a generalized version of the cross-sectional regression test (CSRT) of Shanken (1985). It is based on a quadratic form in the model's deviations, $\hat{Q}_c = \hat{e}'\hat{V}(\hat{e})^+\hat{e}$, where $\hat{V}(\hat{e})$ is a consistent estimator of the asymptotic variance of the sample pricing errors and $\hat{V}(\hat{e})^+$ its pseudo-inverse. When the model is correctly specified (that is, $e = 0_N$ or $\rho^2 = 1$), we have $T\hat{Q}_c \overset{A}{\sim} \chi^2_{N-K-1}$. Following Shanken (1985), the reported p-value, $p(Q_c = 0)$, is for a transformation of $\hat{Q}_c$ that has an approximate F distribution: $\hat{Q}_c \overset{\text{app.}}{\sim} \left(\frac{N-K-1}{T-N+1}\right) F_{N-K-1,T-N+1}$.

Table II.1
Sample Cross-Sectional R^2s and Specification Tests of the Models Using the 25 Size and Book-to-Market Portfolios as Test Assets

The table presents the sample cross-sectional R^2 ($\hat{\rho}^2$) and the generalized CSRT ($\hat{Q}_c$) of nine asset-pricing models. The models include the ICAPM specifications proposed by Hahn and Lee (2006) (HL), Petkova (2006) (P), Campbell and Vuolteenaho (2004) (CV), Koijen et al. (2017) (KLVN), the Fama and French (1993) three-factor model (FF3), the Carhart (1997) model (C), the Pástor and Stambaugh (2003) model (PS), the Fama and French (1993) three-factor model augmented by TERM and DEF (FFTD), and the Fama and French (2015) five-factor model (FF5). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios. The data are from July 1963 to December 2018 (666 observations). $p(\rho^2 = 1)$ is the p-value for the test of $H_0 : \rho^2 = 1$. $p(\rho^2 = 0)$ is the p-value for the test of $H_0 : \rho^2 = 0$. se($\hat{\rho}^2$) is the standard error of $\hat{\rho}^2$ under the assumption that $0 < \rho^2 < 1$. $p(Q_c = 0)$ is the p-value for the approximate F-test of $H_0 : Q_c = 0$. No. of para. is the number of parameters in the model.

	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
Panel A: OLS									
$\hat{\rho}^2$	0.970	0.979	0.968	0.969	0.966	0.982	0.971	0.973	0.979
$p(\rho^2 = 1)$	0.231	0.678	0.069	0.199	0.000	0.214	0.000	0.007	0.000
$p(\rho^2 = 0)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
se($\hat{\rho}^2$)	0.020	0.018	0.021	0.023	0.020	0.015	0.017	0.017	0.013
$\hat{Q}_c$	0.049	0.025	0.052	0.054	0.144	0.052	0.099	0.061	0.092
$p(Q_c = 0)$	0.095	0.696	0.048	0.046	0.000	0.045	0.000	0.008	0.000
No. of para.	3	5	4	3	3	4	4	5	5
Panel B: GLS	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
$\hat{\rho}^2$	0.288	0.433	0.310	0.355	0.212	0.491	0.246	0.313	0.336
$p(\rho^2 = 1)$	0.002	0.252	0.001	0.016	0.000	0.040	0.000	0.000	0.000
$p(\rho^2 = 0)$	0.005	0.014	0.019	0.002	0.000	0.000	0.003	0.008	0.000
se($\hat{\rho}^2$)	0.150	0.209	0.146	0.155	0.074	0.152	0.092	0.129	0.095
$\hat{Q}_c$	0.081	0.037	0.080	0.061	0.145	0.062	0.115	0.088	0.096
$p(Q_c = 0)$	0.001	0.259	0.000	0.015	0.000	0.010	0.000	0.000	0.000
No. of para.	3	5	4	3	3	4	4	5	5
Panel C: WLS	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
$\hat{\rho}^2$	0.979	0.983	0.979	0.980	0.974	0.986	0.976	0.979	0.982
$p(\rho^2 = 1)$	0.320	0.582	0.124	0.295	0.000	0.154	0.000	0.004	0.000
$p(\rho^2 = 0)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
se($\hat{\rho}^2$)	0.014	0.015	0.014	0.015	0.015	0.010	0.014	0.013	0.011
$\hat{Q}_c$	0.066	0.038	0.072	0.070	0.145	0.058	0.110	0.071	0.095
$p(Q_c = 0)$	0.007	0.239	0.002	0.004	0.000	0.018	0.000	0.001	0.000
No. of para.	3	5	4	3	3	4	4	5	5

In Panels A, B, and C of Table II.1, we provide results for the OLS, GLS, and WLS CSRs, respectively, for the case when the 25 size and book-to-market sorted portfolios are used as test assets. When estimation is done using OLS (Panel A), four models out of nine, namely FF3, PS, FFTD, and FF5, are rejected by the R^2 test, while the F-test indicates rejection of all but the HL and P models at the 5% level. The same four models are also rejected by the R^2 test when using WLS (Panel C) and the same significance level, while in this case the F-test rejects all but the

P model. However, when estimation is made using GLS (Panel B), all the models except the P model are rejected at the 5% significance level by both the R^2 and the F-test. The null hypothesis that the model does not explain any of the variation in expected returns, ($H_0 : \rho^2 = 0$), is rejected at the 5% level for all the models and under all estimation methods.

Table II.2 is for the 25 size and momentum sorted portfolios. Based on the OLS R^2 (Panel A), four out of nine models, namely HL, CV, FF3 and C, are rejected at the 5% level, while the F-test indicates rejection of the same four models except the CV model. Using WLS (Panel C), all but the P, KLVN and FF5 models are rejected by the R^2 test, while the WLS F-test rejects four out of the nine models at the 5% level, namely the HL, CV, FF3 and C models. As for the test of $H_0 : \rho^2 = 0$, the null of no explanatory power is strongly rejected at the 1% level for all the models both in the OLS case (Panel A) and the WLS case (Panel C). When estimation is done using GLS (Panel B), the null hypothesis that the model is correctly specified ($H_0 : \rho^2 = 1$) is rejected at the 5% level for all the models by both the R^2 test and the F-test. Additionally, the hypothesis of no explanatory power $H_0 : \rho^2 = 0$ cannot be rejected at the 5% level in three instances, and indicates that this choice of test assets is particularly challenging for the HL, CV and KLVN models.

Although the results are sensitive to the criterion minimized in estimation as well as to the set of test assets used, there is widespread evidence of model misspecification. These are situations in which the use of misspecification-robust standard errors is likely to affect the outcomes of the parameter and multivariate inequality tests.

3.3.4 Properties of the λ estimates under correctly specified and potentially misspecified models

In this section, we follow what has been done in the literature and compare the signs of the time-series estimates with the signs of the cross-sectional estimates. In addition, we require the market price of covariance risk to be positive. We draw conclusions on sign consistency regardless of the statistical significance of the estimates.

Table II.2
Sample Cross-Sectional R^2s and Specification Tests of the Models Using the 25 Size and Momentum Portfolios as Test Assets

The table presents the sample cross-sectional R^2 ($\hat{\rho}^2$) and the generalized CSRT ($\hat{Q}_c$) of nine asset-pricing models. The models include the ICAPM specifications proposed by Hahn and Lee (2006) (HL), Petkova (2006) (P), Campbell and Vuolteenaho (2004) (CV), Koijen et al. (2017) (KLVN), the Fama and French (1993) three-factor model (FF3), the Carhart (1997) model (C), the Pástor and Stambaugh (2003) model (PS), the Fama and French (1993) three-factor model augmented by TERM and DEF (FFTD), and the Fama and French (2015) five-factor model (FF5). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios. The data are from July 1963 to December 2018 (666 observations). $p(\rho^2 = 1)$ is the p-value for the test of $H_0 : \rho^2 = 1$. $p(\rho^2 = 0)$ is the p-value for the test of $H_0 : \rho^2 = 0$. se($\hat{\rho}^2$) is the standard error of $\hat{\rho}^2$ under the assumption that $0 < \rho^2 < 1$. $p(Q_c = 0)$ is the p-value for the approximate F-test of $H_0 : Q_c = 0$. No. of para. is the number of parameters in the model.

	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
Panel A: OLS									
$\hat{\rho}^2$	0.856	0.928	0.880	0.910	0.787	0.971	0.851	0.925	0.979
$p(\rho^2 = 1)$	0.002	0.065	0.039	0.442	0.000	0.000	0.207	0.122	0.115
$p(\rho^2 = 0)$	0.001	0.001	0.001	0.001	0.001	0.000	0.003	0.001	0.000
se($\hat{\rho}^2$)	0.095	0.056	0.094	0.089	0.116	0.017	0.169	0.064	0.015
$\hat{Q}_c$	0.063	0.032	0.037	0.021	0.164	0.112	0.019	0.032	0.043
$p(Q_c = 0)$	0.011	0.420	0.304	0.917	0.000	0.000	0.928	0.437	0.117
No. of para.	3	5	4	3	3	4	4	5	5
Panel B: GLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
$\hat{\rho}^2$	0.116	0.347	0.087	0.145	0.127	0.369	0.164	0.202	0.476
$p(\rho^2 = 1)$	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.008
$p(\rho^2 = 0)$	0.089	0.009	0.314	0.173	0.003	0.000	0.022	0.028	0.000
se($\hat{\rho}^2$)	0.076	0.181	0.053	0.131	0.059	0.095	0.088	0.096	0.133
$\hat{Q}_c$	0.170	0.066	0.200	0.141	0.229	0.115	0.162	0.154	0.066
$p(Q_c = 0)$	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.003
No. of para.	3	5	4	3	3	4	4	5	5
Panel C: WLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
$\hat{\rho}^2$	0.907	0.942	0.912	0.933	0.917	0.978	0.933	0.952	0.989
$p(\rho^2 = 1)$	0.000	0.071	0.012	0.388	0.000	0.000	0.032	0.036	0.160
$p(\rho^2 = 0)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
se($\hat{\rho}^2$)	0.056	0.048	0.065	0.068	0.046	0.013	0.056	0.038	0.007
$\hat{Q}_c$	0.104	0.041	0.062	0.028	0.186	0.113	0.045	0.048	0.049
$p(Q_c = 0)$	0.000	0.157	0.010	0.690	0.000	0.000	0.117	0.062	0.057
No. of para.	3	5	4	3	3	4	4	5	5

In Table III, we report estimates of the price of covariance risk $\hat{\lambda}$ and associated t-ratios under correctly specified and potentially misspecified models. For correctly specified models, we give the t-ratio of Fama and MacBeth (1973), followed by that of Shanken (1992) and Jagannathan and Wang (1998), which account for estimation error in the covariances. Last, we report the t-ratio under potentially misspecified models, based on the results presented in Section 2. The various t-ratios are identified by subscripts fm, s, jw, and pm, respectively. Additionally, we also report the signs with which the underlying state variables predict future returns in the multiple time-series regressions (as a superscript in $\hat{\lambda}$, that is, $\hat{\lambda}^{(\pm)}$).

In Panels A, B, and C of Table III.1, we provide results for the OLS, GLS, and WLS CSRs, respectively, for the case when the 25 size and book-to-market sorted portfolios are used as test assets. When estimation is done using OLS (Table III.1, Panel A), only three models, namely FF3, C and FFTD, appear to be consistent with an ICAPM interpretation. However, none of the estimates with an inconsistent sign is statistically significant at the 5% level, except for the profitability (rmw) factor in the FF5 model. The estimate on the risk-free (rf) factor in the P model is also sign-inconsistent and statistically significant when Fama and MacBeth (1973) standard errors are used in the estimation, but becomes insignificant when using EIV-corrected and misspecification-robust standard errors.

For GLS (Table III.1, Panel B), six out of nine models appear to be consistent with an ICAPM interpretation. Only three models, namely P, CV and FF5, have estimated prices of covariance risk whose signs are inconsistent with the signs of the estimated slopes from the predictive regressions. As for the statistical significance of the estimates with an inconsistent sign, most of them are not statistically significant after accounting for the estimation error in the covariances and for potential model misspecification. For instance, in case of the estimates on the market (rm) and dividend-yield (dy) factors in the P model, the Fama and MacBeth (1973) t-ratios indicate statistical significance at the 5% level but this is no longer the case if one considers the Shanken (1992), Jagannathan and Wang (1998) and misspecification-robust t-ratios. The only factor whose price of covariance risk is significant, as indicated by the set of all t-ratios, is the profitability (rmw) factor in the FF5 model.

Table III.1
Estimates and t-ratios of Prices of Covariance Risk Using the 25 Size and Book-to-Market Portfolios as Test Assets

The table presents the estimation results of nine asset-pricing models. The models include the ICAPM specifications proposed by Hahn and Lee (2006) (HL), Petkova (2006) (P), Campbell and Vuolteenaho (2004) (CV), Koijen et al. (2017) (KLVN), the Fama and French (1993) three-factor model (FF3), the Carhart (1997) model (C), the Pástor and Stambaugh (2003) model (PS), the Fama and French (1993) three-factor model augmented by TERM and DEF (FFTD), and the Fama and French (2015) five-factor model (FF5). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios. The data are from July 1963 to December 2018 (666 observations). We report parameter estimates $\hat{\lambda}$, the Fama and MacBeth (1973) t-ratio under correctly specified models (t-ratio$_{fm}$), the Shanken (1992) and the Jagannathan and Wang (1998) t-ratios under correctly specified models that account for the EIV problem (t-ratio$_s$ and t-ratio$_{jw}$, respectively), and our model misspecification-robust t-ratios (t-ratio$_{pm}$).

Panel A:OLS

	HL			P				
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{dy}^{(+)}$	$\hat{\lambda}_{rf}^{(+)}$
Estimate	-0.63	495.37	182.31	-7.25	339.23	725.36	-11.25	-1363.82
t-ratio$_{fm}$	(-0.55)	(4.66)	(0.65)	(-1.42)	(2.67)	(3.31)	(-1.13)	(-2.38)
t-ratio$_s$	(-0.32)	(2.70)	(0.38)	(-0.70)	(1.32)	(1.63)	(-0.55)	(-1.17)
t-ratio$_{jw}$	(-0.28)	(2.46)	(0.37)	(-0.71)	(1.17)	(1.71)	(-0.57)	(-1.12)
t-ratio$_{pm}$	(-0.29)	(2.53)	(0.35)	(-0.55)	(0.97)	(1.53)	(-0.40)	(-0.58)

	CV				KLVN			FF3		
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{pe}^{(-)}$	$\hat{\lambda}_{vs}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{cp}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$
Estimate	-0.65	485.08	-0.50	0.57	-0.29	430.61	29.08	3.26	1.74	6.56
t-ratio$_{fm}$	(-0.17)	(4.27)	(-0.07)	(0.10)	(-0.17)	(3.20)	(0.66)	(3.36)	(1.26)	(4.40)
t-ratio$_s$	(-0.10)	(2.56)	(-0.04)	(0.06)	(-0.11)	(1.78)	(0.73)	(3.27)	(1.23)	(4.25)
t-ratio$_{jw}$	(-0.10)	(2.27)	(-0.04)	(0.07)	(-0.10)	(1.88)	(0.41)	(3.01)	(1.25)	(4.25)
t-ratio$_{pm}$	(-0.10)	(2.39)	(-0.04)	(0.04)	(-0.11)	(2.03)	(0.39)	(3.01)	(1.25)	(4.23)

	C				PS			
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{umd}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{l}^{(+)}$
Estimate	7.70	1.14	14.19	21.07	-0.86	0.74	6.08	10.14
t-ratio$_{fm}$	(6.49)	(0.82)	(7.18)	(6.21)	(-0.48)	(0.50)	(4.12)	(2.55)
t-ratio$_s$	(4.80)	(0.62)	(5.28)	(4.60)	(-0.42)	(0.44)	(3.55)	(2.21)
t-ratio$_{jw}$	(3.75)	(0.48)	(4.09)	(3.53)	(-0.42)	(0.36)	(3.02)	(1.97)
t-ratio$_{pm}$	(3.55)	(0.49)	(4.18)	(2.91)	(-0.36)	(0.35)	(2.99)	(1.80)

	FFTD					FF5				
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(-)}$	$\hat{\lambda}_{rmw}^{(-)}$	$\hat{\lambda}_{cma}^{(-)}$
Estimate	0.99	2.15	3.09	301.70	430.77	3.96	5.88	3.65	13.98	3.23
t-ratio$_{fm}$	(0.84)	(1.50)	(1.72)	(3.73)	(2.34)	(3.03)	(3.57)	(0.86)	(3.68)	(0.38)
t-ratio$_s$	(0.58)	(1.04)	(1.20)	(2.59)	(1.63)	(2.86)	(3.35)	(0.82)	(3.45)	(0.36)
t-ratio$_{jw}$	(0.48)	(1.07)	(1.12)	(2.38)	(1.77)	(2.72)	(3.33)	(0.81)	(3.04)	(0.37)
t-ratio$_{pm}$	(0.40)	(1.04)	(0.81)	(1.35)	(1.47)	(2.58)	(3.33)	(0.58)	(2.41)	(0.26)

Table III.1 (Continued)
Estimates and t-ratios of Prices of Covariance Risk Using the 25 Size and Book-to-Market Portfolios as Test Assets

Panel B:GLS

	HL			P				
	$\hat{\lambda}^{(+)}_{rm}$	$\hat{\lambda}^{(+)}_{term}$	$\hat{\lambda}^{(+)}_{def}$	$\hat{\lambda}^{(+)}_{rm}$	$\hat{\lambda}^{(+)}_{term}$	$\hat{\lambda}^{(+)}_{def}$	$\hat{\lambda}^{(+)}_{dy}$	$\hat{\lambda}^{(+)}_{rf}$
Estimate	0.73	309.76	84.34	-10.97	279.96	443.78	-21.49	-593.66
t-ratio$_{fm}$	(0.72)	(5.09)	(0.70)	(-3.19)	(3.79)	(3.02)	(-3.23)	(-1.25)
t-ratio$_{s}$	(0.55)	(3.80)	(0.53)	(-1.92)	(2.28)	(1.82)	(-1.95)	(-0.75)
t-ratio$_{jw}$	(0.50)	(3.65)	(0.57)	(-1.93)	(2.21)	(1.88)	(-1.89)	(-0.79)
t-ratio$_{pm}$	(0.50)	(2.69)	(0.37)	(-1.29)	(1.52)	(1.28)	(-1.23)	(-0.46)

	CV				KLVN			FF3		
	$\hat{\lambda}^{(+)}_{rm}$	$\hat{\lambda}^{(+)}_{term}$	$\hat{\lambda}^{(-)}_{pe}$	$\hat{\lambda}^{(+)}_{vs}$	$\hat{\lambda}^{(+)}_{rm}$	$\hat{\lambda}^{(+)}_{term}$	$\hat{\lambda}^{(+)}_{cp}$	$\hat{\lambda}^{(+)}_{rm}$	$\hat{\lambda}^{(+)}_{smb}$	$\hat{\lambda}^{(+)}_{hml}$
Estimate	-1.99	282.35	4.51	6.68	1.15	314.46	55.50	3.38	1.85	5.90
t-ratio$_{fm}$	(-0.83)	(4.49)	(1.12)	(1.58)	(1.14)	(5.17)	(3.00)	(3.55)	(1.37)	(4.03)
t-ratio$_{s}$	(-0.64)	(3.43)	(0.86)	(1.21)	(0.81)	(3.62)	(2.12)	(3.45)	(1.34)	(3.90)
t-ratio$_{jw}$	(-0.71)	(3.31)	(0.97)	(1.14)	(0.75)	(3.41)	(2.05)	(3.15)	(1.37)	(3.87)
t-ratio$_{pm}$	(-0.55)	(2.26)	(0.69)	(0.63)	(0.73)	(2.64)	(1.38)	(3.14)	(1.36)	(3.85)

	C				PS			
	$\hat{\lambda}^{(+)}_{rm}$	$\hat{\lambda}^{(+)}_{smb}$	$\hat{\lambda}^{(+)}_{hml}$	$\hat{\lambda}^{(+)}_{umd}$	$\hat{\lambda}^{(+)}_{rm}$	$\hat{\lambda}^{(+)}_{smb}$	$\hat{\lambda}^{(+)}_{hml}$	$\hat{\lambda}^{(+)}_{l}$
Estimate	6.67	1.59	12.06	17.13	0.92	1.06	5.87	6.34
t-ratio$_{fm}$	(6.06)	(1.18)	(6.72)	(5.95)	(0.61)	(0.76)	(4.01)	(2.06)
t-ratio$_{s}$	(4.83)	(0.96)	(5.34)	(4.74)	(0.57)	(0.71)	(3.71)	(1.92)
t-ratio$_{jw}$	(3.85)	(0.82)	(4.26)	(4.05)	(0.54)	(0.66)	(3.30)	(1.70)
t-ratio$_{pm}$	(3.58)	(0.82)	(4.07)	(3.21)	(0.37)	(0.60)	(3.29)	(1.11)

	FFTD					FF5				
	$\hat{\lambda}^{(+)}_{rm}$	$\hat{\lambda}^{(+)}_{smb}$	$\hat{\lambda}^{(+)}_{hml}$	$\hat{\lambda}^{(+)}_{term}$	$\hat{\lambda}^{(+)}_{def}$	$\hat{\lambda}^{(+)}_{rm}$	$\hat{\lambda}^{(+)}_{smb}$	$\hat{\lambda}^{(-)}_{hml}$	$\hat{\lambda}^{(-)}_{rmw}$	$\hat{\lambda}^{(-)}_{cma}$
Estimate	1.54	1.23	3.00	238.54	128.14	4.97	4.74	1.31	12.38	9.13
t-ratio$_{fm}$	(1.34)	(0.87)	(1.71)	(3.27)	(1.03)	(4.01)	(2.96)	(0.34)	(3.76)	(1.19)
t-ratio$_{s}$	(1.09)	(0.71)	(1.40)	(2.66)	(0.84)	(3.77)	(2.80)	(0.33)	(3.55)	(1.14)
t-ratio$_{jw}$	(1.00)	(0.69)	(1.37)	(2.52)	(0.91)	(3.60)	(2.80)	(0.31)	(2.91)	(1.17)
t-ratio$_{pm}$	(0.92)	(0.69)	(1.11)	(1.61)	(0.59)	(3.17)	(2.64)	(0.21)	(2.29)	(0.75)

The results for the WLS case (Table III.1, Panel C) indicate that the same six models, namely HL, KLVN, FF3, C, PS, and FFTD are consistent with the sign restrictions imposed by the ICAPM. Out of the models with price of covariance risk estimates whose signs do not coincide with the signs of the time-series estimates, only the FF5 model contains a coefficient that remains statistically significant at the 5% level using all sets of standard errors.

Table III.2 presents results for the 25 size and momentum sorted portfolios. In the OLS case (Table III.2, Panel A), only the C model appears to be consistent with an ICAPM interpretation as the signs of its price of covariance risk estimates coincide with the signs of the time-series estimates and it also satisfies the requirement that the market price of covariance risk is positive. The P, FF3 and PS models contain estimates with inconsistent signs that become insignificant when using EIV-

Table III.1 (Continued)
Estimates and t-ratios of Prices of Covariance Risk Using the 25 Size and Book-to-Market Portfolios as Test Assets

Panel C:WLS

	HL			P				
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{dy}^{(+)}$	$\hat{\lambda}_{rf}^{(+)}$
Estimate	0.13	392.99	83.67	-2.43	257.46	425.95	-3.17	-1288.26
t-ratio$_{fm}$	(0.11)	(3.79)	(0.30)	(-0.43)	(2.02)	(2.09)	(-0.29)	(-2.18)
t-ratio$_{s}$	(0.08)	(2.55)	(0.20)	(-0.25)	(1.19)	(1.23)	(-0.17)	(-1.28)
t-ratio$_{jw}$	(0.07)	(2.43)	(0.19)	(-0.25)	(1.09)	(1.15)	(-0.16)	(-1.25)
t-ratio$_{pm}$	(0.07)	(2.44)	(0.18)	(-0.22)	(0.99)	(0.96)	(-0.14)	(-0.68)

	CV				KLVN			FF3		
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{pe}^{(-)}$	$\hat{\lambda}_{vs}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{cp}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$
Estimate	-0.41	348.48	1.05	3.81	0.72	329.01	26.14	3.39	1.67	6.08
t-ratio$_{fm}$	(-0.10)	(3.13)	(0.13)	(0.67)	(0.50)	(2.92)	(0.69)	(3.51)	(1.22)	(4.07)
t-ratio$_{s}$	(-0.07)	(2.24)	(0.09)	(0.48)	(0.37)	(2.14)	(0.51)	(3.41)	(1.19)	(3.94)
t-ratio$_{jw}$	(-0.07)	(2.12)	(0.09)	(0.51)	(0.35)	(2.03)	(0.48)	(3.13)	(1.21)	(3.93)
t-ratio$_{pm}$	(-0.07)	(2.06)	(0.09)	(0.33)	(0.36)	(2.00)	(0.47)	(3.14)	(1.21)	(3.92)

	C				PS			
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{umd}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{l}^{(+)}$
Estimate	7.16	1.26	12.69	18.58	0.22	0.92	5.62	7.73
t-ratio$_{fm}$	(6.31)	(0.92)	(6.67)	(5.90)	(0.13)	(0.63)	(3.81)	(2.15)
t-ratio$_{s}$	(4.89)	(0.73)	(5.16)	(4.58)	(0.12)	(0.58)	(3.45)	(1.95)
t-ratio$_{jw}$	(3.90)	(0.58)	(4.18)	(3.66)	(0.12)	(0.51)	(3.01)	(1.70)
t-ratio$_{pm}$	(3.61)	(0.59)	(4.01)	(2.82)	(0.08)	(0.46)	(2.97)	(1.14)

	FFTD					FF5				
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(-)}$	$\hat{\lambda}_{rmw}^{(-)}$	$\hat{\lambda}_{cma}^{(-)}$
Estimate	0.55	0.57	1.82	343.94	69.40	4.14	5.50	2.56	13.05	4.76
t-ratio$_{fm}$	(0.46)	(0.40)	(1.02)	(4.29)	(0.41)	(3.21)	(3.37)	(0.62)	(3.63)	(0.57)
t-ratio$_{s}$	(0.33)	(0.29)	(0.73)	(3.07)	(0.30)	(3.04)	(3.19)	(0.59)	(3.43)	(0.54)
t-ratio$_{jw}$	(0.30)	(0.27)	(0.74)	(2.87)	(0.32)	(2.87)	(3.19)	(0.58)	(2.94)	(0.56)
t-ratio$_{pm}$	(0.26)	(0.28)	(0.54)	(1.65)	(0.23)	(2.66)	(3.14)	(0.41)	(2.35)	(0.40)

corrected and misspecification-robust standard errors. On the other hand, the HL, CV, KLVN, FFTD and FF5 models contain estimates with an inconsistent sign that are statistically significant even after accounting for potential model misspecification.

Table III.2
Estimates and t-ratios of Prices of Covariance Risk Using the 25 Size and Momentum Portfolios as Test Assets

The table presents the estimation results of nine asset-pricing models. The models include the ICAPM specifications proposed by Hahn and Lee (2006) (HL), Petkova (2006) (P), Campbell and Vuolteenaho (2004) (CV), Koijen et al. (2017) (KLVN), the Fama and French (1993) three-factor model (FF3), the Carhart (1997) model (C), the Pástor and Stambaugh (2003) model (PS), the Fama and French (1993) three-factor model augmented by TERM and DEF (FFTD), and the Fama and French (2015) five-factor model (FF5). The models are estimated using monthly excess returns on the 25 Fama-French size and momentum ranked portfolios. The data are from July 1963 to December 2018 (666 observations). We report parameter estimates $\hat{\lambda}$, the Fama and MacBeth (1973) t-ratio under correctly specified models (t-ratio$_{fm}$), the Shanken (1992) and the Jagannathan and Wang (1998) t-ratios under correctly specified models that account for the EIV problem (t-ratio$_s$ and t-ratio$_{jw}$, respectively), and our model misspecification-robust t-ratios (t-ratio$_{pm}$).

Panel A:OLS

	HL			P				
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{dy}^{(+)}$	$\hat{\lambda}_{rf}^{(+)}$
Estimate	6.88	-500.30	-341.54	-0.22	-607.32	51.95	-4.92	-3014.52
t-ratio$_{fm}$	(5.63)	(-4.36)	(-1.58)	(-0.07)	(-5.05)	(0.28)	(-0.80)	(-5.94)
t-ratio$_s$	(3.15)	(-2.44)	(-0.89)	(-0.03)	(-2.28)	(0.13)	(-0.36)	(-2.68)
t-ratio$_{jw}$	(2.27)	(-1.93)	(-0.82)	(-0.03)	(-1.67)	(0.12)	(-0.35)	(-2.09)
t-ratio$_{pm}$	(2.49)	(-2.26)	(-0.71)	(-0.02)	(-1.55)	(0.10)	(-0.19)	(-1.68)

	CV				KLVN			FF3		
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{pe}^{(-)}$	$\hat{\lambda}_{vs}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{cp}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$
Estimate	5.05	-719.79	6.01	21.23	10.94	-745.59	182.96	1.00	2.69	-8.33
t-ratio$_{fm}$	(1.37)	(-4.55)	(0.77)	(2.87)	(6.64)	(-5.48)	(4.62)	(0.91)	(1.84)	(-2.55)
t-ratio$_s$	(0.57)	(-1.89)	(0.32)	(1.20)	(2.20)	(-1.82)	(1.54)	(0.88)	(1.77)	(-2.45)
t-ratio$_{jw}$	(0.49)	(-1.59)	(0.30)	(1.00)	(1.68)	(-1.55)	(1.43)	(0.83)	(1.71)	(-2.43)
t-ratio$_{pm}$	(0.42)	(-2.21)	(0.26)	(0.71)	(2.09)	(-2.02)	(1.66)	(0.84)	(1.66)	(-1.86)

	C				PS			
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{umd}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{l}^{(+)}$
Estimate	5.59	1.17	12.44	6.67	-17.92	-1.83	-17.30	44.39
t-ratio$_{fm}$	(5.21)	(0.81)	(4.51)	(6.34)	(-5.56)	(-1.19)	(-4.15)	(7.03)
t-ratio$_s$	(4.75)	(0.75)	(4.13)	(5.73)	(-2.16)	(-0.46)	(-1.61)	(2.72)
t-ratio$_{jw}$	(3.81)	(0.70)	(3.80)	(3.92)	(-1.41)	(-0.37)	(-1.03)	(1.64)
t-ratio$_{pm}$	(3.82)	(0.69)	(3.68)	(3.97)	(-1.47)	(-0.29)	(-1.39)	(1.55)

	FFTD					FF5				
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(-)}$	$\hat{\lambda}_{rmw}^{(-)}$	$\hat{\lambda}_{cma}^{(-)}$
Estimate	10.20	7.88	14.02	-870.28	398.16	10.46	8.18	-43.38	19.48	79.66
t-ratio$_{fm}$	(6.61)	(4.72)	(4.39)	(-5.97)	(2.30)	(6.58)	(3.61)	(-5.93)	(3.43)	(6.14)
t-ratio$_s$	(2.59)	(1.86)	(1.73)	(-2.34)	(0.91)	(4.33)	(2.40)	(-3.91)	(2.28)	(4.05)
t-ratio$_{jw}$	(1.73)	(1.68)	(1.47)	(-1.59)	(0.71)	(3.93)	(1.83)	(-3.24)	(1.57)	(3.70)
t-ratio$_{pm}$	(2.07)	(1.74)	(1.57)	(-2.01)	(0.50)	(3.24)	(1.67)	(-3.14)	(1.36)	(2.95)

For GLS (Table III.2, Panel B), six of the nine models have estimates that are not consistent with an ICAPM interpretation. The models that satisfy the ICAPM sign requirements are FF3, C, and PS. However, most of the coefficient estimates with an inconsistent sign are not statistically significant at the 5% level. Only the P and FF5 models contain a statistically significant estimate

Table III.2 (Continued)
Estimates and t-ratios of Prices of Covariance Risk Using the 25 Size and Momentum Portfolios as Test Assets

Panel B:GLS

	HL			P				
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{dy}^{(+)}$	$\hat{\lambda}_{rf}^{(+)}$
Estimate	3.76	-45.13	277.89	-2.33	-170.88	350.89	-6.76	-1875.44
t-ratio$_{fm}$	(3.76)	(-0.80)	(2.20)	(-1.01)	(-2.84)	(2.66)	(-1.57)	(-5.23)
t-ratio$_{s}$	(3.51)	(-0.76)	(2.06)	(-0.66)	(-1.86)	(1.74)	(-1.03)	(-3.41)
t-ratio$_{jw}$	(3.31)	(-0.72)	(2.04)	(-0.62)	(-1.56)	(1.66)	(-1.06)	(-3.23)
t-ratio$_{pm}$	(2.86)	(-0.47)	(1.14)	(-0.50)	(-1.23)	(1.19)	(-0.77)	(-2.03)

	CV				KLVN			FF3		
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{pe}^{(-)}$	$\hat{\lambda}_{vs}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{cp}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$
Estimate	1.65	-50.98	3.21	-1.57	4.19	-70.59	55.85	3.20	2.48	4.83
t-ratio$_{fm}$	(0.69)	(-0.89)	(0.77)	(-0.37)	(4.11)	(-1.24)	(3.02)	(3.20)	(1.76)	(2.08)
t-ratio$_{s}$	(0.68)	(-0.87)	(0.75)	(-0.36)	(3.45)	(-1.05)	(2.55)	(3.12)	(1.72)	(2.04)
t-ratio$_{jw}$	(0.66)	(-0.85)	(0.74)	(-0.34)	(3.07)	(-1.02)	(2.45)	(2.88)	(1.79)	(2.08)
t-ratio$_{pm}$	(0.38)	(-0.51)	(0.40)	(-0.18)	(2.44)	(-0.62)	(1.00)	(2.86)	(1.74)	(1.65)

	C				PS			
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{umd}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{l}^{(+)}$
Estimate	4.89	2.23	10.54	6.07	0.53	1.90	3.28	6.17
t-ratio$_{fm}$	(4.69)	(1.58)	(4.19)	(5.92)	(0.34)	(1.33)	(1.36)	(2.32)
t-ratio$_{s}$	(4.36)	(1.49)	(3.90)	(5.45)	(0.32)	(1.25)	(1.28)	(2.17)
t-ratio$_{jw}$	(3.58)	(1.50)	(3.69)	(3.90)	(0.30)	(1.22)	(1.08)	(1.85)
t-ratio$_{pm}$	(3.59)	(1.47)	(3.29)	(3.84)	(0.19)	(1.17)	(0.87)	(1.09)

	FFTD					FF5				
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(-)}$	$\hat{\lambda}_{rmw}^{(-)}$	$\hat{\lambda}_{cma}^{(-)}$
Estimate	4.37	4.28	5.84	-119.76	366.15	8.05	8.84	-28.89	20.97	53.16
t-ratio$_{fm}$	(3.86)	(2.83)	(2.33)	(-1.97)	(2.71)	(5.57)	(4.27)	(-5.21)	(4.16)	(5.16)
t-ratio$_{s}$	(3.39)	(2.49)	(2.05)	(-1.74)	(2.39)	(4.30)	(3.31)	(-4.02)	(3.23)	(3.99)
t-ratio$_{jw}$	(2.98)	(2.64)	(2.11)	(-1.52)	(2.20)	(4.08)	(2.92)	(-3.94)	(2.61)	(3.93)
t-ratio$_{pm}$	(2.65)	(2.39)	(1.58)	(-1.06)	(1.32)	(3.34)	(2.60)	(-2.92)	(2.05)	(2.76)

with an inconsistent sign as indicated by the set of all four t-ratios.

When WLS is used (Table III.2, Panel C), only the C model appears to satisfy the sign restrictions of the ICAPM. The HL, P, and PS models contain estimates for which the sign consistency is violated and which are statistically significant when using standard errors under correctly specified models but insignificant when using misspecification-robust standard errors. On the other hand, the CV, KLVN, FFTD, and FF5 models contain sign inconsistent estimates that are statistically significant even after controlling for model misspecification.

To summarize, if we simply compare the signs of the cross-sectional estimates with the signs of the time-series estimates and we require the market price of covariance risk to be positive, then the sign restrictions imposed by the ICAPM are satisfied in only 20 out of 54 cases. However, it

Table III.2 (Continued)
Estimates and t-ratios of Prices of Covariance Risk Using the 25 Size and Momentum Portfolios as Test Assets

Panel C:WLS

	HL			P				
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{dy}^{(+)}$	$\hat{\lambda}_{rf}^{(+)}$
Estimate	5.02	-277.13	-426.03	-1.14	-430.90	-228.62	-4.03	-2694.15
t-ratio$_{fm}$	(4.10)	(-2.61)	(-1.84)	(-0.37)	(-4.10)	(-1.12)	(-0.66)	(-5.58)
t-ratio$_{s}$	(2.93)	(-1.87)	(-1.32)	(-0.19)	(-2.09)	(-0.57)	(-0.34)	(-2.84)
t-ratio$_{jw}$	(2.21)	(-1.52)	(-1.27)	(-0.17)	(-1.56)	(-0.59)	(-0.32)	(-2.30)
t-ratio$_{pm}$	(2.12)	(-1.43)	(-1.05)	(-0.09)	(-0.98)	(-0.34)	(-0.12)	(-1.30)

	CV				KLVN			FF3		
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{pe}^{(-)}$	$\hat{\lambda}_{vs}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{cp}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$
Estimate	1.51	-521.52	9.83	12.59	9.49	-577.07	165.58	1.66	3.61	-5.50
t-ratio$_{fm}$	(0.40)	(-3.81)	(1.27)	(1.74)	(6.21)	(-4.71)	(4.39)	(1.59)	(2.49)	(-1.84)
t-ratio$_{s}$	(0.22)	(-2.05)	(0.69)	(0.94)	(2.40)	(-1.83)	(1.70)	(1.54)	(2.41)	(-1.79)
t-ratio$_{jw}$	(0.19)	(-1.74)	(0.64)	(0.85)	(1.90)	(-1.63)	(1.68)	(1.45)	(2.44)	(-1.84)
t-ratio$_{pm}$	(0.16)	(-2.29)	(0.57)	(0.50)	(2.77)	(-2.38)	(1.95)	(1.48)	(2.42)	(-1.65)

	C				PS			
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{umd}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{l}^{(+)}$
Estimate	5.22	1.84	11.97	6.40	-9.00	1.13	-10.84	24.63
t-ratio$_{fm}$	(4.90)	(1.28)	(4.41)	(6.14)	(-4.07)	(0.77)	(-3.12)	(6.01)
t-ratio$_{s}$	(4.50)	(1.20)	(4.06)	(5.59)	(-2.43)	(0.46)	(-1.87)	(3.57)
t-ratio$_{jw}$	(3.64)	(1.15)	(3.84)	(3.93)	(-1.69)	(0.40)	(-1.21)	(2.17)
t-ratio$_{pm}$	(3.64)	(1.14)	(3.74)	(3.90)	(-0.95)	(0.29)	(-1.35)	(1.17)

	FFTD					FF5				
	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(+)}$	$\hat{\lambda}_{term}^{(+)}$	$\hat{\lambda}_{def}^{(+)}$	$\hat{\lambda}_{rm}^{(+)}$	$\hat{\lambda}_{smb}^{(+)}$	$\hat{\lambda}_{hml}^{(-)}$	$\hat{\lambda}_{rmw}^{(-)}$	$\hat{\lambda}_{cma}^{(-)}$
Estimate	8.01	7.53	9.14	-656.02	334.33	9.09	9.77	-39.90	19.93	71.97
t-ratio$_{fm}$	(5.67)	(4.45)	(3.24)	(-5.35)	(1.79)	(5.38)	(4.14)	(-5.29)	(3.23)	(5.05)
t-ratio$_{s}$	(2.78)	(2.18)	(1.59)	(-2.63)	(0.88)	(3.72)	(2.87)	(-3.66)	(2.25)	(3.50)
t-ratio$_{jw}$	(1.88)	(2.03)	(1.45)	(-1.79)	(0.75)	(3.51)	(2.20)	(-3.29)	(1.57)	(3.47)
t-ratio$_{pm}$	(2.44)	(2.11)	(1.35)	(-2.72)	(0.49)	(3.03)	(2.07)	(-3.21)	(1.44)	(3.00)

should be noted that most cross-sectional estimates with an inconsistent sign are not statistically significant. In addition, accounting for model misspecification often makes a qualitative difference in terms of the conclusions reached.

3.3.5 Test of multiple sign restrictions

In this section, we employ the test of multiple sign restrictions discussed in Section 2.3 to assess whether the various models satisfy the time-series and cross-sectional restrictions imposed by the ICAPM. The use of this test will lead us to conclusions that are substantially different from the ones based on the comparative analysis of the previous section. The reason, in a nutshell, is that the test of multiple sign restrictions accounts for the estimation error in the parameters, for the joint significance of the estimates, and for potential model misspecification. Specifically, a coefficient estimate that is not statistically significant is consistent with the true coefficient being either positive or negative. This has two implications in the current setting. On the one hand, it is not clear whether a statistically insignificant cross-sectional estimate is consistent with the sign restriction obtained from the time series. On the other hand, and perhaps more importantly, it is not clear what sign restriction should be tested, if any, when the time-series estimate is statistically insignificant. In the following analysis, we will explore two cases: (i) when our *a priori* knowledge is simply based on the signs of the time-series estimates, regardless of their statistical significance, and (ii) when our *a priori* knowledge is only based on the signs of the time-series estimates that are statistically significant.

Imposing sign restrictions on all λ's

For a K-factor model, the test of multiple sign restrictions is a test of the null hypothesis $H_0 : \mathcal{Q}\lambda \geq 0_K$ versus the alternative $H_1 : \lambda \in \Re^K$, where $\mathcal{Q}$ is a $K \times K$ matrix of constraints with rank K. Specifically, when testing whether the coefficient associated with the k^{th} factor is positive (negative), we set the (k, k)-element of the $\mathcal{Q}$ matrix equal to one (minus one), while the other elements in the k^{th} row of $\mathcal{Q}$ are set equal to zero.

In Table IV, we report the values of the test statistic, LR, and associated p-values under correctly specified and potentially misspecified models. The specific form of $V(\hat{\lambda})$ in LR depends on whether the Fama and MacBeth (1973), Shanken (1992), Jagannathan and Wang (1998), or misspecification-robust asymptotic variances of the $\hat{\lambda}$'s are used (see Section 2). The corresponding likelihood ratio tests and their p-values are identified by the subscripts fm, s, jw, and pm, respectively.

In Panels A, B, and C of Table IV.1, we present results for the OLS, GLS, and WLS tests of multiple sign restrictions, respectively, for the case when the 25 size and book-to-market sorted portfolios are used as test assets. Under all estimation methods, FF5 is the only model for which the set of sign restrictions is systematically rejected at the 5% level by all four test statistics. For the other eight models, we are unable to reject the null hypothesis that the sign restrictions imposed by the ICAPM hold. These results are consistent with the analysis of the estimates of the prices of covariance risk in Table III.1. Out of all the estimates with an inconsistent sign, only the estimate associated with the profitability (rmw) factor in FF5 was statistically significant. Statistical precision of the λ estimates is clearly the key driver of the power of the test of sign restrictions.

It is worth noting that conducting inference under potential model misspecification often leads

Table IV.1
Multiple Sign Restriction Tests of the Models Using the 25 Size and Book-to-Market Portfolios as Test Assets

The table presents the results of the multiple sign restriction test of nine asset-pricing models for the case where the restrictions imposed are based on the signs of the respective coefficients from the long-horizon predictive regressions, regardless of their statistical significance. This is a likelihood ratio test of the null hypothesis that the models satisfy the sign restrictions placed by the ICAPM $H_0 : \mathcal{Q}\lambda \geq 0_K$. The models include the ICAPM specifications proposed by Hahn and Lee (2006) (HL), Petkova (2006) (P), Campbell and Vuolteenaho (2004) (CV), Koijen et al. (2017) (KLVN), the Fama and French (1993) three-factor model (FF3), the Carhart (1997) model (C), the Pástor and Stambaugh (2003) model (PS), the Fama and French (1993) three-factor model augmented by TERM and DEF (FFTD), and the Fama and French (2015) five-factor model (FF5). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios. The data are from July 1963 to December 2018 (666 observations). We report the values of the likelihood ratio statistics and corresponding p-values obtained using the Fama and MacBeth (1973) variances under correctly specified models (LR_{fm} and p-value$_{fm}$), the Shanken (1992) and the Jagannathan and Wang (1998) variances under correctly specified models that account for the EIV problem (LR_s and p-value$_s$, and LR_{jw} and p-value$_{jw}$, respectively), and our model misspecification-robust variances (LR_{pm} and p-value$_{jw}$).

Panel A: OLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.30	8.12	0.03	0.03	0.00	0.00	0.23	0.00	27.89
p-value$_{fm}$	(0.67)	(0.02)	(0.77)	(0.89)	(0.87)	(0.88)	(0.87)	(0.97)	(0.00)
LR_s	0.10	1.96	0.01	0.01	0.00	0.00	0.18	0.00	24.17
p-value$_s$	(0.78)	(0.36)	(0.79)	(0.91)	(0.87)	(0.88)	(0.89)	(0.97)	(0.00)
LR_{jw}	0.08	1.99	0.01	0.01	0.00	0.00	0.17	0.00	17.53
p-value$_{jw}$	(0.79)	(0.37)	(0.79)	(0.91)	(0.86)	(0.89)	(0.88)	(0.97)	(0.00)
LR_{pm}	0.08	1.25	0.01	0.01	0.00	0.00	0.13	0.00	16.82
p-value$_{pm}$	(0.78)	(0.56)	(0.81)	(0.91)	(0.86)	(0.85)	(0.90)	(0.98)	(0.00)
Panel B: GLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.00	18.32	1.26	0.00	0.00	0.00	0.00	0.00	31.23
p-value$_{fm}$	(0.90)	(0.00)	(0.44)	(0.89)	(0.87)	(0.88)	(0.98)	(0.98)	(0.00)
LR_s	0.00	6.64	0.75	0.00	0.00	0.00	0.00	0.00	27.13
p-value$_{fm}$	(0.90)	(0.07)	(0.57)	(0.89)	(0.87)	(0.88)	(0.98)	(0.98)	(0.00)
LR_{jw}	0.00	7.32	0.93	0.00	0.00	0.00	0.00	0.00	22.03
p-value$_{jw}$	(0.89)	(0.05)	(0.51)	(0.90)	(0.86)	(0.89)	(0.97)	(0.98)	(0.00)
LR_{pm}	0.00	3.82	0.48	0.00	0.00	0.00	0.00	0.00	20.67
p-value$_{pm}$	(0.88)	(0.23)	(0.66)	(0.90)	(0.86)	(0.87)	(0.98)	(0.98)	(0.00)
Panel C: WLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.00	5.42	0.02	0.00	0.00	0.00	0.00	0.00	25.18
p-value$_{fm}$	(0.89)	(0.08)	(0.78)	(0.93)	(0.87)	(0.88)	(0.98)	(0.98)	(0.00)
LR_s	0.00	1.87	0.01	0.00	0.00	0.00	0.00	0.00	22.13
p-value$_{fm}$	(0.89)	(0.37)	(0.80)	(0.93)	(0.87)	(0.88)	(0.98)	(0.98)	(0.00)
LR_{jw}	0.00	1.87	0.01	0.00	0.00	0.00	0.00	0.00	16.34
p-value$_{jw}$	(0.88)	(0.38)	(0.79)	(0.92)	(0.86)	(0.89)	(0.97)	(0.97)	(0.00)
LR_{pm}	0.00	0.76	0.01	0.00	0.00	0.00	0.00	0.00	14.94
p-value$_{pm}$	(0.87)	(0.68)	(0.82)	(0.92)	(0.86)	(0.85)	(0.98)	(0.98)	(0.00)

to qualitatively different conclusions. In general, the amount of evidence against the null hypothesis decreases when using misspecification-robust standard errors and we observe an increase in p-values. This closely matches the pattern of statistical significance of the λ estimates in the cross-sectional analysis of Table III.1. For instance, OLS and GLS test results (Table IV.1, Panels A and B, respectively) for the P model indicate that the sets of sign restrictions is rejected when the test statistics are based on the Fama and MacBeth (1973) standard errors, but not when inference is robustified against potential model misspecification. This is consistent with the pattern observed in Panels A and B of Table III.1, showing that the P model contains estimates with inconsistent signs that are statistically significant when using Fama and MacBeth (1973) standard errors but insignificant when using misspecification robust standard errors.

In Panels A, B, and C of Table IV.2, we employ the 25 size and momentum sorted portfolios as test assets. Only the FF5 model is systematically rejected under all estimation methods and by all test statistics at the 5% level. Additionally, in the OLS case (Table IV.2, Panel A) the consistency of the HL and CV models with the ICAPM implications is also rejected at the 5% level when misspecification-robust standard errors are used in the estimation. Interestingly, the misspecification-robust test statistic cannot reject the hypothesis of sign consistency in case of the KLVN and FFTD models despite the fact that these models contained a statistically significant estimate with an inconsistent sign as shown in Panel A of Table III.2. This is due to the fact that we are employing a test of joint restrictions across multiple factors; given that the other coefficient estimates in these models have the predicted sign and are statistically significant the evidence against the null is weakened. Similarly, when using GLS (Table IV.2, Panel B) and misspecification-robust standard errors the P model barely misses rejection with a p-value$_{pm}$ of 7%, which reflects the feature of imposing joint sign restrictions across multiple factors.

For WLS (Table IV.2, Panel C), the set of all four test statistics indicates that only the FF3 and C models satisfy the ICAPM restrictions, while the CV, KLVN, FFTD, and FF5 models do not satisfy the restrictions when considering a 5% confidence level. In line with the patterns of diminishing statistical significance shown in Panel A of Table III.3 the test indicates rejection of the HL, P and PS models when Fama and MacBeth (1973) standard errors are used in the estimation but this is no longer the case once misspecification-robust errors are employed.

Several observations emerge from the analysis. First, in 43 out of 54 cases, there is not enough evidence against the null of consistency with the ICAPM when using a 5% significance level and

Table IV.2
Multiple Sign Restriction Tests of the Models Using the 25 Size and Momentum Portfolios as Test Assets

The table presents the results of the multiple sign restriction test of nine asset-pricing models for the case where the restrictions imposed are based on the signs of the respective coefficients from the long-horizon predictive regressions, regardless of their statistical significance. This is a likelihood ratio test of the null hypothesis that the models satisfy the sign restrictions placed by the ICAPM $H_0 : \mathcal{Q}\lambda \geq 0_K$. The models include the ICAPM specifications proposed by Hahn and Lee (2006) (HL), Petkova (2006) (P), Campbell and Vuolteenaho (2004) (CV), Koijen et al. (2017) (KLVN), the Fama and French (1993) three-factor model (FF3), the Carhart (1997) model (C), the Pástor and Stambaugh (2003) model (PS), the Fama and French (1993) three-factor model augmented by TERM and DEF (FFTD), and the Fama and French (2015) five-factor model (FF5). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios. The data are from July 1963 to December 2018 (666 observations). We report the values of the likelihood ratio statistics and corresponding p-values obtained using the Fama and MacBeth (1973) variances under correctly specified models (LR_{fm} and p-value$_{fm}$), the Shanken (1992) and the Jagannathan and Wang (1998) variances under correctly specified models that account for the EIV problem (LR_s and p-value$_s$, and LR_{jw} and p-value$_{jw}$, respectively), and our model misspecification-robust variances (LR_{pm} and p-value$_{jw}$).

	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
Panel A: OLS									
LR_{fm}	19.38	43.69	20.69	30.08	6.50	0.00	30.94	35.65	58.24
p-value$_{fm}$	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.91)	(0.00)	(0.00)	(0.00)
LR_s	6.08	8.88	3.59	3.32	5.98	0.00	4.65	5.49	25.02
p-value$_s$	(0.04)	(0.02)	(0.09)	(0.16)	(0.04)	(0.90)	(0.15)	(0.13)	(0.00)
LR_{jw}	4.15	4.84	2.52	2.40	5.91	0.00	1.99	2.53	15.98
p-value$_{jw}$	(0.10)	(0.10)	(0.18)	(0.25)	(0.04)	(0.89)	(0.41)	(0.41)	(0.00)
LR_{pm}	5.92	6.13	4.88	4.07	3.48	0.00	3.31	4.05	13.71
p-value$_{pm}$	(0.04)	(0.07)	(0.05)	(0.11)	(0.12)	(0.89)	(0.24)	(0.22)	(0.01)
Panel B: GLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.65	34.16	1.15	1.54	0.00	0.00	0.00	3.87	54.50
p-value$_{fm}$	(0.56)	(0.00)	(0.39)	(0.36)	(0.87)	(0.91)	(0.97)	(0.24)	(0.00)
LR_s	0.58	14.44	1.10	1.11	0.00	0.00	0.00	3.02	31.80
p-value$_{fm}$	(0.58)	(0.00)	(0.40)	(0.45)	(0.86)	(0.90)	(0.97)	(0.33)	(0.00)
LR_{jw}	0.52	13.27	1.05	1.04	0.00	0.00	0.00	2.30	21.13
p-value$_{jw}$	(0.62)	(0.00)	(0.42)	(0.46)	(0.86)	(0.88)	(0.96)	(0.42)	(0.00)
LR_{pm}	0.22	6.23	0.34	0.39	0.00	0.00	0.00	1.12	15.04
p-value$_{pm}$	(0.73)	(0.07)	(0.58)	(0.67)	(0.85)	(0.89)	(0.98)	(0.65)	(0.00)
Panel C: WLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	10.12	36.17	14.52	22.15	3.40	0.00	16.71	28.67	47.89
p-value$_{fm}$	(0.01)	(0.00)	(0.00)	(0.00)	(0.13)	(0.91)	(0.00)	(0.00)	(0.00)
LR_s	5.21	9.35	4.22	3.33	3.20	0.00	5.96	6.90	22.57
p-value$_{fm}$	(0.06)	(0.01)	(0.06)	(0.16)	(0.14)	(0.90)	(0.08)	(0.07)	(0.00)
LR_{jw}	4.57	6.12	3.02	2.64	3.40	0.00	2.85	3.19	15.60
p-value$_{jw}$	(0.08)	(0.06)	(0.14)	(0.22)	(0.13)	(0.89)	(0.27)	(0.32)	(0.00)
LR_{pm}	4.68	6.79	5.22	5.66	2.74	0.00	1.98	7.40	13.66
p-value$_{pm}$	(0.08)	(0.06)	(0.04)	(0.05)	(0.18)	(0.89)	(0.42)	(0.05)	(0.01)

misspecification-robust standard errors.[5] Second, accounting for model misspecification can make a significant difference in terms of conclusions: when the test is implemented using the Fama and MacBeth (1973) standard errors then we observe 32 out of 54 instances in which there is not enough evidence to reject the null of sign consistency. Third, the amount of evidence against the null is driven by the statistical significance of the cross-sectional estimates. Finally, the statistical significance of the individual cross-sectional estimates is only indicative of the test results since the null hypothesis being tested is composite.

Imposing sign restrictions conditional on the state variables being robust predictors

As previously mentioned, it is not clear whether and what sign restrictions should be imposed when the time-series estimates are not statistically significant. The results presented in Table IV relate to the case in which the restrictions imposed are purely based on the signs of the time-series estimates, regardless of their statistical significance. We now explore the case in which sign restrictions are imposed conditional on the state variables being robust predictors. Specifically, if the state variable corresponding to the k^{th} factor in a K-factor model is not a robust predictor of future aggregate returns, we eliminate the corresponding row from the matrix of constraints $\mathcal{Q}$. Thus, the test of multiple sign restrictions is a test of the null hypothesis $H_0: \ \mathcal{Q}\lambda \geq 0_p$ versus the alternative $H_1: \ \lambda \in \Re^K$, where $\mathcal{Q}$ is the $p \times K$ matrix of constraints and $p \leq K$ is the number of restrictions being imposed.

In Table V, we report results by imposing sign restrictions only on the factors whose associated state variables have estimated time-series coefficients that are statistically significant at the 5% level, as shown in Panel C of Tables I.1 and I.2. More specifically, we maintain the sign restriction associated with the *term* factor in the P, CV and FFTD models, the *def* factor in the HL model, the *dy* factor in the P model, the *pe* factor in the CV model, the *smb* factor in the C model, the *hml* factor in the FF3, C, PS and FFTD models, and the *rmw* factor in the FF5 model. Additionally, we maintain the restriction that the market price of covariance risk is positive across all the models.

In Panels A, B, and C of Table V.1, we provide results for the OLS, GLS, and WLS tests of multiple sign restrictions, respectively, for the case when the 25 size and book-to-market sorted portfolios are used as test assets. The results are qualitatively similar to the baseline case presented

[5] We also explore the impact of autocorrelation on our results by using the automatic lag length selection procedure of Newey and West (1994) and reach similar conclusions. Specifically, failure to reject the null is now observed in 47 out of 54 cases, and the models that exhibit inconsistencies with the ICAPM at the 5% level using misspecification-robust estimation are the FF5 model for OLS and GLS using the 25 size and book-to-market sorted portfolios, the P model for OLS and GLS, the KLVN model for OLS and WLS, and the FFTD model for OLS using the 25 size and momentum sorted portfolios.

Table V.1
Robust Multiple Sign Restriction Tests of the Models Using the 25 Size and Book-to-Market Portfolios as Test Assets
No restrictions are imposed if the state variables are not robust predictors

The table presents the results of the multiple sign restriction test of nine asset-pricing models for the case where we impose sign restrictions only if the respective coefficient estimates from the long-horizon predictive regressions are statistically significant; otherwise no restriction is imposed. This is a likelihood ratio test of the null hypothesis that the models satisfy the sign restrictions placed by the ICAPM $H_0 : \mathcal{Q}\lambda \geq 0_K$. The models include the ICAPM specifications proposed by Hahn and Lee (2006) (HL), Petkova (2006) (P), Campbell and Vuolteenaho (2004) (CV), Koijen et al. (2017) (KLVN), the Fama and French (1993) three-factor model (FF3), the Carhart (1997) model (C), the Pástor and Stambaugh (2003) model (PS), the Fama and French (1993) three-factor model augmented by TERM and DEF (FFTD), and the Fama and French (2015) five-factor model (FF5). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios. The data are from July 1963 to December 2018 (666 observations). We report the values of the likelihood ratio statistics and corresponding p-values obtained using the Fama and MacBeth (1973) variances under correctly specified models (LR_{fm} and p-value$_{fm}$), the Shanken (1992) and the Jagannathan and Wang (1998) variances under correctly specified models that account for the EIV problem (LR_s and p-value$_s$, and LR_{jw} and p-value$_{jw}$, respectively), and our model misspecification-robust variances (LR_{pm} and p-value$_{jw}$).

	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
Panel A: OLS									
LR_{fm}	0.30	2.02	0.03	0.03	0.00	0.00	0.23	0.00	13.52
p-value$_{fm}$	(0.50)	(0.18)	(0.62)	(0.43)	(0.71)	(0.85)	(0.52)	(0.93)	(0.00)
LR_s	0.10	0.49	0.01	0.01	0.00	0.00	0.18	0.00	11.94
p-value$_{fm}$	(0.61)	(0.45)	(0.65)	(0.46)	(0.71)	(0.85)	(0.55)	(0.93)	(0.00)
LR_{jw}	0.08	0.50	0.01	0.01	0.00	0.00	0.17	0.00	9.23
p-value$_{jw}$	(0.63)	(0.47)	(0.64)	(0.46)	(0.70)	(0.85)	(0.57)	(0.92)	(0.00)
LR_{pm}	0.08	0.30	0.01	0.01	0.00	0.00	0.13	0.00	5.81
p-value$_{pm}$	(0.62)	(0.58)	(0.66)	(0.46)	(0.71)	(0.81)	(0.58)	(0.95)	(0.02)
Panel B: GLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.00	10.58	1.26	0.00	0.00	0.00	0.00	0.00	14.15
p-value$_{fm}$	(0.72)	(0.00)	(0.29)	(0.50)	(0.72)	(0.85)	(0.73)	(0.93)	(0.00)
LR_s	0.00	3.85	0.75	0.00	0.00	0.00	0.00	0.00	12.58
p-value$_{fm}$	(0.72)	(0.08)	(0.39)	(0.50)	(0.71)	(0.85)	(0.73)	(0.93)	(0.00)
LR_{jw}	0.00	3.77	0.93	0.00	0.00	0.00	0.00	0.00	8.45
p-value$_{jw}$	(0.72)	(0.09)	(0.35)	(0.50)	(0.70)	(0.84)	(0.73)	(0.93)	(0.01)
LR_{pm}	0.00	1.66	0.48	0.00	0.00	0.00	0.00	0.00	5.23
p-value$_{pm}$	(0.71)	(0.27)	(0.43)	(0.50)	(0.70)	(0.82)	(0.73)	(0.94)	(0.03)
Panel C: WLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.00	0.18	0.02	0.00	0.00	0.00	0.00	0.00	13.17
p-value$_{fm}$	(0.72)	(0.59)	(0.65)	(0.50)	(0.71)	(0.85)	(0.73)	(0.93)	(0.00)
LR_s	0.00	0.06	0.01	0.00	0.00	0.00	0.00	0.00	11.76
p-value$_{fm}$	(0.72)	(0.67)	(0.66)	(0.50)	(0.71)	(0.85)	(0.73)	(0.93)	(0.00)
LR_{jw}	0.00	0.06	0.01	0.00	0.00	0.00	0.00	0.00	8.64
p-value$_{jw}$	(0.72)	(0.68)	(0.66)	(0.50)	(0.70)	(0.84)	(0.74)	(0.92)	(0.01)
LR_{pm}	0.00	0.05	0.01	0.00	0.00	0.00	0.00	0.00	5.50
p-value$_{pm}$	(0.72)	(0.71)	(0.67)	(0.50)	(0.70)	(0.82)	(0.73)	(0.94)	(0.02)

in Table V.1. The only model that is systematically rejected across estimation methods and by the set of all four test statistics is the FF5 model. This accurately reflects the fact that for this model we kept the sign restriction on the price of covariance risk estimate associated with the profitability (rmw) factor, which as shown in Table III.1 was statistically significant and had an inconsistent sign. Generally, we observe a decrease in p-values relative to the baseline case when removing a restriction from coefficient estimates with a consistent sign (which is the case for the HL, CV, KLVN, FF3, C, PS, and FFTD models). However, when a restrictions is removed from a coefficient estimate with an inconsistent sign (which is the case for the P model) we observe an increase in p-values since the amount of evidence against the null of sign consistency decreases.

Table V.2 presents results for the 25 size and momentum sorted portfolios. Worth noting is the fact that the FF5 model is no longer systematically rejected, and we only observe rejection of the null when estimation is done using GLS (Table V.2, Panel B). Otherwise, the evidence against the null follows closely the pattern of statistical significance of the restricted sign cross-sectional coefficient, rmw, shown in Table III.2. Other notable differences relative to the baseline case are the HL model for OLS estimation, and the KLVN model for WLS estimation, both of which are now consistent with an ICAPM interpretation. Comparing Panel A of Table V.2 with Panel A of Table IV.2, we note that the HL now satisfies the ICAPM restrictions (p-value$_{pm}$ of 39% versus p-value$_{pm}$ of 4% in the baseline case). This is due to the removal of the sign restriction on the $term$ factor which, as shown in Panel A of Table III.2, has a statistically significant cross-sectional estimate with an inconsistent sign. Similarly, the WLS p-value$_{pm}$ associated with the KLVN model increases beyond the point of rejecting the null (from 5% in the baseline case to 50%), which is due to the removal of the constraint on the $term$ factor, whose estimate was marginally significant and had an inconsistent sign (Table III.2, Panel C).

Overall, we find that reducing the number of restrictions imposed on the model coefficients reduces the evidence against the null hypothesis of sign consistency relative to the baseline case whereby restrictions are imposed on all coefficients. The null hypothesis of consistency with an ICAPM interpretation is rejected in only 7 out of 54 cases when using a 5% significance level and misspecification-robust standard errors.[6]

Finally, we explore an alternative way of setting up the matrix of constraints, $\mathcal{Q}$, when the time-series estimates are statistically insignificant. We implement the new set of restrictions by setting equal to zero the price of covariance risk corresponding to a state variable that is not a

[6]With a Newey and West (1994) automatic lag length selection adjustment, only in 3 out of 54 cases the misspecification-robust test statistics reject the null hypothesis of consistency with the ICAPM at the 5% level. Specifically, when using the size and momentum sorted portfolios as test assets the CV model for OLS estimation and the FFTD model for OLS and WLS estimation exhibit inconsistency with the ICAPM.

Table V.2
Robust Multiple Sign Restriction Tests of the Models Using the 25 Size and Momentum Portfolios as Test Assets
No restrictions are imposed if the state variables are not robust predictors

The table presents the results of the multiple sign restriction test of nine asset-pricing models for the case where we impose sign restrictions only if the respective coefficient estimates from the long-horizon predictive regressions are statistically significant; otherwise no restriction is imposed. This is a likelihood ratio test of the null hypothesis that the models satisfy the sign restrictions placed by the ICAPM $H_0 : \mathcal{Q}\lambda \geq 0_K$. The models include the ICAPM specifications proposed by Hahn and Lee (2006) (HL), Petkova (2006) (P), Campbell and Vuolteenaho (2004) (CV), Koijen et al. (2017) (KLVN), the Fama and French (1993) three-factor model (FF3), the Carhart (1997) model (C), the Pástor and Stambaugh (2003) model (PS), the Fama and French (1993) three-factor model augmented by TERM and DEF (FFTD), and the Fama and French (2015) five-factor model (FF5). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios. The data are from July 1963 to December 2018 (666 observations). We report the values of the likelihood ratio statistics and corresponding p-values obtained using the Fama and MacBeth (1973) variances under correctly specified models (LR_{fm} and p-value$_{fm}$), the Shanken (1992) and the Jagannathan and Wang (1998) variances under correctly specified models that account for the EIV problem (LR_s and p-value$_s$, and LR_{jw} and p-value$_{jw}$, respectively), and our model misspecification-robust variances (LR_{pm} and p-value$_{jw}$).

Panel A: OLS

	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	2.49	25.48	20.69	0.00	6.50	0.00	30.94	35.65	11.73
p-value$_{fm}$	(0.13)	(0.00)	(0.00)	(0.50)	(0.01)	(0.86)	(0.00)	(0.00)	(0.00)
LR_s	0.79	5.21	3.59	0.00	5.98	0.00	4.65	5.49	5.20
p-value$_{fm}$	(0.35)	(0.03)	(0.06)	(0.50)	(0.02)	(0.86)	(0.03)	(0.06)	(0.03)
LR_{jw}	0.67	2.78	2.52	0.00	5.91	0.00	1.99	2.53	2.45
p-value$_{jw}$	(0.37)	(0.12)	(0.13)	(0.50)	(0.02)	(0.85)	(0.12)	(0.25)	(0.13)
LR_{pm}	0.50	2.39	4.88	0.00	3.48	0.00	3.31	4.05	1.84
p-value$_{pm}$	(0.39)	(0.13)	(0.04)	(0.50)	(0.07)	(0.85)	(0.07)	(0.12)	(0.17)

Panel B: GLS

	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.00	9.83	1.14	0.00	0.00	0.00	0.00	3.87	17.32
p-value$_{fm}$	(0.72)	(0.00)	(0.31)	(0.50)	(0.70)	(0.86)	(0.68)	(0.12)	(0.00)
LR_s	0.00	4.22	1.09	0.00	0.00	0.00	0.00	3.02	10.45
p-value$_{fm}$	(0.72)	(0.06)	(0.32)	(0.50)	(0.70)	(0.86)	(0.68)	(0.18)	(0.00)
LR_{jw}	0.00	3.37	1.04	0.00	0.00	0.00	0.00	2.30	6.82
p-value$_{jw}$	(0.71)	(0.10)	(0.33)	(0.50)	(0.69)	(0.84)	(0.69)	(0.25)	(0.01)
LR_{pm}	0.00	2.04	0.34	0.00	0.00	0.00	0.00	1.12	4.20
p-value$_{pm}$	(0.70)	(0.20)	(0.49)	(0.50)	(0.69)	(0.84)	(0.66)	(0.47)	(0.05)

Panel C: WLS

	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	3.37	16.82	14.52	0.00	3.40	0.00	16.71	28.67	10.42
p-value$_{fm}$	(0.07)	(0.00)	(0.00)	(0.50)	(0.07)	(0.86)	(0.00)	(0.00)	(0.00)
LR_s	1.75	4.38	4.22	0.00	3.20	0.00	5.96	6.90	5.05
p-value$_{fm}$	(0.18)	(0.05)	(0.05)	(0.50)	(0.07)	(0.86)	(0.01)	(0.03)	(0.03)
LR_{jw}	1.61	2.45	3.02	0.00	3.40	0.00	2.85	3.19	2.46
p-value$_{jw}$	(0.19)	(0.15)	(0.10)	(0.50)	(0.07)	(0.84)	(0.08)	(0.18)	(0.13)
LR_{pm}	1.11	0.96	5.22	0.00	2.74	0.00	1.98	7.40	2.06
p-value$_{pm}$	(0.24)	(0.26)	(0.03)	(0.50)	(0.10)	(0.84)	(0.15)	(0.02)	(0.15)

robust predictor of future equity returns. Specifically, instead of eliminating the corresponding row from the matrix of constraints, we set each element in that row equal to zero. The results are presented in Table VI in the Appendix. The null of consistency with the ICAPM is rejected in only 1 of the 54 cases at the 5% level using misspecification-robust standard errors (namely the FFTD model when using WLS estimation and size and momentum sorted portfolios).[7]

3.4 Concluding remarks

We develop a multivariate inequality framework for testing the consistency of multifactor asset-pricing models with the time-series and cross-sectional restrictions imposed by the ICAPM. Our test is based on results in the statistics literature due to Wolak (1987, 1989) and represents one of the first applications of Wolak's methods in empirical finance, alongside the ones in Kan et al. (2013) and Gospodinov, Kan, and Robotti (2013).

We apply our test to nine multifactor models using two different sets of portfolios as test assets and three alternative estimation schemes. We find little evidence of inconsistency of popular multifactor models with the restrictions imposed by the ICAPM. Our findings are at odds with the results in Maio and Santa-Clara (2012) who argue that most models do not satisfy the restrictions imposed by the ICAPM, but are in line with Boons (2016) and Barroso et al. (2019) who use individual stock level evidence to show that most multifactor models are consistent with an ICAPM interpretation. Interestingly, using our testing framework, we are able to show that most multifactor models are consistent with the ICAPM restrictions even when portfolios instead of individual stocks are used in analysis.

In the extant literature the consistency of the models with the ICAPM restrictions is assessed by eye-balling the signs of the parameter estimates in the time-series and cross-sectional regressions. We go beyond this practice and propose a multivariate inequality test to assess the consistency of several multifactor models with the implications of the ICAPM. Specifically, our methodology accounts for the estimation error in the covariances and for the fact that the consistency of a multifactor model with the implications of the ICAPM should be evaluated using tests of joint sign restrictions across factors. We also take seriously the fact that asset-pricing models are only approximations to reality and are likely to be misspecified. Consistent with this view, we employ inference methods that are robust to model misspecification, in addition to the traditional methods that assume that the underlying model is correctly specified.

[7]When using a Newey and West (1994) automatic lag length selection adjustment all models are found to be consistent with an ICAPM interpretation.

Appendix

A.1 ***Proof of Proposition 1:***

The proof relies on the fact that $\hat{\lambda}$ is a smooth function of $\hat{\mu}$ and $\hat{V}$. Therefore, once we have the asymptotic distribution of $\hat{\mu}$ and $\hat{V}$, we can use the delta method to obtain the asymptotic distribution of $\hat{\lambda}$. Let

$$\varphi = \begin{bmatrix} \mu \\ \text{vec}(V) \end{bmatrix}, \qquad \hat{\varphi} = \begin{bmatrix} \hat{\mu} \\ \text{vec}(\hat{V}) \end{bmatrix}. \tag{A.1}$$

We first note that $\hat{\mu}$ and $\hat{V}$ can be written as the generalized method of moments (GMM) estimator that uses the moment conditions $E[r_t] = 0_{(N+K)(N+K+1)}$, where

$$r_t = \begin{bmatrix} Y_t - \mu \\ \text{vec}((Y_t - \mu)(Y_t - \mu)' - V) \end{bmatrix}. \tag{A.2}$$

Since this is an exactly identified system of moment conditions, it is straightforward to verify that under the assumption that Y_t is stationary and ergodic with finite fourth moment, we have[8]

$$\sqrt{T}(\hat{\varphi} - \varphi) \stackrel{A}{\sim} N(0_{(N+K)(N+K+1)}, S_0), \tag{A.3}$$

where

$$S_0 = \sum_{j=-\infty}^{\infty} E[r_t r_{t+j}']. \tag{A.4}$$

Using the delta method, the asymptotic distribution of $\hat{\lambda}$ under potentially misspecified models is given by

$$\sqrt{T}(\hat{\lambda} - \lambda) \;\stackrel{A}{\sim}\; N\left(0_K, \left[\frac{\partial \lambda}{\partial \varphi'}\right] S_0 \left[\frac{\partial \lambda}{\partial \varphi'}\right]'\right). \tag{A.5}$$

Define $K_{m,n}$ as a commutation matrix (see, for example, Magnus and Neudecker (1999)) such that $K_{m,n}\text{vec}(A) = \text{vec}(A')$ where A is an $m \times n$ matrix. In addition, we denote $K_{n,n}$ by K_n. Let Θ be an $N^2 \times N^2$ matrix such that $\text{vec}(\Sigma_d) = \Theta\text{vec}(\Sigma)$.[9]

[8] Note that S_0 is a singular matrix as $\hat{V}$ is symmetric, so there are redundant elements in $\hat{\varphi}$. We could have written $\hat{\varphi}$ as $[\hat{\mu}',\ \text{vech}(\hat{V})']'$, but the results are the same under both specifications.

[9] Specifically, Θ is a matrix with (i,i)-th element equal to one, where $i = 1, 1+1(N+1), 1+2(N+1), \ldots, 1+(N-1)(N+1)$, and zeros elsewhere.

(a) The partial derivatives of λ with respect to μ are given by

$$\frac{\partial\lambda}{\partial\mu_f'} = 0_{K\times K}, \tag{A.6}$$

$$\frac{\partial\lambda}{\partial\mu_R'} = A. \tag{A.7}$$

It is easy to obtain:

$$\frac{\partial \text{vec}(V_{R,f})}{\partial \text{vec}(V)'} = [I_K,\ 0_{K\times N}] \otimes [0_{N\times K},\ I_N]. \tag{A.8}$$

For the derivative of $\lambda = A\mu_R$ with respect to vec(V), we use the product rule to obtain

$$\frac{\partial\lambda}{\partial \text{vec}(V)'} = (\mu_R' W V_{R,f} \otimes I_K)\frac{\partial \text{vec}(H)}{\partial \text{vec}(V)'} + (\mu_R' W \otimes H)\frac{\partial \text{vec}(V_{f,R})}{\partial \text{vec}(V)'}. \tag{A.9}$$

The second term is given by

$$(\mu_R' W \otimes H)\frac{\partial \text{vec}(V_{f,R})}{\partial \text{vec}(V)'} = [H,\ 0_{K\times N}] \otimes [0_K',\ \mu_R' W]. \tag{A.10}$$

For the first term, we use the chain rule to obtain

$$\begin{aligned}
& (\mu_R' W V_{R,f} \otimes I_K)\frac{\partial \text{vec}(H)}{\partial \text{vec}(V)'} \\
= \; & (\mu_R' W V_{R,f} \otimes I_K)\frac{\partial \text{vec}(H)}{\partial \text{vec}(H^{-1})'}\frac{\partial \text{vec}(H^{-1})}{\partial \text{vec}(V)'} \\
= \; & -(\mu_R' W V_{R,f} \otimes I_K)(H \otimes H)\left[(V_{f,R} W \otimes I_K)\frac{\partial \text{vec}(V_{f,R})}{\partial \text{vec}(V)'} + (I_K \otimes V_{f,R} W)\frac{\partial \text{vec}(V_{R,f})}{\partial \text{vec}(V)'}\right] \\
= \; & -(\lambda' \otimes H)\left\{([0_{K\times K},\ V_{f,R} W] \otimes [I_K,\ 0_{K\times N}])\, K_{N+K} + [I_K,\ 0_{K\times N}] \otimes [0_{K\times K},\ V_{f,R} W]\right\} \\
= \; & [H,\ 0_{K\times N}] \otimes [0_K',\ -\lambda' V_{f,R} W] + [-\lambda',\ 0_N'] \otimes [0_{K\times K},\ A].
\end{aligned} \tag{A.11}$$

Combining the two terms, we have

$$\frac{\partial\lambda}{\partial \text{vec}(V)'} = [H,\ 0_{K\times N}] \otimes \left[0_K',\ e' W\right] - \left[\lambda', 0_N'\right] \otimes [0_{K\times K},\ A]. \tag{A.12}$$

Using the expression of $\partial\lambda/\partial\varphi'$, we can simplify the asymptotic variance of $\hat{\lambda}$ to

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[h_t(\varphi) h_{t+j}(\varphi)'], \tag{A.13}$$

where

$$\begin{aligned} h_t(\varphi) &= \frac{\partial\lambda}{\partial\varphi'} r_t(\varphi) \\ &= A(R_t-\mu_R) + \text{vec}\left([0_K',\ e'W][(Y_t-\mu)(Y_t-\mu)'-V] \begin{bmatrix} H \\ 0_{N\times K} \end{bmatrix} \right) \\ &\quad - \text{vec}\left([0_{K\times K},\ A][(Y_t-\mu)(Y_t-\mu)'-V] \begin{bmatrix} \lambda \\ 0_N \end{bmatrix} \right) \\ &= (\lambda_t-\lambda) + H(f_t-\mu_f)u_t - A(R_t-\mu_R)(f_t-\mu_f)'\lambda + AV_{R,f}\lambda \\ &= (\lambda_t-\lambda) + AG_t\lambda + H(f_t-\mu_f)u_t. \end{aligned} \tag{A.14}$$

This completes the proof of part (a).

(b) The partial derivatives of λ with respect to μ are the same as in the fixed weighting matrix case. For the derivative of λ with respect to $\text{vec}(V)$, we use the product rule to obtain

$$\frac{\partial\lambda}{\partial\text{vec}(V)'} = (\mu_R' V_R^{-1} V_{R,f} \otimes I_K)\frac{\partial\text{vec}(H)}{\partial\text{vec}(V)'} + (\mu_R' V_R^{-1} \otimes H)\frac{\partial\text{vec}(V_{f,R})}{\partial\text{vec}(V)'} + (\mu_R' \otimes HV_{f,R})\frac{\partial\text{vec}(V_R^{-1})}{\partial\text{vec}(V)'}. \tag{A.15}$$

The last two terms are given by

$$(\mu_R' V_R^{-1} \otimes H)\frac{\partial\text{vec}(V_{f,R})}{\partial\text{vec}(V)'} = [H,\ 0_{K\times N}] \otimes [0_K',\ \mu_R' V_R^{-1}], \tag{A.16}$$

$$(\mu_R' \otimes HV_{f,R})\frac{\partial\text{vec}(V_R^{-1})}{\partial\text{vec}(V)'} = -[0_K',\ \mu_R' V_R^{-1}] \otimes [0_{K\times K},\ A]. \tag{A.17}$$

For the first term, we use the chain rule to obtain

$$
\begin{aligned}
& (\mu_R' V_R^{-1} V_{R,f} \otimes I_K) \frac{\partial \mathrm{vec}(H)}{\partial \mathrm{vec}(V)'} \\
= \;& (\mu_R' V_R^{-1} V_{R,f} \otimes I_K) \frac{\partial \mathrm{vec}(H)}{\partial \mathrm{vec}(H^{-1})'} \frac{\partial \mathrm{vec}(H^{-1})}{\partial \mathrm{vec}(V)'} \\
= \;& -(\mu_R' V_R^{-1} V_{R,f} \otimes I_K)(H \otimes H) \left[(V_{f,R} V_R^{-1} \otimes I_K) \frac{\partial \mathrm{vec}(V_{f,R})}{\partial \mathrm{vec}(V)'} \right. \\
& \left. + (V_{f,R} \otimes V_{f,R}) \frac{\partial \mathrm{vec}(V_R^{-1})}{\partial \mathrm{vec}(V)'} + (I_K \otimes V_{f,R} V_R^{-1}) \frac{\partial \mathrm{vec}(V_{R,f})}{\partial \mathrm{vec}(V)'} \right] \\
= \;& -(\lambda' \otimes H) \left\{ \left([0_{K\times K},\ V_{f,R} V_R^{-1}] \otimes [I_K,\ 0_{K\times N}] \right) K_{N+K} \right. \\
& - [0_{K\times K},\ V_{f,R} V_R^{-1}] \otimes [0_{K\times K},\ V_{f,R} V_R^{-1}] \\
& \left. + [I_K,\ 0_{K\times N}] \otimes [0_{K\times K},\ V_{f,R} V_R^{-1}] \right\} \\
= \;& [H,\ 0_{K\times N}] \otimes [0_K',\ -\lambda' V_{f,R} V_R^{-1}] \\
& + [-\lambda',\ \lambda' V_{f,R} V_R^{-1}] \otimes [0_{K\times K},\ A]. \qquad \text{(A.18)}
\end{aligned}
$$

Combining the three terms, we have

$$
\frac{\partial \lambda}{\partial \mathrm{vec}(V)'} = [H,\ 0_{K\times N}] \otimes \left[0_K',\ e' V_R^{-1}\right] - \left[\lambda',\ e' V_R^{-1}\right] \otimes [0_{K\times K},\ A]. \qquad \text{(A.19)}
$$

Using the expression of $\partial\lambda/\partial\varphi'$, we can simplify the asymptotic variance of $\hat{\lambda}$ to

$$
V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[h_t(\varphi) h_{t+j}(\varphi)'], \qquad \text{(A.20)}
$$

where

$$
\begin{aligned}
h_t(\varphi) &= \frac{\partial \lambda}{\partial \varphi'} r_t(\varphi) \\
&= A(R_t - \mu_R) + \mathrm{vec}\left([0_K',\ e' V_R^{-1}][(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} H \\ 0_{N\times K} \end{bmatrix} \right) \\
&\quad - \mathrm{vec}\left([0_{K\times K},\ A][(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} \lambda \\ V_R^{-1} e \end{bmatrix} \right) \\
&= (\lambda_t - \lambda) + H(f_t - \mu_f)u_t - A(R_t - \mu_R)(f_t - \mu_f)'\lambda - A(R_t - \mu_R)u_t + A V_{R,f}\lambda \\
&= (\lambda_t - \lambda) + A G_t \lambda + H(f_t - \mu_f)u_t - (\lambda_t - \lambda)u_t. \qquad \text{(A.21)}
\end{aligned}
$$

This completes the proof of part (b).

(c) The partial derivatives of λ with respect to μ are the same as in the fixed weighting matrix case. For the derivative of λ with respect to $\text{vec}(V)$, we use the product rule to obtain

$$\frac{\partial\lambda}{\partial\text{vec}(V)'} = (\mu_R'\Sigma_d^{-1}V_{R,f}\otimes I_K)\frac{\partial\text{vec}(H)}{\partial\text{vec}(V)'}+(\mu_R'\Sigma_d^{-1}\otimes H)\frac{\partial\text{vec}(V_{f,R})}{\partial\text{vec}(V)'}+(\mu_R'\otimes HV_{f,R})\frac{\partial\text{vec}(\Sigma_d^{-1})}{\partial\text{vec}(V)'}. \quad \text{(A.22)}$$

The last two terms are given by

$$(\mu_R'\Sigma_d^{-1}\otimes H)\frac{\partial\text{vec}(V_{f,R})}{\partial\text{vec}(V)'} = [H, 0_{K\times N}]\otimes[0_K',\ \mu_R'\Sigma_d^{-1}], \quad \text{(A.23)}$$

$$(\mu_R'\otimes HV_{f,R})\frac{\partial\text{vec}(\Sigma_d^{-1})}{\partial\text{vec}(V)'} = -(\mu_R'\Sigma_d^{-1}\otimes A)\Theta([-\beta,\ I_N]\otimes[-\beta,\ I_N]). \quad \text{(A.24)}$$

For the first term, we use the chain rule to obtain

$$\begin{aligned}
&(\mu_R'\Sigma_d^{-1}V_{R,f}\otimes I_K)\frac{\partial\text{vec}(H)}{\partial\text{vec}(V)'}\\
=\ &(\mu_R'\Sigma_d^{-1}V_{R,f}\otimes I_K)\frac{\partial\text{vec}(H)}{\partial\text{vec}(H^{-1})'}\frac{\partial\text{vec}(H^{-1})}{\partial\text{vec}(V)'}\\
=\ &-(\mu_R'\Sigma_d^{-1}V_{R,f}\otimes I_K)(H\otimes H)\left[(V_{f,R}\Sigma_d^{-1}\otimes I_K)\frac{\partial\text{vec}(V_{f,R})}{\partial\text{vec}(V)'}\right.\\
&\left.+(V_{f,R}\otimes V_{f,R})\frac{\partial\text{vec}(\Sigma_d^{-1})}{\partial\text{vec}(V)'}+(I_K\otimes V_{f,R}\Sigma_d^{-1})\frac{\partial\text{vec}(V_{R,f})}{\partial\text{vec}(V)'}\right]\\
=\ &-(\lambda'\otimes H)\left\{\left([0_{K\times K},\ V_{f,R}\Sigma_d^{-1}]\otimes[I_K,\ 0_{K\times N}]\right)K_{N+K}\right.\\
&-(V_{f,R}\Sigma_d^{-1}\otimes V_{f,R}\Sigma_d^{-1})\Theta([-\beta,\ I_N]\otimes[-\beta,\ I_N])\\
&\left.+[I_K,\ 0_{K\times N}]\otimes[0_{K\times K},\ V_{f,R}\Sigma_d^{-1}]\right\}\\
=\ &[H,\ 0_{K\times N}]\otimes[0_K',\ -\lambda'V_{f,R}\Sigma_d^{-1}]\\
&+(\lambda'V_{f,R}\Sigma_d^{-1}\otimes A)\Theta([-\beta,\ I_N]\otimes[-\beta,\ I_N])-[\lambda',\ 0_N']\otimes[0_{K\times K},\ A]. \quad \text{(A.25)}
\end{aligned}$$

Combining the three terms, we have

$$\begin{aligned}
\frac{\partial\lambda}{\partial\text{vec}(V)'} =\ &[H,\ 0_{K\times N}]\otimes\left[0_K',\ e'\Sigma_d^{-1}\right]\\
&-[\lambda',\ 0_N']\otimes[0_{K\times K},\ A]-(e'\Sigma_d^{-1}\otimes A)\Theta([-\beta,\ I_N]\otimes[-\beta,\ I_N]). \quad \text{(A.26)}
\end{aligned}$$

Using the expression of $\partial\lambda/\partial\varphi'$, we can simplify the asymptotic variance of $\hat{\lambda}$ to

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[h_t(\varphi)h_{t+j}(\varphi)'], \tag{A.27}$$

where

$$\begin{aligned}
h_t(\varphi) &= \frac{\partial\lambda}{\partial\varphi'} r_t(\varphi) \\
&= A(R_t - \mu_R) + \mathrm{vec}\left([0_K',\ e'\Sigma_d^{-1}][(Y_t-\mu)(Y_t-\mu)' - V] \begin{bmatrix} H \\ 0_{N\times K} \end{bmatrix} \right) \\
&\quad - \mathrm{vec}\left([0_{K\times K},\ A][(Y_t-\mu)(Y_t-\mu)' - V] \begin{bmatrix} \lambda \\ 0_N \end{bmatrix} \right) \\
&\quad - (e'\Sigma_d^{-1} \otimes A)\Theta \mathrm{vec}\left([-\beta,\ I_N][(Y_t-\mu)(Y_t-\mu)' - V] \begin{bmatrix} -\beta' \\ I_N \end{bmatrix} \right) \\
&= (\lambda_t - \lambda) + H(f_t - \mu_f)u_t - A(R_t - \mu_R)(f_t - \mu_f)'\lambda \\
&\quad + AV_{R,f}\lambda - (e'\Sigma_d^{-1} \otimes A)\Theta \mathrm{vec}(\epsilon_t\epsilon_t' - \Sigma) \\
&= (\lambda_t - \lambda) + AG_t\lambda + H(f_t - \mu_f)u_t - A\Psi_t\Sigma_d^{-1}e.
\end{aligned} \tag{A.28}$$

The second last equality follows from the first order condition $V_{f,R}\Sigma_d^{-1}e = 0_K$. This completes the proof of part (c).

Note that when the model is correctly specified, we have $e = 0_N$ and $u_t = 0$. In this case, we have

$$h_t(\varphi) = (\lambda_t - \lambda) + AG_t\lambda. \tag{A.29}$$

This completes the proof of Proposition 1.

A.2 *Proof of Proposition 2:*

(a) We first derive the asymptotic distribution of

$$T\hat{Q} = T(\hat{\mu}_R'\hat{W}\hat{\mu}_R - \hat{\mu}_R'\hat{W}\hat{\beta}(\hat{\beta}'\hat{W}\hat{\beta})^{-1}\hat{\beta}'\hat{W}\hat{\mu}_R) \tag{A.30}$$

under $H_0 : \rho^2 = 1$, where $\hat{W} \xrightarrow{\text{a.s.}} W$ (this includes the known weighting matrix case as a special case). This can be accomplished by using the GMM results of Hansen (1982).

Let $\theta = (\theta_1',\ \theta_2')'$, where $\theta_1 = (\alpha',\ \text{vec}(\beta)')'$ and $\theta_2 = \gamma$. Define

$$g_t(\theta) \equiv \begin{bmatrix} g_{1t}(\theta_1) \\ g_{2t}(\theta) \end{bmatrix} = \begin{bmatrix} l_t \otimes \epsilon_t \\ R_t - \beta\gamma \end{bmatrix}, \tag{A.31}$$

where $l_t = [1,\ f_t']'$ and $\epsilon_t = R_t - \alpha - \beta f_t$. When the model is correctly specified, we have $E[g_t(\theta)] = 0_{p+N}$, where $p = N(K+1)$. The sample moments of $g_t(\theta)$ are given by

$$\bar{g}_T(\theta) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} g_{1t}(\theta_1) \\ \frac{1}{T}\sum_{t=1}^{T} g_{2t}(\theta) \end{bmatrix}. \tag{A.32}$$

Let $\hat{\theta} = (\hat{\theta}_1',\ \hat{\theta}_2')'$, where $\hat{\theta}_1 = (\hat{\alpha}',\ \text{vec}(\hat{\beta})')'$ is the OLS estimator of α and β, and

$$\hat{\theta}_2 = \hat{\gamma} = (\hat{\beta}'\hat{W}\hat{\beta})^{-1}\hat{\beta}'\hat{W}\hat{\mu}_R \tag{A.33}$$

is the second-pass CSR estimator of γ. Note that $\hat{\theta}$ is the solution to the following first-order condition

$$B_T\bar{g}_T(\theta) = 0_{p+K}, \tag{A.34}$$

where

$$B_T = \begin{bmatrix} I_p & 0_{p\times N} \\ 0_{K\times p} & \hat{\beta}'\hat{W} \end{bmatrix} \xrightarrow{\text{a.s.}} \begin{bmatrix} I_p & 0_{p\times N} \\ 0_{K\times p} & \beta'W \end{bmatrix} \equiv B. \tag{A.35}$$

Writing

$$\begin{aligned} l_t \otimes \epsilon_t &= \text{vec}(\epsilon_t l_t') = (l_t \otimes I_N)\text{vec}(\epsilon_t), && \text{(A.36)} \\ \epsilon_t &= R_t - \alpha - \beta f_t = R_t - (l_t' \otimes I_N)\theta_1, && \text{(A.37)} \\ \beta\gamma &= (\gamma' \otimes I_N)\text{vec}(\beta), && \text{(A.38)} \end{aligned}$$

we have:

$$\begin{aligned}
\frac{\partial g_{1t}(\theta_1)}{\partial \theta_1'} &= -l_t l_t' \otimes I_N, & \text{(A.39)}\\
\frac{\partial g_{1t}(\theta_1)}{\partial \theta_2'} &= 0_{p\times K}, & \text{(A.40)}\\
\frac{\partial g_{2t}(\theta)}{\partial \theta_1'} &= [0,\ -\gamma'] \otimes I_N, & \text{(A.41)}\\
\frac{\partial g_{2t}(\theta)}{\partial \theta_2'} &= -\beta. & \text{(A.42)}
\end{aligned}$$

Let

$$\begin{aligned}
D_T &= \frac{\partial \bar{g}_T(\theta)}{\partial \theta'} \\
&= \begin{bmatrix} -\left(\frac{1}{T}\sum_{t=1}^{T} l_t l_t'\right) \otimes I_N & 0_{p\times K} \\ [0,\ -\gamma'] \otimes I_N & -\beta \end{bmatrix} \\
&\xrightarrow{\text{a.s.}} \begin{bmatrix} -E[l_t l_t'] \otimes I_N & 0_{p\times K} \\ [0,\ -\gamma'] \otimes I_N & -\beta \end{bmatrix} \equiv D. \qquad \text{(A.43)}
\end{aligned}$$

Hansen (1982, Lemma 4.1) shows that when the model is correctly specified, we have:

$$\sqrt{T}\bar{g}_T(\hat{\theta}) \overset{A}{\sim} N(0_{p+N}, [I_{p+N} - D(BD)^{-1}B]S[I_{p+N} - D(BD)^{-1}B]'), \qquad \text{(A.44)}$$

where

$$S = \sum_{j=-\infty}^{\infty} E[g_t(\theta) g_{t+j}(\theta)']. \qquad \text{(A.45)}$$

Using the partitioned matrix inverse formula, it is easy to verify that

$$E[l_t l_t']^{-1} = \begin{bmatrix} 1 + \mu_f' V_f^{-1} \mu_f & -\mu_f' V_f^{-1} \\ -V_f^{-1}\mu_f & V_f^{-1} \end{bmatrix}. \qquad \text{(A.46)}$$

It follows that

$$
\begin{aligned}
BD &= \begin{bmatrix} -E[l_t l_t'] \otimes I_N & 0_{p\times K} \\ [0,\ -\gamma'] \otimes \beta' W & -H^{-1} \end{bmatrix}, & \text{(A.47)} \\
(BD)^{-1} &= \begin{bmatrix} -E[l_t l_t']^{-1} \otimes I_N & 0_{p\times K} \\ [-\gamma' V_f^{-1}\mu_f,\ \gamma' V_f^{-1}] \otimes A & -H \end{bmatrix}, & \text{(A.48)} \\
D(BD)^{-1}B &= \begin{bmatrix} I_p & 0_{p\times N} \\ [-\gamma' V_f^{-1}\mu_f,\ \gamma' V_f^{-1}] \otimes (I_N - \beta A) & -\beta A \end{bmatrix}, & \text{(A.49)} \\
I_N - D(BD)^{-1}B &= \begin{bmatrix} 0_{p\times p} & 0_{p\times N} \\ [\gamma' V_f^{-1}\mu_f,\ -\gamma' V_f^{-1}] \otimes (I_N - \beta A) & I_N - \beta A \end{bmatrix}. & \text{(A.50)}
\end{aligned}
$$

We now provide a simplification of the asymptotic distribution of $\bar{g}_{2T}(\hat{\theta})$. From (A.44), we have:

$$\sqrt{T}\bar{g}_{2T}(\hat{\theta}) \overset{A}{\sim} N(0_N, V_q), \tag{A.51}$$

where

$$V_q = \sum_{j=-\infty}^{\infty} E[q_t(\theta) q_{t+j}(\theta)'], \tag{A.52}$$

and

$$
\begin{aligned}
q_t(\theta) &= [0_{N\times p},\ I_N][I_{p+N} - D(BD)^{-1}B] g_t(\theta) \\
&= -(I_N - \beta A)\epsilon_t \gamma' V_f^{-1}(f_t - \mu_f) + (I_N - \beta A)(R_t - \beta\gamma) \\
&= (I_N - \beta A)[R_t - \epsilon_t \gamma' V_f^{-1}(f_t - \mu_f)] \\
&= (I_N - \beta A)\epsilon_t y_t \\
&= [I_N - \beta(\beta' W \beta)^{-1}\beta' W]\epsilon_t y_t \\
&= W^{-\frac{1}{2}}[I_N - W^{\frac{1}{2}}\beta(\beta' W\beta)^{-1}\beta' W^{\frac{1}{2}}]W^{\frac{1}{2}}\epsilon_t y_t \\
&= W^{-\frac{1}{2}}[I_N - W^{\frac{1}{2}} V_{R,f}(V_{f,R} W V_{R,f})^{-1} V_{f,R} W^{\frac{1}{2}}]W^{\frac{1}{2}}\epsilon_t y_t \\
&= W^{-\frac{1}{2}} P P' W^{\frac{1}{2}} \epsilon_t y_t, & \text{(A.53)}
\end{aligned}
$$

where $y_t = 1 - \lambda'(f_t - \mu_f) = 1 - \gamma' V_f^{-1}(f_t - \mu_f)$. The fourth equality follows from the fact that, under $H_0 : \rho^2 = 1$, $(I_N - \beta A)R_t = (I_N - \beta A)\epsilon_t$. With this expression of q_t, we can write V_q as

$$V_q = W^{-\frac{1}{2}} PP' W^{\frac{1}{2}} S W^{\frac{1}{2}} PP' W^{-\frac{1}{2}}, \tag{A.54}$$

where S is the asymptotic covariance matrix of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t y_t$. Having derived the asymptotic distribution of $\bar{g}_{2T}(\hat{\theta})$, the asymptotic distribution of $\hat{Q}$ is given by

$$T\hat{Q} = T\bar{g}_{2T}(\hat{\theta})' \hat{W} \bar{g}_{2T}(\theta) \overset{A}{\sim} \sum_{j=1}^{N-K} \xi_j x_j, \tag{A.55}$$

where the x_j's are independent χ^2_1 random variables, and the ξ_j's are the $N-K$ nonzero eigenvalues of

$$W^{\frac{1}{2}} V_q W^{\frac{1}{2}} = PP' W^{\frac{1}{2}} S W^{\frac{1}{2}} PP'. \tag{A.56}$$

Equivalently, the ξ_j's are the eigenvalues of $P' W^{\frac{1}{2}} S W^{\frac{1}{2}} P$. Since $\hat{Q}_0 \xrightarrow{\text{a.s.}} Q_0 > 0$, we have:

$$T(\hat{\rho}^2 - 1) = -\frac{T\hat{Q}}{\hat{Q}_0} \overset{A}{\sim} -\sum_{j=1}^{N-K} \frac{\xi_j}{Q_0} x_j. \tag{A.57}$$

This completes the proof of part (a).

(b) The proof uses the same notation and delta method employed in Proposition 1 to obtain the asymptotic distribution of $\hat{\rho}^2$ as

$$\sqrt{T}(\hat{\rho}^2 - \rho^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[n_t n_{t+j}]\right), \tag{A.58}$$

where

$$n_t = \frac{\partial \rho^2}{\partial \varphi'} r_t(\varphi). \tag{A.59}$$

Obtaining an explicit expression for n_t requires computing $\partial \rho^2 / \partial \varphi'$. For the known weighting matrix case and the estimated GLS and WLS cases, we have

$$\frac{\partial \rho^2}{\partial \mu_f} = 0_K, \tag{A.60}$$

$$\frac{\partial \rho^2}{\partial \mu_R} = 2Q_0^{-1} W[(1-\rho^2)\mu_R - e]. \tag{A.61}$$

Equation (A.60) follows because ρ^2 does not depend on μ_f. For (A.61), using the first order condition $\beta' W e = 0_K$ and letting $Q_0 = \mu_R' W \mu_R$, we have

$$\frac{\partial Q_0}{\partial \mu_R} = 2W\mu_R, \qquad \frac{\partial Q}{\partial \mu_R} = 2We. \tag{A.62}$$

It follows that

$$\frac{\partial \rho^2}{\partial \mu_R} = -Q_0^{-1}\frac{\partial Q}{\partial \mu_R} + Q_0^{-2}Q\frac{\partial Q_0}{\partial \mu_R} = -2Q_0^{-1}We + 2QQ_0^{-2}W\mu_R = 2Q_0^{-1}W[(1-\rho^2)\mu_R - e]. \tag{A.63}$$

The expression for $\partial\rho^2/\partial\text{vec}(V)'$, however, depends on whether we use a known W or an estimate of W, say $\hat{W}$, as the weighting matrix. We start with the known weighting matrix W case. Differentiating $Q = e'We$ with respect to $\text{vec}(V)$, we obtain:

$$\frac{\partial Q}{\partial \text{vec}(V)'} = 2e'W\frac{\partial(\mu_R - \beta\gamma)}{\partial \text{vec}(V)'} = -2e'W\left[(\gamma' \otimes I_N)\frac{\partial \text{vec}(\beta)}{\partial \text{vec}(V)'} + \beta\frac{\partial \gamma}{\partial \text{vec}(V)'}\right]. \tag{A.64}$$

Note that the second term vanishes because of the first order condition $\beta' W e = 0_K$. Using

$$\frac{\partial \text{vec}(\beta)}{\partial \text{vec}(V)'} = [V_f^{-1},\ 0_{K\times N}] \otimes [-\beta,\ I_N]. \tag{A.65}$$

for the first term and the fact that $\beta' W e = 0_K$ gives

$$\frac{\partial Q}{\partial \text{vec}(V)'} = -2e'W\left([\gamma' V_f^{-1},\ 0_N'] \otimes [-\beta,\ I_N]\right) = -2\left([\gamma' V_f^{-1},\ 0_N'] \otimes [0_K',\ e'W]\right). \tag{A.66}$$

Since $Q_0 = \mu_R' W \mu_R$ does not depend on V, we have:

$$\frac{\partial \rho^2}{\partial \text{vec}(V)'} = -Q_0^{-1}\frac{\partial Q}{\partial \text{vec}(V)'} = 2Q_0^{-1}\left[\gamma' V_f^{-1},\ 0_N'\right] \otimes \left[0_K',\ e'W\right]. \tag{A.67}$$

Therefore, for the known weighting matrix W case, n_t is given by

$$\begin{aligned} n_t &= \frac{\partial \rho^2}{\partial \varphi'} r_t(\varphi) \\ &= 2Q_0^{-1}[(1-\rho^2)\mu_R' - e']W(R_t - \mu_R) + 2Q_0^{-1}e'W(R_t - \mu_R)(f_t - \mu_f)'V_f^{-1}\gamma \\ &= 2Q_0^{-1}[-u_t y_t + (1-\rho^2)v_t]. \end{aligned} \tag{A.68}$$

We now turn to the $\hat{W} = \hat{V}_R^{-1}$ case. Differentiating $Q = e'V_R^{-1}e$ with respect to $\text{vec}(V)$, we obtain:

$$\begin{aligned}\frac{\partial Q}{\partial \text{vec}(V)'} &= 2e'V_R^{-1}\frac{\partial(\mu_R-\beta\gamma)}{\partial \text{vec}(V)'} + (e'\otimes e')\frac{\partial \text{vec}(V_R^{-1})}{\partial \text{vec}(V)'} \\ &= -2\left([\gamma' V_f^{-1},\ 0_N']\otimes[0_K',\ e'V_R^{-1}]\right) - (e'\otimes e')\left([0_{N\times K},\ V_R^{-1}]\otimes[0_{N\times K},\ V_R^{-1}]\right) \\ &= -[2\gamma' V_f^{-1},\ e'V_R^{-1}]\otimes[0_K',\ e'V_R^{-1}]. \qquad (A.69)\end{aligned}$$

Similarly, we have:

$$\frac{\partial Q_0}{\partial \text{vec}(V)'} = -[0_K',\ \mu_R'V_R^{-1}]\otimes[0_K',\ \mu_R'V_R^{-1}]. \qquad (A.70)$$

It follows that

$$\begin{aligned}\frac{\partial \rho^2}{\partial \text{vec}(V)'} &= -Q_0^{-1}\frac{\partial Q}{\partial \text{vec}(V)'} + Q_0^{-2}Q\frac{\partial Q_0}{\partial \text{vec}(V)'} \\ &= Q_0^{-1}\left[2\gamma' V_f^{-1},\ e'V_R^{-1}\right]\otimes\left[0_K',\ e'V_R^{-1}\right] \\ &\quad - Q_0^{-1}(1-\rho^2)\left[0_K',\ \mu_R'V_R^{-1}\right]\otimes\left[0_K',\ \mu_R'V_R^{-1}\right]. \qquad (A.71)\end{aligned}$$

Therefore, we have:

$$\begin{aligned} n_t &= \frac{\partial \rho^2}{\partial \varphi'} r_t(\varphi) \\ &= 2Q_0^{-1}[(1-\rho^2)\mu_R' - e']V_R^{-1}(R_t-\mu_R) + Q_0^{-1}e'V_R^{-1}(R_t-\mu_R)[2\gamma' V_f^{-1}(f_t-\mu_f) \\ &\quad + e'V_R^{-1}(R_t-\mu_R)] - Q_0^{-1}(1-\rho^2)[\mu_R'V_R^{-1}(R_t-\mu_R)]^2 - Q_0^{-1}Q + Q_0^{-1}(1-\rho^2)Q_0 \\ &= Q_0^{-1}[u_t^2 - 2u_t y_t + (1-\rho^2)(2v_t - v_t^2)]. \qquad (A.72)\end{aligned}$$

Finally, for the WLS case, we can use

$$\frac{\partial \text{vec}(\Sigma_d^{-1})}{\partial \text{vec}(V)'} = \frac{\partial \text{vec}(\Sigma_d^{-1})}{\partial \text{vec}(\Sigma_d)'}\frac{\partial \text{vec}(\Sigma_d)}{\partial \text{vec}(\Sigma)'}\frac{\partial \text{vec}(\Sigma)}{\partial \text{vec}(V)'} = -(\Sigma_d^{-1}\otimes\Sigma_d^{-1})\Theta([-\beta,\ I_N]\otimes[-\beta,\ I_N]). \qquad (A.73)$$

and show that

$$\begin{aligned}\frac{\partial \rho^2}{\partial \text{vec}(V)'} &= Q_0^{-1}\left\{\left[2\gamma' V_f^{-1},\ 0_N'\right]\otimes\left[0_K',\ e'\Sigma_d^{-1}\right] + \left(e'\Sigma_d^{-1}\otimes e'\Sigma_d^{-1}\right)\Theta\left([-\beta,\ I_N]\otimes[-\beta,\ I_N]\right)\right\} \\ &\quad - Q_0^{-1}(1-\rho^2)\left(\mu_R'\Sigma_d^{-1}\otimes\mu_R'\Sigma_d^{-1}\right)\Theta\left([-\beta,\ I_N]\otimes[-\beta,\ I_N]\right). \qquad (A.74)\end{aligned}$$

It is then straightforward to obtain

$$
\begin{aligned}
n_t &= \frac{\partial\rho^2}{\partial\varphi'} r_t(\varphi) \\
&= 2Q_0^{-1}[(1-\rho^2)v_t - u_t] + 2Q_0^{-1}u_t\gamma' V_f^{-1}(f_t - \mu_f) + Q_0^{-1}e'\Sigma_d^{-1}\mathrm{Diag}(\epsilon_t\epsilon_t')\Sigma_d^{-1}e \\
&\quad - Q_0^{-1}(1-\rho^2)\mu_R'\Sigma_d^{-1}\mathrm{Diag}(\epsilon_t\epsilon_t')\Sigma_d^{-1}\mu_R - Q_0^{-1}Q + Q_0^{-1}(1-\rho^2)Q_0 \\
&= Q_0^{-1}\left[-2u_t y_t + e'\Gamma_t e + (1-\rho^2)(2v_t - \mu_R'\Gamma_t\mu_R)\right]. \qquad \text{(A.75)}
\end{aligned}
$$

This completes the proof of part (b).

(c) We start by rewriting $Q_0 - Q$ as

$$
\begin{aligned}
Q_0 - Q &= \mu_R' W V_{R,f}(V_{f,R} W V_{R,f})^{-1} V_{f,R} W \mu_R \\
&= \lambda'(V_{f,R} W V_{R,f})\lambda. \qquad \text{(A.76)}
\end{aligned}
$$

The matrix in the middle is positive definite because $V_{R,f}$ is assumed to be of full column rank. Therefore, the necessary and sufficient condition for $Q_0 = Q$ (that is, $\rho^2 = 0$) is $\lambda = 0_K$. Note that (A.76) also holds for its sample counterpart. As a consequence, we can write $\hat{\rho}^2$ as

$$
\hat{\rho}^2 = 1 - \frac{\hat{Q}}{\hat{Q}_0} = \frac{\hat{Q}_0 - \hat{Q}}{\hat{Q}_0} = \frac{\hat{\lambda}'(\hat{V}_{f,R}\hat{W}\hat{V}_{R,f})\hat{\lambda}}{\hat{Q}_0}. \qquad \text{(A.77)}
$$

Under the null hypothesis $H_0 : \lambda = 0_K$, we have:

$$
\sqrt{T}\hat{\lambda} \overset{A}{\sim} N(0_K, V(\hat{\lambda})), \qquad \text{(A.78)}
$$

where $V(\hat{\lambda})$ is the asymptotic variance of $\hat{\lambda}$ obtained under the potentially misspecified model. As $\hat{Q}_0 \xrightarrow{\text{a.s.}} Q_0 > 0$ and

$$
\hat{V}_{f,R}\hat{W}\hat{V}_{R,f} \xrightarrow{\text{a.s.}} V_{f,R} W V_{R,f}, \qquad \text{(A.79)}
$$

it follows that

$$
T\hat{\rho}^2 \overset{A}{\sim} \sum_{j=1}^{K} \frac{\xi_j}{Q_0} x_j, \qquad \text{(A.80)}
$$

where the x_j's are independent χ_1^2 random variables and the ξ_j's are the eigenvalues of

$$
(V_{f,R} W V_{R,f}) V(\hat{\lambda}). \qquad \text{(A.81)}
$$

This completes the proof of part (c).

Table VI.1
Robust Multiple Sign Restriction Tests of the Models Using the 25 Size and Book-to-Market Portfolios as Test Assets
Zero restrictions are imposed if the state variables are not robust predictors

The table presents the results of the multiple sign restriction test of nine asset-pricing models for the case where we impose sign restrictions only if the respective coefficient estimates from the long-horizon predictive regressions are statistically significant; otherwise a zero restriction is imposed. This is a likelihood ratio test of the null hypothesis that the models satisfy the sign restrictions placed by the ICAPM $H_0 : \mathcal{Q}\lambda \geq 0_K$. The models include the ICAPM specifications proposed by Hahn and Lee (2006)(HL), Petkova (2006) (P), Campbell and Vuolteenaho (2004) (CV), Koijen, Lustig, and Van Nieuwerburgh (2017) (KLVN), the Fama and French (1993) three-factor model (FF3), the Carhart (1997) model (C), the Pastor and Stambaugh (2003) model (PS), the Fama and French (1993) three-factor model augmented by TERM and DEF (FFTD), and the Fama and French (2015) five-factor model (FF5). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios. The data are from July 1963 to December 2018 (666 observations). We report the values of the likelihood ratio statistics and corresponding p-values obtained using the Fama and MacBeth (1973) variances under correctly specified models (LR_{fm} and p-value$_{fm}$), the Shanken (1992) and the Jagannathan and Wang (1998) variances under correctly specified models that account for the EIV problem (LR_s and p-value$_s$, and LR_{jw} and p-value$_{jw}$, respectively), and our model misspecification-robust variances (LR_{pm} and p-value$_{jw}$).

Panel A: OLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.30	2.02	0.03	0.03	0.00	0.00	0.23	0.00	13.52
p-value$_{fm}$	(0.66)	(0.43)	(0.80)	(0.82)	(0.86)	(0.93)	(0.79)	(0.98)	(0.00)
LR_s	0.10	0.49	0.01	0.01	0.00	0.00	0.18	0.00	11.94
p-value$_{fm}$	(0.76)	(0.76)	(0.82)	(0.84)	(0.86)	(0.92)	(0.82)	(0.98)	(0.01)
LR_{jw}	0.08	0.50	0.01	0.01	0.00	0.00	0.17	0.00	9.23
p-value$_{jw}$	(0.78)	(0.76)	(0.81)	(0.84)	(0.85)	(0.93)	(0.82)	(0.98)	(0.02)
LR_{pm}	0.08	0.30	0.01	0.01	0.00	0.00	0.13	0.00	5.81
p-value$_{pm}$	(0.77)	(0.83)	(0.82)	(0.84)	(0.85)	(0.91)	(0.84)	(0.99)	(0.10)
Panel B: GLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.00	10.58	1.26	0.00	0.00	0.00	0.00	0.00	14.15
p-value$_{fm}$	(0.86)	(0.01)	(0.45)	(0.88)	(0.86)	(0.93)	(0.93)	(0.98)	(0.00)
LR_s	0.00	3.85	0.75	0.00	0.00	0.00	0.00	0.00	12.58
p-value$_{fm}$	(0.86)	(0.22)	(0.57)	(0.88)	(0.86)	(0.92)	(0.93)	(0.98)	(0.01)
LR_{jw}	0.00	3.77	0.93	0.00	0.00	0.00	0.00	0.00	8.45
p-value$_{jw}$	(0.86)	(0.23)	(0.52)	(0.88)	(0.85)	(0.92)	(0.93)	(0.98)	(0.03)
LR_{pm}	0.00	1.66	0.48	0.00	0.00	0.00	0.00	0.00	5.23
p-value$_{pm}$	(0.85)	(0.51)	(0.63)	(0.88)	(0.85)	(0.91)	(0.93)	(0.99)	(0.13)
Panel C: WLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.00	0.18	0.02	0.00	0.00	0.00	0.00	0.00	13.17
p-value$_{fm}$	(0.86)	(0.85)	(0.82)	(0.88)	(0.86)	(0.92)	(0.93)	(0.98)	(0.00)
LR_s	0.00	0.06	0.01	0.00	0.00	0.00	0.00	0.00	11.76
p-value$_{fm}$	(0.86)	(0.90)	(0.83)	(0.88)	(0.85)	(0.92)	(0.93)	(0.98)	(0.01)
LR_{jw}	0.00	0.06	0.01	0.00	0.00	0.00	0.00	0.00	8.64
p-value$_{jw}$	(0.86)	(0.90)	(0.82)	(0.88)	(0.85)	(0.92)	(0.93)	(0.98)	(0.03)
LR_{pm}	0.00	0.05	0.01	0.00	0.00	0.00	0.00	0.00	5.50
p-value$_{pm}$	(0.86)	(0.91)	(0.83)	(0.88)	(0.85)	(0.91)	(0.93)	(0.99)	(0.11)

Table VI.2
Robust Multiple Sign Restriction Tests of the Models Using the 25 Size and Momentum Portfolios as Test Assets
Zero restrictions are imposed if the state variables are not robust predictors

The table presents the results of the multiple sign restriction test of nine asset-pricing models for the case where we impose sign restrictions only if the respective coefficient estimates from the long-horizon predictive regressions are statistically significant; otherwise a zero restriction is imposed. This is a likelihood ratio test of the null hypothesis that the models satisfy the sign restrictions placed by the ICAPM $H_0 : \mathcal{Q}\lambda \geq 0_K$. The models include the ICAPM specifications proposed by Hahn and Lee (2006)(HL), Petkova (2006) (P), Campbell and Vuolteenaho (2004) (CV), Koijen, Lustig, and Van Nieuwerburgh (2017) (KLVN), the Fama and French (1993) three-factor model (FF3), the Carhart (1997) model (C), the Pastor and Stambaugh (2003) model (PS), the Fama and French (1993) three-factor model augmented by TERM and DEF (FFTD), and the Fama and French (2015) five-factor model (FF5). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios. The data are from July 1963 to December 2018 (666 observations). We report the values of the likelihood ratio statistics and corresponding p-values obtained using the Fama and MacBeth (1973) variances under correctly specified models (LR_{fm} and p-value$_{fm}$), the Shanken (1992) and the Jagannathan and Wang (1998) variances under correctly specified models that account for the EIV problem (LR_s and p-value$_s$, and LR_{jw} and p-value$_{jw}$, respectively), and our model misspecification-robust variances (LR_{pm} and p-value$_{jw}$).

Panel A: OLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	2.49	25.48	20.69	0.00	6.50	0.00	30.94	35.65	11.73
p-value$_{fm}$	(0.21)	(0.00)	(0.00)	(0.88)	(0.03)	(0.93)	(0.00)	(0.00)	(0.01)
LR_s	0.79	5.21	3.59	0.00	5.98	0.00	4.65	5.49	5.20
p-value$_{fm}$	(0.50)	(0.12)	(0.15)	(0.88)	(0.04)	(0.93)	(0.11)	(0.12)	(0.13)
LR_{jw}	0.67	2.78	2.52	0.00	5.91	0.00	1.99	2.53	2.45
p-value$_{jw}$	(0.52)	(0.33)	(0.25)	(0.88)	(0.04)	(0.92)	(0.35)	(0.39)	(0.39)
LR_{pm}	0.50	2.39	4.88	0.00	3.48	0.00	3.31	4.05	1.84
p-value$_{pm}$	(0.56)	(0.37)	(0.09)	(0.88)	(0.12)	(0.92)	(0.21)	(0.21)	(0.49)
Panel B: GLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	0.00	9.83	1.14	0.00	0.00	0.00	0.00	3.87	17.32
p-value$_{fm}$	(0.86)	(0.02)	(0.48)	(0.88)	(0.85)	(0.93)	(0.92)	(0.23)	(0.00)
LR_s	0.00	4.22	1.09	0.00	0.00	0.00	0.00	3.02	10.45
p-value$_{fm}$	(0.86)	(0.19)	(0.49)	(0.88)	(0.85)	(0.93)	(0.92)	(0.31)	(0.01)
LR_{jw}	0.00	3.37	1.04	0.00	0.00	0.00	0.00	2.30	6.82
p-value$_{jw}$	(0.86)	(0.26)	(0.50)	(0.88)	(0.85)	(0.92)	(0.92)	(0.42)	(0.07)
LR_{pm}	0.00	2.04	0.34	0.00	0.00	0.00	0.00	1.12	4.20
p-value$_{pm}$	(0.85)	(0.44)	(0.68)	(0.88)	(0.85)	(0.92)	(0.92)	(0.65)	(0.20)
Panel C: WLS									
	HL	P	CV	KLVN	FF3	C	PS	FFTD	FF5
LR_{fm}	3.37	16.82	14.52	0.00	3.40	0.00	16.71	28.67	10.42
p-value$_{fm}$	(0.13)	(0.00)	(0.00)	(0.88)	(0.13)	(0.93)	(0.00)	(0.00)	(0.01)
LR_s	1.75	4.38	4.22	0.00	3.20	0.00	5.96	6.90	5.05
p-value$_{fm}$	(0.30)	(0.17)	(0.12)	(0.88)	(0.14)	(0.93)	(0.06)	(0.06)	(0.14)
LR_{jw}	1.61	2.45	3.02	0.00	3.40	0.00	2.85	3.19	2.46
p-value$_{jw}$	(0.31)	(0.37)	(0.20)	(0.88)	(0.13)	(0.92)	(0.25)	(0.30)	(0.39)
LR_{pm}	1.11	0.96	5.22	0.00	2.74	0.00	1.98	7.40	2.06
p-value$_{pm}$	(0.40)	(0.62)	(0.08)	(0.88)	(0.18)	(0.92)	(0.36)	(0.05)	(0.45)

Bibliography

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz, 2016, Fragility of asymptotic agreement under Bayesian learning, *Theoretical Economics* 11, 187–225.

Acharya, Viral V, and Tanju Yorulmazer, 2008, Information contagion and bank herding, *Journal of Money, Credit and Banking* 40, 215–231.

Ahnert, Toni, and Ali Kakhbod, 2017, Information choice and amplification of financial crises, *The Review of Financial Studies* 30, 2130–2178.

Allen, Franklin, Ana Babus, and Elena Carletti, 2012, Asset commonality, debt maturity and systemic risk, *Journal of Financial Economics* 104, 519–534.

Allen, Franklin, and Douglas Gale, 2000, Financial contagion, *Journal of Political Economy* 108, 1–33.

Allen, Franklin, and Douglas Gale, 2004, Financial fragility, liquidity, and asset prices, *Journal of the European Economic Association* 2, 1015–1048.

Allen, Franklin, and Douglas Gale, 2009, *Understanding financial crises* (Oxford University Press).

Alvarez, Fernando E, Francesco Lippi, and Luigi Paciello, 2015, Monetary shocks in models with inattentive producers, *The Review of Economic Studies* 83, 421–459.

Andersen, Torben G, Tim Bollerslev, Francis X Diebold, and Clara Vega, 2007, Real-time price discovery in global stock, bond and foreign exchange markets, *Journal of International Economics* 73, 251–277.

Anderson, Evan W, Eric Ghysels, and Jennifer L Juergens, 2009, The impact of risk and uncertainty on expected returns, *Journal of Financial Economics* 94, 233–263.

Ang, Andrew, and Joseph Chen, 2002, Asymmetric correlations of equity portfolios, *Journal of Financial Economics* 63, 443–494.

Angeletos, George-Marios, and Alessandro Pavan, 2004, Transparency of information and coordination in economies with investment complementarities, *American Economic Review* 94, 91–98.

Antoniou, Constantinos, Richard DF Harris, and Ruogu Zhang, 2015, Ambiguity aversion and stock market participation: An empirical analysis, *Journal of Banking & Finance* 58, 57–70.

Barroso, Pedro, Martijn Boons, and Paul Karehnke, 2019, Time-varying state variable risk premia in an ICAPM, *Working Paper* .

Bekaert, Geert, Michael Ehrmann, Marcel Fratzscher, and Arnaud Mehl, 2014, The global crisis and equity market contagion, *The Journal of Finance* 69, 2597–2649.

Bekaert, Geert, Campbell R Harvey, and Angela Ng, 2005, Market integration and contagion, *Journal of Business* 78.

Ben-David, Itzhak, John R Graham, and Campbell R Harvey, 2013, Managerial miscalibration, *The Quarterly Journal of Economics* 128, 1547–1584.

Benoit, Sylvain, Jean-Edouard Colliard, Christophe Hurlin, and Christophe Pérignon, 2017, Where the risks lie: A survey on systemic risk, *Review of Finance* 21, 109–152.

Bollerslev, Tim, and Viktor Todorov, 2011, Tails, fears, and risk premia, *The Journal of Finance* 66, 2165–2211.

Boons, Martijn, 2016, State variables, macroeconomic activity, and the cross section of individual stocks, *Journal of Financial Economics* 119, 489–511.

Boyer, Brian H, Tomomi Kumagai, and Kathy Yuan, 2006, How do crises spread? Evidence from accessible and inaccessible stock indices, *The Journal of Finance* 61, 957–1003.

Caballero, Ricardo J, and Arvind Krishnamurthy, 2008, Collective risk management in a flight to quality episode, *The Journal of Finance* 63, 2195–2230.

Caballero, Ricardo J, and Alp Simsek, 2013, Fire sales in a model of complexity, *The Journal of Finance* 68, 2549–2587.

Calomiris, Charles W, Inessa Love, and Maria Soledad Martinez Peria, 2010, Crisis "shock factors" and the cross-section of global equity returns, Technical report, National Bureau of Economic Research.

Calomiris, Charles W, Inessa Love, and María Soledad Martínez Pería, 2012, Stock returns sensitivities to crisis shocks: Evidence from developed and emerging markets, *Journal of International Money and Finance* 31, 743–765.

Calvo, Guillermo A, 2004, *Contagion in emerging markets: when Wall Street is a carrier* (Springer).

Calvo, Guillermo A, and Enrique G Mendoza, 2000, Rational contagion and the globalization of securities markets, *Journal of International Economics* 51, 79–113.

Campbell, John Y, and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94, 1249–1275.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

Chen, Nai-Fu, Richard Roll, and Stephen A Ross, 1986, Economic forces and the stock market, *Journal of business* 383–403.

Chudik, Alexander, and Marcel Fratzscher, 2011a, Identifying the global transmission of the 2007–2009 financial crisis in a GVAR model, *European Economic Review* 55, 325–339.

Chudik, Alexander, and Marcel Fratzscher, 2011b, Identifying the global transmission of the 2007–2009 financial crisis in a gvar model, *European Economic Review* 55, 325–339.

Claessens, Stijn, 2014, An overview of macroprudential policy tools .

Claessens, Stijn, Rudiger Dornbusch, and Yung Chul Park, 2001, Contagion: Why crises spread and how this can be stopped, in *International financial contagion*, 19–41 (Springer).

Cochrane, John H, 2005, *Asset Pricing* (Princeton University Press).

Cochrane, John H, and Monika Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138–160.

Committee on Capital Markets Regulation, 2014, What to do about contagion?, `http://capmktsreg.org/reports/what-to-do-about-contagion/`.

Connolly, Robert A, and F Albert Wang, 2003, International equity market comovements: Economic fundamentals or contagion?, *Pacific-Basin Finance Journal* 11, 23–43.

Conrad, Jennifer, Bradford Cornell, and Wayne R Landsman, 2002, When is bad news really bad news?, *The Journal of Finance* 57, 2507–2532.

Cooper, Ilan, and Paulo Maio, 2018, Asset growth, profitability, and investment opportunities, *Management Science* .

Cukierman, Alex, and Yehuda Izhakian, 2015, Bailout uncertainty in a microfounded general equilibrium model of the financial system, *Journal of Banking & Finance* 52, 160–179.

Dasgupta, Amil, 2004, Financial contagion through capital connections: A model of the origin and spread of bank panics, *Journal of the European Economic Association* 2, 1049–1084.

Diamond, Douglas W, and Philip H Dybvig, 1983, Bank runs, deposit insurance, and liquidity, *The Journal of Political Economy* 401–419.

Dicks, David L, and Paolo Fulghieri, 2019, Uncertainty aversion and systemic risk, *Journal of Political Economy* 127, 000–000.

Didier, Tatiana, Paolo Mauro, and Sergio L Schmukler, 2008, Vanishing financial contagion?, *Journal of Policy Modeling* 30, 775–791.

Easley, David, and Maureen O'Hara, 2010, Liquidity and valuation in an uncertain world, *Journal of Financial Economics* 97, 1–11.

Eichengreen, Barry, Ashoka Mody, Milan Nedeljkovic, and Lucio Sarno, 2012, How the subprime crisis went global: evidence from bank credit default swap spreads, *Journal of International Money and Finance* 31, 1299–1318.

Eichengreen, Barry, Andrew K Rose, and Charles Wyplosz, 1996, Contagious currency crises, Technical report, National Bureau of Economic Research.

Ellsberg, Daniel, 1961, Risk, ambiguity, and the Savage axioms, *The quarterly journal of economics* 643–669.

Epstein, Larry G, and Martin Schneider, 2007, Learning under ambiguity, *The Review of Economic Studies* 74, 1275–1303.

Epstein, Larry G, and Martin Schneider, 2008, Ambiguity, information quality, and asset pricing, *The Journal of Finance* 63, 197–228.

Epstein, Larry G, and Martin Schneider, 2010, Ambiguity and asset markets, Technical report, National Bureau of Economic Research.

Fama, Eugene F., 1970, Multiperiod consumption-investment decisions, *American Economic Review* 60, 163–174.

Fama, Eugene F, 1991, Efficient capital markets: Ii, *The Journal of Finance* 46, 1575–1617.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of financial economics* 33, 3–56.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of financial economics* 116, 1–22.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Forbes, Kristin, 2012, The big C: identifying and mitigating contagion, in *Proceedings-Economic Policy Symposium-Jackson Hole*, 23–87, Federal Reserve Bank of Kansas City.

Forbes, Kristin, and Roberto Rigobon, 2001, Measuring contagion: conceptual and empirical issues, in *International financial contagion*, 43–66 (Springer).

Forbes, Kristin J, 2004, The asian flu and russian virus: the international transmission of crises in firm-level data, *Journal of International Economics* 63, 59–92.

Forbes, Kristin J, and Roberto Rigobon, 2002, No contagion, only interdependence: measuring stock market comovements, *The Journal of Finance* 57, 2223–2261.

Forbes, Kristin J, and Francis E Warnock, 2012, Capital flow waves: Surges, stops, flight, and retrenchment, *Journal of International Economics* 88, 235–251.

Fratzscher, Marcel, 2012a, Capital flows, push versus pull factors and the global financial crisis, *Journal of International Economics* 88, 341–356.

Fratzscher, Marcel, 2012b, Capital flows, push versus pull factors and the global financial crisis, *Journal of International Economics* 88, 341–356.

Freixas, Xavier, Bruno M Parigi, and Jean-Charles Rochet, 2000, Systemic risk, interbank relations, and liquidity provision by the central bank, *Journal of Money, Credit and Banking* 611–638.

Gibbons, Michael R, Stephen A Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Gilboa, Itzhak, and David Schmeidler, 1989, Maxmin expected utility with non-unique prior, *Journal of mathematical economics* 18, 141–153.

Glick, Reuven, and Andrew K Rose, 1999, Contagion and trade: Why are currency crises regional?, *Journal of International Money and Finance* 18, 603–617.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2013, Chi-squared tests for evaluation and comparison of asset pricing models, *Journal of Econometrics* 173, 108–125.

Guidolin, Massimo, and Francesca Rinaldi, 2010, A simple model of trading and pricing risky assets under ambiguity: any lessons for policy-makers?, *Applied Financial Economics* 20, 105–135.

Guidolin, Massimo, and Francesca Rinaldi, 2013, Ambiguity in asset pricing and portfolio choice: A review of the literature, *Theory and Decision* 74, 183–217.

Hahn, Jaehoon, and Hangyong Lee, 2006, Yield spreads as alternative risk factors for size and book-to-market, *Journal of Financial and Quantitative Analysis* 41, 245–269.

Ilut, Cosmin, and Rosen Valchev, 2017, Economic agents as imperfect problem solvers, *Working Paper* .

Iyer, Rajkamal, and Jose-Luis Peydro, 2011, Interbank contagion at work: Evidence from a natural experiment, *The Review of Financial Studies* 24, 1337–1377.

Jagannathan, Ravi, Keiichi Kubota, and Hitoshi Takehara, 1998, Relationship between labor-income risk and average return: Empirical evidence from the japanese stock market, *The Journal of Business* 71, 319–347.

Jagannathan, Ravi, and Yong Wang, 2007, Lazy investors, discretionary consumption, and the cross-section of stock returns, *The Journal of Finance* 62, 1623–1661.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional capm and the cross-section of expected returns, *The Journal of finance* 51, 3–53.

Jagannathan, Ravi, and Zhenyu Wang, 1998, An asymptotic theory for estimating beta-pricing models using cross-sectional regression, *The Journal of Finance* 53, 1285–1309.

Jotikasthira, Chotibhak, Christian Lundblad, and Tarun Ramadorai, 2012, Asset fire sales and purchases and the international transmission of funding shocks, *The Journal of Finance* 67, 2015–2050.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2016, A rational theory of mutual funds' attention allocation, *Econometrica* 84, 571–626.

Kaminsky, Graciela L, Carmen M Reinhart, and Carlos A Vegh, 2003, The unholy trinity of financial contagion, *Journal of Economic Perspectives* 17, 51–74.

Kan, Raymond, and Cesare Robotti, 2011, On the estimation of asset pricing models using univariate betas, *Economics Letters* 110, 117–121.

Kan, Raymond, Cesare Robotti, and Jay Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *The Journal of Finance* 68, 2617–2649.

Kan, Raymond, and Chu Zhang, 1999, Two-pass tests of asset pricing models with useless factors, *the Journal of Finance* 54, 203–235.

Kandel, Eugene, and Neil D Pearson, 1995, Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy* 831–872.

Kandel, Shmuel, and Robert F Stambaugh, 1995, Portfolio inefficiency and the cross-section of expected returns, *The Journal of Finance* 50, 157–184.

Kannan, Prakash, and Friederike Fritzi N Köhler-Geib, 2009, The uncertainty channel of contagion .

Karolyi, G Andrew, 2003, Does international financial contagion really exist?, *International Finance* 6, 179–199.

Kasperson, Roger E, Ortwin Renn, Paul Slovic, Halina S Brown, Jacque Emel, Robert Goble, Jeanne X Kasperson, and Samuel Ratick, 1988, The social amplification of risk: A conceptual framework, *Risk Analysis* 8, 177–187.

Kearney, Colm, and Sha Liu, 2014, Textual sentiment in finance: A survey of methods and models, *International Review of Financial Analysis* 33, 171–185.

Kim, Oliver, and Robert E. Verrecchia, 1994, Market liquidity and volume around earnings announcements, *Journal of Accounting and Economics* 17, 41 – 67.

King, Mervyn A., and Sushil Wadhwani, 1990, Transmission of volatility between stock markets, *Review of Financial Studies* 3, 5–33.

Kodres, Laura E, and Matthew Pritsker, 2002, A rational expectations model of financial contagion, *The Journal of Finance* 57, 769–799.

Koijen, Ralph SJ, Hanno Lustig, and Stijn Van Nieuwerburgh, 2017, The cross-section and time series of stock and bond returns, *Journal of Monetary Economics* 88, 50–69.

Krishnamurthy, Arvind, 2010, Amplification mechanisms in liquidity crises, *American Economic Journal: Macroeconomics* 2, 1–30.

Kuhnen, Camelia M, 2015, Asymmetric learning from financial information, *The Journal of Finance* 70, 2029–2062.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175–194.

Li, C Wei, Ashish Tiwari, and Lin Tong, 2016, Investment decisions under ambiguity: Evidence from mutual fund investor behavior, *Management Science* .

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *The Review of Economics and Statistics* 47, 13–37.

Luo, Yulei, 2008, Consumption dynamics under information processing constraints, *Review of Economic Dynamics* 11, 366–385.

Lutzenberger, Fabian T, 2015, Multifactor models and their consistency with the ICAPM: Evidence from the european stock market, *European Financial Management* 21, 1014–1052.

Mackowiak, Bartosz, and Mirko Wiederholt, 2009, Optimal sticky prices under rational inattention, *American Economic Review* 99, 769–803.

Maćkowiak, Bartosz, and Mirko Wiederholt, 2015, Inattention to rare events, Technical report, ECB Working Paper.

Maio, Paulo, and Pedro Santa-Clara, 2012, Multifactor models and their consistency with the ICAPM, *Journal of Financial Economics* 106, 586–613.

Manconi, Alberto, Massimo Massa, and Ayako Yasuda, 2012, The role of institutional investors in propagating the crisis of 2007–2008, *Journal of Financial Economics* 104, 491–518.

Matějka, Filip, 2015, Rationally inattentive seller: Sales and discrete pricing, *The Review of Economic Studies* 83, 1125–1155.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Mondria, Jordi, 2010, Portfolio choice, attention allocation, and price comovement, *Journal of Economic Theory* 145, 1837.

Mondria, Jordi, and Climent Quintana-Domeque, 2013, Financial contagion and attention allocation, *The Economic Journal* 123, 429–454.

Mossin, Jan, 1966, Equilibrimn in a capital asset market, *Econometrica* 34, 768–783.

Newey, Whitney K, and Kenneth D West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Newey, Whitney K, and Kenneth D West, 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* 61, 631–653.

Nimark, Kristoffer, and Savitar Sundaresan, 2018, Inattention and belief polarization, *Working Paper* .

Oh, Frederick Dongchuhl, 2013, Contagion of a liquidity crisis between two firms, *Journal of Financial Economics* 107, 386–400.

Ozsoy, Sati Mehmet, 2013, *Asymmetric Correlations in Financial Markets*, Ph.D. thesis, Duke University.

Paciello, Luigi, and Mirko Wiederholt, 2013, Exogenous information, endogenous information, and optimal monetary policy, *Review of Economic Studies* 81, 356–388.

Pástor, L'uboš, and Robert F Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Pavlova, Anna, and Roberto Rigobon, 2008, The role of portfolio constraints in the international propagation of shocks, *The Review of Economic Studies* 75, 1215–1256.

Pericoli, Marcello, and Massimo Sbracia, 2003, A primer on financial contagion, *Journal of Economic Surveys* 17, 571–608.

Petkova, Ralitsa, 2006, Do the Fama–French factors proxy for innovations in predictive variables?, *The Journal of Finance* 61, 581–612.

Pritsker, Matthew, 2013, Knightian uncertainty and interbank lending, *Journal of Financial Intermediation* 22, 85–105.

Renn, Ortwin, William J Burns, Jeanne X Kasperson, Roger E Kasperson, and Paul Slovic, 1992, The social amplification of risk: Theoretical foundations and empirical applications, *Journal of Social Issues* 48, 137–160.

Rigobon, Roberto, and Shang-Jin Wei, 2003, News, contagion, and anticipation understanding the differences between the Asian and the Latin American crises, *LAEBAWorking Paper Series* .

Rose, Andrew K, and Mark M Spiegel, 2010, Cross-country causes and consequences of the 2008 crisis: International Linkages and American Exposure, *Pacific Economic Review* 15, 340–363.

Rose, Andrew K, and Mark M Spiegel, 2011, Cross-country causes and consequences of the crisis: An update, *European Economic Review* 55, 309–324.

Routledge, Bryan R, and Stanley E Zin, 2009, Model uncertainty and liquidity, *Review of Economic dynamics* 12, 543–566.

Savage, Leonard, 1972, *The Foundations of Statistics (2nd revised edition)* (Dover Publications, New York).

Shanken, Jay, 1985, Multivariate tests of the zero-beta capm, *Journal of Financial Economics* 14, 327–348.

Shanken, Jay, 1992, On the estimation of beta-pricing models, *The Review of Financial Studies* 5, 1–33.

Sharpe, William F, 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *The Journal of Finance* 19, 425–442.

Sims, Christopher A, 2003, Implications of rational inattention, *Journal of Monetary Economics* 50, 665–690.

Sixteenth Jacques Polak Annual Research Conference, 2015, Unconventional monetary and exchange rate policies, `http://www.finanssivalvonta.fi/en/Publications/Annual_reports/Documents/FIN-FSA_Annual_Report_2009.pdf`.

Slovic, Paul, 1987, Perception of risk, *Science* 236, 280–285.

Stevens, Luminita, 2015, Coarse pricing policies, Technical report, Federal Reserve Bank of Minneapolis.

Sundaresan, Savitar, 2018, Emergency preparedness: Rare events and the persistence of uncertainty, *Working Paper* .

Trautmann, Stefan T, and Gijs Van De Kuilen, 2015, Ambiguity attitudes, in Gideon Keren, and George Wu, eds., *The Wiley Blackwell Handbook of Judgment and Decision Making*, chapter 3, 89–116 (John Wiley & Sons, Ltd).

Treynor, Jack L., 1961, Toward a theory of market value of risky assets, *Unpublished manuscript* .

Tutino, Antonella, 2013, Rationally inattentive consumption choices, *Review of Economic Dynamics* 16, 421–439.

Uhlig, Harald, 2010, A model of a systemic bank run, *Journal of Monetary Economics* 57, 78–96.

Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2010, Information acquisition and under-diversification, *The Review of Economic Studies* 77, 779–805.

Williams, Christopher D, 2014, Asymmetric responses to earnings news: A case for ambiguity, *The Accounting Review* 90, 785–817.

Wolak, Frank A, 1987, An exact test for multiple inequality and equality constraints in the linear regression model, *Journal of the American Statistical Association* 82, 782–793.

Wolak, Frank A, 1989, Testing inequality constraints in linear econometric models, *Journal of Econometrics* 41, 205–235.

Woodford, Michael, 2001, Imperfect common knowledge and the effects of monetary policy, Technical report, National Bureau of Economic Research.

Woodford, Michael, 2009, Information-constrained state-dependent pricing, *Journal of Monetary Economics* 56, S100–S124.

Yuan, Kathy, 2005, Asymmetric price movements and borrowing constraints: A rational expectations equilibrium model of crises, contagion, and confusion, *The Journal of Finance* 60, 379–411.

Zhou, John, 2015, The good, the bad, and the ambiguous: The aggregate stock market dynamics around macroeconomic news, *Working Paper* .