



Estimating the Number of Latent Factors: A Comparative Analysis

Marco Avarucci , Valentina Raponi & Paolo Zaffaroni

To cite this article: Marco Avarucci , Valentina Raponi & Paolo Zaffaroni (13 May 2026): Estimating the Number of Latent Factors: A Comparative Analysis, *Econometric Reviews*, DOI: [10.1080/07474938.2026.2648946](https://doi.org/10.1080/07474938.2026.2648946)

To link to this article: <https://doi.org/10.1080/07474938.2026.2648946>



© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 13 May 2026.



Submit your article to this journal [↗](#)



Article views: 186



View related articles [↗](#)



View Crossmark data [↗](#)

Estimating the Number of Latent Factors: A Comparative Analysis

Marco Avarucci^a, Valentina Raponi^b, and Paolo Zaffaroni^c

^aAdam Smith Business School, University of Glasgow, Glasgow, UK; ^bIESE Business School, University of Navarra, Barcelona, Spain; ^cImperial College Business School, London, UK and University of Rome La Sapienza, Rome, Italy

ABSTRACT

This paper evaluates a set of widely used methodologies for determining the number of latent factors in large-dimensional factor models. Its contribution is a comprehensive and systematic comparison of their performance. We assess these estimators not only under the data-generating processes for which they were originally designed, but also across a broader set of environments. Our analysis encompasses static, dynamic, and generalized dynamic factor models, considering factor strength that ranges from strong to semi-strong and semi-weak. Our results show that with strong factors, most estimators across all three classes deliver near-perfect identification when both the cross-section n and time dimension T are large, providing practitioners with a wide set of reliable choices. As factor strength weakens, performance diverges: only a few estimators remain comparatively robust, while other estimators tend to underestimate the true number of factors or shocks, particularly when the idiosyncratic components are not i.i.d. Overall, no single estimator dominates across all settings. Our findings provide practical guidance for applied work and highlight the advantages and limitations of existing methodologies.

ARTICLE HISTORY

Received 25 July 2024
Accepted 11 March 2026

KEYWORDS

Factor models; Generalized dynamic factor models; Estimator of the number of factors; Frequency bands; Weak factors

1. Introduction

Factor models have become the predominant methodology for estimating large-dimensional datasets, particularly when the time- and cross-sectional dimensions are large. These models assume that observed variables can be decomposed into two unobserved, orthogonal components: a *common* component, capturing comovements, and an *idiosyncratic* component, reflecting noise or individual features. In *static* factor models (SFM), the common component is expressed as a linear combination of r latent factors, which only have a contemporaneous effect on the observed data. The common component may also follow a dynamic process, where $q \leq r$ *dynamic factors* (or *common shocks*) influence the data through their lags. This gives rise to the *generalized dynamic factor model* (GDFM) when modeled as a linear process, or to the *dynamic factor model* (DFM) when restrictions are imposed on the filter.

While the distributional theory for estimators of factor models is well established (see Bai, 2003; Barigozzi et al. 2024; Forni et al. 2000, among others), it typically relies on the unrealistic assumption that the number of latent factors is *known* to the practitioner. Therefore, regardless of the chosen model, determining r or q is a crucial first step for the effective use of large-dimensional factor models.¹

CONTACT Paolo Zaffaroni  pz@ic.ac.uk  Finance, Imperial College Business School and University of Rome La Sapienza, UK
 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07474938.2026.2648946>.

¹This issue is central for empirical asset pricing, including portfolio construction and risk management (see, e.g., Giglio and Xiu 2021; Kozak, Nagel, and Santosh 2018; Lettau and Pelger 2020b). In macroeconomics, it is equally important for identifying the shocks driving business cycle fluctuations (see Angeletos, Collard, and Dellas, 2020; Bai and Ng 2007; Forni et al. 2023; Giannone, Reichlin, and Sala, 2004; Kydland and Prescott 1982; Sargent et al. 1977; among many others).

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Starting from the seminal contribution of Bai and Ng (2002), who introduced an information-criteria estimator for the number of static factors, a large literature has followed. Alessi, Barigozzi, and Capasso (2010) revised the procedure to improve its performance, while Amengual and Watson (2007) and Bai and Ng (2007) extended the approach to dynamic factors. Hallin and Liška (2007) adapted it to the GDFM using a frequency domain approach. Another approach builds on the fact that, as the cross-sectional dimension (n) grows, the largest r eigenvalues of the covariance matrix in static models (or the q eigenvalues of the spectral density matrix in GDFMs) diverge, whereas the others remain bounded. For the static case, see Onatski (2010) and Ahn and Horenstein (2013), while Avarucci et al. (2026) extend the approach to the GDFM. More recently, Barigozzi and Farnè (2024) developed a frequency-domain estimator of q based on a regularized smoothed periodogram. This list is not exhaustive: additional contributions include Kapetanios (2010), Lam, Yao, and Bathia (2011), Trapani (2018), and Caro and Peña (2025), among others.

Most of the estimators mentioned above have been evaluated under data-generating processes (DGPs) consistent with the assumed model, but in practice, assessing the suitability of any DGP is challenging. It is therefore crucial to evaluate their performance in finite samples and under alternative assumptions. The main objective and contribution of this paper is to determine how accurately existing methods perform across a variety of DGPs, not only in the settings for which they were originally designed, thereby providing practical guidance for empirical work.

Moreover, most of the existing methodologies assume *strong* factors, where the r (or q) eigenvalues of the covariance (or spectral density) matrix diverge linearly in n . However, recent works (e.g., Anatolyev and Mikusheva, 2022; Lettau and Pelger, 2020a; Pesaran and Smith, 2021) consider weaker structures, where eigenvalues diverge more slowly (*semi-strong* or *semi-weak* factors) or may not diverge at all (*weak* factors). Such cases are empirically relevant, especially in finance and macroeconomics: weak or semi-strong factors affect risk premia estimation, investment strategies, and risk management, as well as the accuracy of macroeconomic forecasts and policy design. To address this issue, Freyaldenhoven (2022) proposed estimators for the number of static factors that exploit not only the eigenvalues of the covariance matrix, but also the information contained in the associated eigenvectors. Empirical evidence further shows that different estimators can yield different numbers of factors on the same dataset, partly due to their varying sensitivity to weak factors. Understanding how these estimators perform in the presence of weaker factors is therefore essential for robust modeling and reliable inference.

Our study delivers several clear findings. With strong factors, most estimators (whether designed for SFMs, DFMs, or GDFMs) achieve very high accuracy, often approaching perfection when both n and T are large. As factor strength weakens, performance diverges across classes: some estimators remain comparatively robust while others deteriorate. Several methods designed for DFMs tend to over-select due to correlation in the idiosyncratic component and within the GDFM class, a subset retains good accuracy even under weak factors, particularly when T is large. Reliability improves as both the cross-sectional and time dimensions grow, though in the presence of more elevated idiosyncratic dependence, some methods require n and T to be very large. Overall, no single estimator dominates across all settings. While certain methods designed for SFMs and GDFMs emerge as the most reliable under weaker factor structures, they also rest on more restrictive assumptions and involve user-chosen parameters, whose sensitivity has not been fully assessed. Our results provide practical guidance for applied work: (i) Methods designed for SFMs and GDFMs are often preferable under strong factors; (ii) Only a few methods remain reliable when factors are weak; and (iii) Several estimators (especially those designed for DFMs) should be applied with caution unless idiosyncratic auto- and cross-sectional dependence is negligible.

The paper is organized as follows. Section 2 introduces SFMs, DFMs, and GDFMs, together with their assumptions. Section 3 presents the estimators for each case: static (Section 3.2), dynamic (Section 3.3), and generalized dynamic factor models (Section 3.4). Section 4 describes the simulation designs, while Section 5 reports the main results for the three experiments, namely the static experiment (Section 5.1), the dynamic experiment (Section 5.2), and the GDFM experiment (Section 5.3). Section 6 concludes. Additional material is provided in Online Appendices A and B.

Throughout the paper, we are going to use the following notation. For a generic matrix A , we denote by $\|A\|$ its Frobenius norm and by A' its transpose. If A is positive semi-definite, then $\psi_k(A)$ is its k th largest eigenvalue. We use L to indicate the lag operator and $c(L) = \sum_{j=1}^{\infty} c_j L^j$, for the one-sided infinite lag (real) polynomial. The $p \times q$ matrix lag polynomial $C(L)$ has entries $c_{ij,k}(L)$, for $i = 1, \dots, p$ and $j = 1, \dots, q$. Finally, $z_{it} \sim \mathcal{N}(0, 1)$ and $z_{it} \sim \mathcal{U}[a, b]$, mean that the sequence of i.i.d. random variables z_{it} has either a standard normal distribution or a uniform distribution with support in $[a, b]$, respectively. C denotes a positive bounded constant whose value may change from one occurrence to another. $\mathbb{E}(\cdot)$ denotes the expectation operator.

2. The static and the (generalized) dynamic factor models

The idea of factor structures in time series dates back to Chamberlain (1983) and Chamberlain and Rothschild (1983), who postulated that an n -dimensional time series $X_t = (X_{1t}, \dots, X_{nt})'$ can be decomposed into a linear function of an r -dimensional vector of latent factors $F_t = (F_{1t}, \dots, F_{rt})'$ and an idiosyncratic component $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{nt})'$, i.e.

$$X_{it} = \chi_{it}^{\text{stat}} + \epsilon_{it} = b_i' F_t + \epsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (2.1)$$

where $b_i = (b_{i1}, \dots, b_{ir})'$ is the r -vector of unknown factor loadings, and $\chi_t^{\text{stat}} = (\chi_{1t}^{\text{stat}}, \dots, \chi_{nt}^{\text{stat}})'$ denotes the n -dimensional static common component. The number of factors $r \geq 0$ is assumed to be fixed and much smaller than n . Because (21) restricts F_t to contemporaneous effects on X_{it} , it is termed a static factor model (SFM), with F_t the associated static factor.²

Despite its name, a static factor may itself follow a dynamic process (for example, an r -dimensional stationary VAR). Since the idiosyncratic component $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{nt})'$ is allowed to exhibit mild serial and cross-sectional correlation, model (21) is referred to as an *approximate* factor model, in contrast to an *exact* factor model where ϵ_t has a diagonal covariance matrix. The identification of an approximate factor model requires an asymptotic framework in which both n and T diverge. Hence, the consistency of all the estimators requires $n, T \rightarrow \infty$.

In all the papers reviewed here, factor models are statistical models defined by specific sets of assumptions. For the static model, we list the assumptions in Bai and Ng (2002) as a reference, using the prefix (S) to indicate the static case.

- (S.1) $\mathbb{E} \|F_t\|^4 < \infty$ and $\text{plim}_{T \rightarrow \infty} T^{-1} F' F$ is positive definite;
- (S.2) $\sup_i \|b_i\| < \infty$ and $\|n^{-1} B' B - D\| \rightarrow 0$ as $n \rightarrow \infty$ for some $(r \times r)$ positive definite matrix D where $B = (b_1, \dots, b_n)'$ is a nonrandom matrix of dimension $(n \times r)$;
- (S.3) (i) $\mathbb{E} |\epsilon_{it}|^8 \leq C$; (ii) $\mathbb{E} (\epsilon_s' \epsilon_t / n) = \gamma_n(s, t)$, $|\gamma_n(s, s)| \leq C$ for all s , and $T^{-1} \sum_{s=1}^T \sum_{t=1}^T |\gamma_n(s, t)| \leq C$; (iii) $\mathbb{E} (\epsilon_{it} \epsilon_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t , in addition $n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\tau_{ij}| \leq C$; (iv) $\mathbb{E} (\epsilon_{it} \epsilon_{js}) = \tau_{ij,ts}$ and $(nT)^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^T \sum_{t=1}^T |\tau_{ij,ts}| \leq C$; (v) for every (t, s) , $\mathbb{E} |n^{-1/2} \sum_{i=1}^n [\epsilon_{is} \epsilon_{it} - \mathbb{E} (\epsilon_{is} \epsilon_{it})]|^4 \leq C$;
- (S.4) $\mathbb{E} (n^{-1} \sum_{i=1}^n \|T^{-1/2} F' \epsilon_i\|) \leq C$.

Assumptions (S.1) and (S.2) imply that the factors are strong, or pervasive (only a fixed number of units, non-increasing with n , is not affected by the factors), whereas Assumption (S.3) entails that the eigenvalues of the covariance matrix of ϵ_t remain bounded when $n \rightarrow \infty$. Assumption (S.4) imposes weak dependence between the factors and the idiosyncratic component.

To allow factors to affect X_t through lags as well as contemporaneously, Geweke (1977) and Sargent et al. (1977) introduced *dynamic loadings*, replacing $b_i' F_t$ with linear processes $b_i'(L) u_t = b_{i1}(L) u_{1t} + \dots + b_{iq}(L) u_{qt}$, where q is a positive integer and $u_t = (u_{1t}, \dots, u_{qt})'$ is a q -dimensional white-noise vector process. The *exact* specification, with i.i.d. and mutually orthogonal idiosyncratic components,

²Throughout this section, we assume that the n -dimensional time series $\{X_t : t = 1, \dots, T\}$ is a finite- (n, T) realization of the double sequence $\{X_{it} : i \in \mathbb{N}, t \in \mathbb{Z}\}$, which, as is standard in the literature and without loss of generality, is second-order stationary with $\mathbb{E}(X_{it}) = 0$ and $0 < \mathbb{E}(X_{it}^2) < \infty$ for all $i = 1, \dots, n$ and $t = 1, \dots, T$. The same assumptions hold for the factors F_t and the idiosyncratic component ϵ_t .

was later generalized to the Generalized Dynamic Factor Model (GDFM) by Forni et al. (2000) and Forni and Lippi (2001):

$$X_{it} = \chi_{it}^{\text{gdyn}} + \epsilon_{it} = b'_i(L)u_t + \epsilon_{it} \quad (2.2)$$

where $\chi_{nt}^{\text{gdyn}} = (\chi_{1t}^{\text{gdyn}}, \dots, \chi_{nt}^{\text{gdyn}})' = b'_i(L)u_t$ denotes the n -dimensional (*generalized dynamic*) *common component*, with $b_{ij}(L)$, being an infinite one-sided square summable polynomial, for $j = 1, \dots, q$. The GDFM is typically accompanied by the following assumptions, denoted by the prefix (D) to indicate the dynamic specification:

- (D.1) χ_t^{gdyn} admits an $n \times n$ spectral density matrix $\Sigma_n^X(\theta)$, whose q -largest eigenvalue $\lambda_{qn}^X(\theta)$ diverges for all $\theta \in [-\pi, \pi]$ as $n \rightarrow \infty$, and $\liminf_{n \rightarrow \infty} n^{-1} \lambda_{qn}^X(\theta) > 0$.
- (D.2) $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{nt})'$ has an $n \times n$ spectral density matrix $\Sigma_n^\epsilon(\theta)$, whose largest eigenvalue $\lambda_{1n}^\epsilon(\theta)$ is bounded for all $\theta \in [-\pi, \pi]$, for any $n \in \mathbb{N}$.
- (D.3) $\mathbb{E}(\epsilon_{it}, u_{jt-k}) = 0$, for any $i \in \mathbb{N}, j = 1, \dots, q$ and $k \in \mathbb{Z}$.

Assumption (D.1) entails that the common shocks are pervasive, and (D.2) and (D.3) have a similar interpretation to (S.3) and (S.4), respectively.

Restricted versions of the GDFM have been considered in the literature. For example, Forni et al. (2005) assume that

$$X_{it} = \chi_{it}^{\text{dyn}} + \epsilon_{it} = C(L)f_t + \epsilon_{it} \quad (2.3)$$

where $\chi_{nt}^{\text{dyn}} = (\chi_{1t}^{\text{dyn}}, \dots, \chi_{nt}^{\text{dyn}})' = C(L)f_t$ denotes the n -dimensional *dynamic common component*, with $C(L)$ being such that $B(L) = C(L)A^{-1}(L)$, and where the factors f_t follow a vector autoregressive scheme of order h of the form $A(L)f_t = u_t$, with all solutions of $\det[A(z)] = 0$ lying outside the unit ball. The orders of the $c_{ij}(L)$, $j = 1, \dots, q$, have a finite maximum $s \geq 0$, i.e., $C(L) = C_0 + C_1L + \dots + C_sL^s$ in a finite $n \times q$ matrix polynomial. The entries of the q -dimensional vector $f_t = (f_{1t}, \dots, f_{qt})'$ are referred to as *dynamic factors*, and the specification in (23) is termed a *dynamic factor model* (DFM).

When model (23) holds, then X_t admits a static representation with $r = (s+1)q$ *static factors* $F_t = [f'_t, \dots, f'_{t-s}]'$, which follow a (singular) VAR(p) representation with $p = \max(1, h-s)$, i.e.,

$$F_t = \sum_{i=1}^p \Phi_i F_{t-i} + \xi_t, \quad \xi_t = R u_t, \quad (2.4)$$

where R is an $r \times q$ matrix of rank q . Similar models are considered in Stock and Watson (2002a, 2002b, 2005) and Bai and Ng (2007). See Appendix B1.1 for an illustrative example and further discussion.

3. Estimators

In this section, we present the estimators used to determine the number of factors, static or dynamic, depending on the model considered (SFM, DFM, or GDFM). For each DGP, the estimators are introduced in chronological order, highlighting their main deviations from the assumptions discussed in Section 2. Their acronyms are in square brackets. We begin with some preliminary definitions needed for the subsequent analysis.

3.1. Preliminary definitions

The factor model in (21) can be written compactly as $X = BF' + \epsilon$, where $F = (F_1, \dots, F_T)'$ is a $T \times r$ matrix of factors and $X = (X_1, \dots, X_T)$ is an $n \times T$ matrix of observed data. We denote by $\hat{B}^{(k)}$ and $\hat{F}^{(k)}$ the estimators of the loadings and factors, respectively, obtained *via* asymptotic principal components

(PCs), with the superscript k indicating the number of PCs used. The estimator $\hat{\mathbf{B}}^{(k)}$ can be constructed as $\sqrt{n}$ times the eigenvectors corresponding to the k largest eigenvalues of the $n \times n$ matrix $\mathbf{X}\mathbf{X}'$.³ We define $\hat{\mu}_k = \psi_k(\mathbf{X}'\mathbf{X}/T)$ as the k -th largest eigenvalue of the empirical covariance matrix of $\mathbf{X}_t$, omitting for notational convenience the dependence on T and n . Adopting the normalization $n^{-1}\hat{\mathbf{B}}^{(k)'}\hat{\mathbf{B}}^{(k)} = \mathbf{I}_r$ implies that $\hat{\mathbf{F}}^{(k)} = n^{-1}\mathbf{X}'\hat{\mathbf{B}}^{(k)}$. We define the periodogram-smoothing estimator of the spectral density matrix of $\mathbf{X}_t$ as

$$\hat{\Sigma}(\theta_j) = \frac{1}{2m_T + 1} \sum_{k=-m_T}^{m_T} \mathcal{I}(\theta_{j+k}), \quad (3.1)$$

where the $n \times n$ matrix $\mathcal{I}(\omega_j) = (2\pi T)^{-1} \left[\sum_{t=1}^T \mathbf{X}_t e^{-i\omega_j t} \right] \left[\sum_{t=1}^T \mathbf{X}_t' e^{i\omega_j t} \right]$ represents the periodogram of $\mathbf{X}_t$ at the Fourier frequencies $-\pi < \theta_j := 2\pi j/T \leq \pi$, with i (in Roman font) denoting the imaginary unit, and $m_T > 0$ is a bandwidth parameter. Let $\hat{\mathbf{\Gamma}}(h)$ denote the sample covariance matrix of $\mathbf{X}_t$ and $\mathbf{X}_{t-h}$ based on T observations. The lag-window estimator $\tilde{\Sigma}_n(\theta_j)$ is then defined as:

$$\tilde{\Sigma}(\theta_j^*) = \frac{1}{2\pi} \sum_{h=-m_T}^{m_T} \mathcal{K}\left(\frac{h}{m_T}\right) \hat{\mathbf{\Gamma}}(h) e^{-i\theta_j^* h} \quad (3.2)$$

where $\mathcal{K}(\cdot)$ is a positive even weight function and $\theta_j^* := 2\pi j/(2m_T + 1)$ for $j = -m_T, \dots, m_T$.⁴ We define $\hat{\lambda}_k(\theta_j) = \psi_k(\hat{\Sigma}(\theta_j))$ and $\tilde{\lambda}_k(\theta_j^*) = \psi_k(\tilde{\Sigma}(\theta_j^*))$.

Finally, we denote by $k_{\max}$ the pre-specified maximum value of r and q specified for each estimator.

3.2. Estimators for static factor models

For the SFM (21), we consider the following estimators of the number of factors.

- (1) $[\mathbf{IC}_{p_2}]$ (Bai and Ng (2002)). In their paper, Bai and Ng propose an information-criterion estimator for r , obtained by minimizing a loss function of the form $IC(k) = V(k) + kg(n, T)$, where $V(k) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left(X_{it} - \hat{\mathbf{B}}_i^{(k)} \hat{\mathbf{F}}_t^{(k)} \right)^2 = \frac{1}{n} \sum_{j=k+1}^n \hat{\mu}_j$ is the average of squared residuals when k factors are estimated, and $g(n, T)$ is a penalty term that increases with k to control for overfitting. The authors proposed different versions of this criterion. Here we focus on IC_{p_2} , which is perhaps the most commonly used in applied work:⁵

$$\hat{r}_{\mathbf{IC}_{p_2}} = \arg \min_{0 \leq r \leq k_{\max}} \left[\ln(V(k)) + k \left(\frac{n+T}{nT} \right) \ln(\min(n, T)) \right] \quad (3.3)$$

- (2) $[\mathbf{ABC}]$ (Alessi, Barigozzi, and Capasso (2010)). To improve the small-sample properties of the estimator $\mathbf{IC}_{p_2}$, Alessi, Barigozzi, and Capasso (2010) propose the following modified version:

$$\hat{r}_{\mathbf{ABC}}^{(c)} = \arg \min_{0 \leq k \leq k_{\max}} \left[\ln(V(k)) + ck \left(\frac{n+T}{nT} \right) \ln(\min(n, T)) \right], \quad (3.4)$$

where c is a positive tuning parameter introduced to improve finite-sample performance. The resulting estimate of the number of factors, therefore, depends on the choice of c , and in finite samples, different values of this free parameter may yield arbitrarily different estimates of r . Hallin and Liška (2007) propose an effective procedure to select c , denoted by $\hat{c}$. In what follows, we set $\hat{r}_{\mathbf{ABC}} = \hat{r}_{\mathbf{ABC}}^{(\hat{c})}$. Further details are provided in [Appendix B1.2](#).

³ An analogous estimator, computationally convenient when $T < n$, can be obtained from the $T \times T$ matrix $\mathbf{X}'\mathbf{X}$, by interchanging the roles of the factors and the loadings.

⁴ With a slight abuse of notation, we use m_T to denote the bandwidth parameter in (31) and (32), even though its value need not be the same in both cases.

⁵ It is also employed as a first-step estimator in Bai and Ng (2007).

- (3) [ED] (Onatski, 2010). Based on the intuition that for any $k > r$ the difference $\hat{\mu}_k - \hat{\mu}_{k+1}$ remains bounded, while $\hat{\mu}_r - \hat{\mu}_{r+1}$ diverges, Onatski (2010) proposes the class of estimators

$$\hat{r}_{\text{ED}}^{(\delta)} \equiv \max \{k \leq k_{\max} : \hat{\mu}_k - \hat{\mu}_{k+1} \geq \delta\}, \quad (3.5)$$

where $\delta > 0$ is a threshold parameter. The result holds under the assumption that $n/T \rightarrow y \in (0, \infty)$ as $n, T \rightarrow \infty$. The calibration of δ is carried out through an iterative algorithm that exploits the square-root shape at the edge of the eigenvalue distribution, hence the acronym ED (Edge Distribution); see Appendix B1.3 for details. Notice that the author relaxes Assumptions (S.1)–(S.2) by requiring only that $\psi_r[(B'B)(F'F/T)] \rightarrow \infty$ in probability. Since no separate restrictions are imposed on $B'B$ or $F'F/T$, factors need not be strong (see Section 4.3) or stationary. On the other hand, Assumption (S.3) is strengthened by imposing that $\epsilon = PVQ$, where P ($n \times n$) and Q ($T \times T$) are largely unrestricted deterministic matrices,⁶ and V is an $n \times T$ matrix with i.i.d. Gaussian entries.

- (4) [ER], (Ahn and Horenstein, 2013). In their paper, Ahn and Horenstein (2013) introduce the eigenvalue ratio (ER) estimator, which relies on the property that when the model has an r -dimensional factor structure, the r largest eigenvalues of the covariance matrix of X diverge with n , whereas the remaining ones remain bounded.⁷ The estimator is defined as

$$\hat{r}_{\text{ER}} = \arg \max_{1 \leq k \leq k_{\max}} \frac{\hat{\mu}_k}{\hat{\mu}_{k+1}}. \quad (3.6)$$

As in the ED case, it is assumed that $n/T \rightarrow y \in (0, \infty)$ as $n, T \rightarrow \infty$, with Assumption (S.3) replaced by the condition $\epsilon = PVQ$, where the entries of V are i.i.d., though not necessarily Gaussian.

- (5) [FR] (Freyaldenhoven, 2022). The estimators discussed above rely on the separation between large eigenvalues, reflecting factor-driven variation, and small eigenvalues, reflecting idiosyncratic noise. When factors are weak and B is sparse, this separation becomes less clear. To address this, Freyaldenhoven (2022) incorporates information from the eigenvectors of $X'X/T$ in addition to the eigenvalues, thereby restoring the eigenvalue gap. Specifically, let $\hat{B}^{(k)}$ denote the PC estimator of the loading matrix with entries $\hat{b}_{ik}$. For $\eta \in [0, 2]$, define

$$\hat{\Upsilon}_k^\eta = \psi_k \left(\frac{X'X}{T} \right) \hat{S}_k, \quad \text{where} \quad \hat{S}_k = \left(\frac{1}{z} \sum_{i=1}^z \frac{\hat{b}_{ik}^2}{\sqrt{n^{-1} \sum_{i=1}^n \hat{b}_{ik}^2}} \right)^\eta.$$

For $\eta > 0$, $\hat{S}_k$ measures the concentration of the k -th eigenvector on its z largest entries, with threshold $z = n^\tau g(n)$, $\tau \in [0, 1]$, $g(n) \rightarrow \infty$, and $g(n)/n^\beta \rightarrow 0$ for any $\beta > 0$. The parameter η governs the relative weight on the eigenvalue versus the eigenvector. Classical estimators such as IC_{p_2} , ED and ER can be defined in term of $\hat{\Upsilon}_k^\eta$ instead of $\hat{\mu}_k = \hat{\Upsilon}_k^0$. The author proposes a class of estimators but suggests using the following modified eigenvalue ratio estimator

$$\hat{r}_{\text{FR}} = \arg \max_{0 \leq k \leq k_{\max}} \frac{\hat{\Upsilon}_k^2}{\hat{\Upsilon}_{k+1}^2}. \quad (3.7)$$

For the consistency of the estimator, Assumptions (S.1) and (S.2) must be strengthened by requiring $B'B \rightarrow D$, a diagonal matrix with nonzero entries $d_1 > d_2 > \dots > d_r$, and $\text{plim}_{T \rightarrow \infty} F'F = I_r$.⁸ At the same time, Assumption (S.2) is relaxed by allowing $0 < d_k/n^{\alpha_k} < \infty$ for $k = 1, \dots, r$ and $\alpha_k > 1/2$, thus accommodating for less pervasive factors.

⁶It is assumed that $Q'Q \otimes PP'$ has bounded eigenvalues for all n, T .

⁷They also propose an alternative estimator, the Growth Ratio. We focus on the ER estimator because it is more widely used in the literature and directly comparable with the FR estimator of Freyaldenhoven (2022), which is also included in our comparative analysis.

⁸Simulation evidence shows that the estimator is robust to violations of this assumption.

3.3. Estimators for dynamic factor models

In this section, we consider the DFM defined in (23) and focus on estimators of the number of dynamic factors q . Specifically, we examine two approaches. Both begin by estimating the VAR(p) representation in (24) via ordinary least squares, replacing F_t with $\hat{F}_t^{(\hat{r})}$, where $\hat{r}$ is obtained from one of Bai and Ng (2002)'s estimators. The resulting coefficient estimates are denoted by $\hat{\Phi}_i$, $i = 1, \dots, p$. In what follows, with a slight abuse of notation, we denote by $B = (C_0, \dots, C_s)$ the loading matrix in the static representation, i.e. $X_t^{\text{dyn}} = BF_t$.

- (6) [AW] (Amengual and Watson, 2007). In their paper, Amengual and Watson note that $Y_t = X_t - \sum_{i=1}^p B\Phi_i F_{t-i} = Gu_t + \epsilon_t$, with $G = BR$, so that Y_t can be represented as a static factor model in terms of the common shocks u_t . Therefore, their estimator of the number of dynamic factors, denoted $\hat{q}_{AW}$, can be obtained by applying the IC_{p_2} criterion to $\hat{Y}_t := X_t - \sum_{i=1}^p \hat{B}^{(\hat{r})} \hat{\Phi}_i \hat{F}_{t-i}^{(\hat{r})}$, where $\hat{r}$ is a consistent estimator of r . In their paper, they employ $\hat{r}_{\text{IC}_{p_2}}$.⁹
- (7) [BNS], Bai and Ng (2007). Consider again Equation (24) and let $\Omega_u := \mathbb{E}(u_t u_t')$ denote the $q \times q$ full-rank covariance matrix of the common shocks. Bai and Ng (2007) exploit the fact that the covariance matrix $\Omega_\xi = \mathbb{E}(\xi_t \xi_t') = R\Omega_u R'$ is of rank q . Hence, the k -th eigenvalue of $\hat{\Omega}_\xi = T^{-1} \sum_{t=1}^T \hat{\xi}_t \hat{\xi}_t'$, with $\hat{\xi}_t = \hat{F}_t^{(\hat{r})} - \sum_{i=1}^p \hat{B}^{(\hat{r})} \hat{\Phi}_i \hat{F}_{t-i}^{(\hat{r})}$, should be small for $k > q$, where $\hat{r}$ is a consistent estimator of r , like, e.g., $\hat{r}_{\text{IC}_{p_2}}$. Let $\hat{v}_k = \psi_k(\hat{\Omega}_\xi)$. Then, for $k = 1, \dots, \hat{r}$,

$$\hat{q}_{\text{BNS}} = \min \left\{ k : \hat{D}(k) < \frac{C}{\min\{n^{1/2-\delta}, T^{1/2-\delta}\}} \right\}, \quad \hat{D}(k) := \left(\frac{\hat{v}_k^2}{\sum_{j=1}^{\hat{r}} \hat{v}_j^2} \right)^{1/2}, \quad (3.8)$$

where the constants C and δ satisfy $0 < C < \infty$ and $0 < \delta < 1/2$. 3.4 Estimators for Generalized Dynamic Factor Models In this section, we consider the GDFM specified in (22) and examine estimators for the number of dynamic factors q .

- (8) [HL] (Hallin and Liška (2007)). Using the frequency domain approach, Hallin and Liška (2007) adapt the information-criterion of Bai and Ng (2002) to the GDFM framework. The estimator is defined as

$$\hat{q}_{\text{HL}}^{(c)} \equiv \arg \min_{1 \leq k \leq k_{\max}} \left[\ln \left(\frac{1}{n} \sum_{j=k+1}^n \frac{1}{2m_T + 1} \sum_{l=-m_T}^{m_T} \tilde{\lambda}_j(\theta_l^*) \right) + kc p(n, T) \right], \quad (3.9)$$

where $p(n, T)$ is a penalty term,¹⁰ while the scaling constant c is selected using the stability interval approach described in Appendix B1.2. In addition to Assumptions (D.1)–(D.3), X_t is also required to admit a Wold representation with summable coefficients and finite fourth-order cumulants.

- (9) [ON] (Onatski, 2009). The author develops a test for the null hypothesis $H_0 : q = k_0$ against the alternative $H_1 : k_0 < q \leq k_1$, based on the statistic

$$\mathcal{R}(\theta_0) := \max_{k_0 < k \leq k_1} \frac{\hat{\lambda}_k(\theta_0) - \hat{\lambda}_{k+1}(\theta_0)}{\hat{\lambda}_{k+1}(\theta_0) - \hat{\lambda}_{k+2}(\theta_0)}, \quad -\pi < \theta_0 \leq \pi,$$

where $\hat{\lambda}_k(\theta_0)$ denotes the k -th eigenvalue of the spectral density matrix at given frequency θ_0 . Intuitively, the statistic measures the curvature of the frequency-domain scree plot (the plot of $\hat{\lambda}_k(\theta_0)$ against k), which attains its maximum at $k = q$. Hence, $\mathcal{R}(\theta_0)$ will be large whenever $k_0 \leq q$.

⁹See Amengual and Watson (2007) for alternative estimators.

¹⁰The penalty $p(n, T) = (m_T^{-2} + m_T^{1/2} T^{-1/2} + n^{-1}) \ln \left[\min \left(n, m_T^2, m_T^{-1/2} T^{1/2} \right) \right]$ offers the best performance in their simulation study, where m_T is the chosen bandwidth.

Setting $k_1 = k_{\max}$, the test can be applied sequentially with $k_0 = 1, 2, \dots, k_1$. Asymptotically, the null is not rejected once $k_0 > q$. The estimator is then defined as

$$\hat{q}_{\text{ON}}(\theta_0) := \min\{k_0 : H_0 \text{ is not rejected}\}. \quad (3.10)$$

The consistency of the estimator requires weaker conditions than those of (D.1). Specifically, Assumption (D.1) is relaxed by requiring the largest q eigenvalues of $\Sigma_n^X(\theta_0)$ to diverge faster than $n^{2/3}$. Assumption (D.2) holds because ϵ_t is a square-summable linear combination of processes with absolutely summable covariance functions. Finally, T is assumed to grow faster than n , with $m_T/n \rightarrow y \in (0, \infty)$, which is not in the spirit of the recent literature.

- (10) [UNALSE], Barigozzi and Farnè (2024). In their paper, the authors propose an estimator for the n -dimensional spectral density matrices of X_t , $\Sigma_n^X(\theta)$, $\theta \in [-\pi, \pi]$. They assume a low-rank plus sparsity decomposition of the form $\Sigma_n^X(\theta) = L(\theta) + S(\theta)$, where $L(\theta)$ has rank $q < n$ and $S(\theta)$ is sparse. The Algebraic Spectral Estimator (ALSE) is then defined as

$$[\hat{L}(\theta), \hat{S}(\theta)] = \arg \min_{L(\theta), S(\theta)} \frac{1}{2} \|\tilde{\Sigma}(\theta) - (L(\theta) + S(\theta))\|^2 + \varphi \|L(\theta)\|_* + \rho \|S(\theta)\|_1, \quad (3.11)$$

where φ and ρ are threshold parameters, and $\|\cdot\|_*$ and $\|\cdot\|_1$ denote the nuclear and ℓ_1 norms, respectively. The authors establish both algebraic and parametric consistency of the estimator.¹¹ Since the decomposition of $\Sigma_n^X(\theta)$ is closely related to the GDFM, they define¹²

$$\hat{q}_{\text{UNALSE}}(\theta) = \text{rank}(\hat{L}(\theta)). \quad (3.12)$$

Assumption (D.2) is relaxed by allowing $\lambda_{kn}(\theta)/n^\alpha$ to remain bounded and bounded away from zero for $\alpha \in (1/2, 1]$. On the other hand, the coefficients $B_0, B_1, \dots$ of the polynomial matrix $B(L)$ are assumed to share the same left null space.¹³

- (11) [DDR] (Avarucci et al. 2026). Building on the scree-plot idea, Avarucci et al. (2026) propose an estimator that applies to Fourier frequencies in the set $\mathcal{L} = [\omega_{\underline{\ell}}, \omega_{\bar{\ell}}]$.¹⁴ The implicit assumption is that q is constant across frequencies in the interval. The estimator is defined as

$$\hat{q}_{\text{DDR}}^{(\mathcal{L})} := \arg \max_{1 \leq k \leq k_{\max}} \frac{\sum_{\ell=\underline{\ell}}^{\bar{\ell}} (\hat{\lambda}_k(\theta_\ell) - \hat{\lambda}_{k+1}(\theta_\ell))}{\max \left\{ \sum_{\ell=\underline{\ell}}^{\bar{\ell}} (\hat{\lambda}_{k+1}(\theta_\ell) - \hat{\lambda}_{k+2}(\theta_\ell)), \sum_{\ell=\underline{\ell}}^{\bar{\ell}} \hat{\lambda}_{2b_T+1}(\theta_\ell) \right\}}, \quad (3.13)$$

and we denote by $\hat{q}_{\text{DDR}}$ the special case where $\mathcal{L} = [0, \pi]$. The assumptions on the idiosyncratic component are similar to those in Onatski (2009), and it is further required that $n/T \rightarrow y \in (0, \infty)$ as $n, T \rightarrow \infty$.

4. Simulation experiments

This section describes the design of the simulation experiments used to evaluate the performance of the proposed methods. We generate T observations of an n -dimensional process $\mathbf{X}_t$ with the following factor structure:

$$\mathbf{X}_t = \boldsymbol{\chi}_t^* + \boldsymbol{\epsilon}_t, \quad (4.1)$$

¹¹ Algebraic consistency means that, with probability approaching one, $\hat{S}(\theta)$ is positive definite with the correct sparsity pattern, while $\hat{L}(\theta)$ is positive definite with rank q . The parametric consistency refers to convergence of $\hat{S}(\theta)$ and $\hat{L}(\theta)$ to their population counterparts.

¹² UNALSE means Unshrunk ALSE. The unshrinking does not affect the parametric consistency, so $\hat{q}_{\text{UNALSE}}(\theta) = \hat{q}_{\text{ALSE}}(\theta)$. We use UNALSE for consistency with the author's notation.

¹³ Despite being a restrictive assumption, our simulation indicates that the assumption may not be necessary.

¹⁴ They also introduce two additional estimators extending the ER and GR estimators of Ahn and Horenstein (2013) to the GDFM.

where χ_t^* and ε_t represent the common and idiosyncratic components, respectively. The superscript $*$ indicates that the common component varies across experiments, while the data-generating process of the idiosyncratic part ϵ_t remains fixed throughout. We normalize the variance of the zero-mean processes χ_{it}^* and ϵ_{it} to be 0.5, such that for every unit, the common component explains exactly 50% of the variation in the data and the variance of X_{it} is equal to 1.

This setup allows us to adopt a *unified DGP* that spans different factor model specifications. We start from the most general case (the GDFM) and obtain more restrictive models by imposing suitable parameter restrictions: the DFM as an intermediate case, and the SFM as the most restrictive specification. In this way, the framework enables a systematic assessment of performance across models with varying flexibility of the common component. Moreover, for each simulation experiment, we consider different *factor strengths* (see Section 4.3), in order to examine how the methods perform under stronger or weaker contributions of the common component.

4.1. The idiosyncratic component

The idiosyncratic component maintains the same DGP across all simulation experiments. Specifically, for each unit i and time t we define $\epsilon_{it} = \vartheta \varepsilon_{it}$, where

$$\varepsilon_{it} = \rho \varepsilon_{i,t-1} + v_{it} + \sum_{h=\max(i-J,1)}^{i-1} \beta v_{ht} + \sum_{h=i+1}^{\min(i+J,n)} \beta v_{ht}. \quad (4.2)$$

The scaling factor is given by $\vartheta = [0.5(1 - \rho^2)/(1 + 2J\beta^2)]^{1/2}$, with innovations $v_{it} \sim \mathcal{N}(0, 1)$. In the baseline specification, we set $\rho = 0.5$, $\beta = 0.2$, and $J = 10$. This construction ensures that ϵ_{it} has variance 0.5 for most cross-sectional units.¹⁵ The specification induces a mild, yet non-negligible, degree of both temporal and cross-sectional dependence in the idiosyncratic component. Slight variations of this data-generating process have been previously considered in the literature, including Bai and Ng (2002), Alessi, Barigozzi, and Capasso (2010), Onatski (2010), Kapetanios (2010), Ahn and Horenstein (2013), among others. In particular, some contributions adopt $J = \max\{\lfloor n/20 \rfloor, 10\}$; however, we fix $J = 10$ throughout to ensure that assumptions (D.1) and (S.3) are satisfied.

4.2. The common component

We consider three nested specifications of the common component χ_t^* , starting from the most general GDFM and progressively restricting it to obtain the DFM and the simplest SFM.

4.2.1. Generalized dynamic factor model (GDFM)

Using the notation introduced in Section 2, we define the common component in (41) under the GDFM as

$$\chi_{it}^* = \chi_{it}^{\text{gdyn}} = \varphi_i(a, c) \sum_{j=1}^q b_{ij}(L) u_{jt}, \quad \text{with} \quad b_{ij}(L) = \frac{c_{ij,0} + c_{ij,1}L + c_{ij,2}L^2}{1 - a_{ij}L}, \quad (4.3)$$

where $c_{ij,k} \sim \mathcal{N}(0, 1)$ for $k = 0, 1, 2$, $a_{ij} \sim \mathcal{U}[-0.8, 0.8]$, and all draws are mutually independent across i, j, k . The scalar $\varphi_i(a, c)$ depends on the ARMA parameters and ensures standardization such that $\text{Var}(\chi_{it}^{\text{gdyn}}) = 0.5$.

¹⁵More precisely, this holds for units with index $J + 1 \leq i \leq n - J$.

4.2.2. Dynamic factor model (DFM)

The DFM is obtained as a restricted version of the GDFM by fixing $a_{ij} = 0.5$ for all i, j , while keeping the other parameters unchanged. Substituting into (43) yields

$$\chi_{it}^* = \chi_{it}^{\text{dfm}} = \varphi_i(a, c) \sum_{j=1}^q (c_{ij,0} + c_{ij,1}L + c_{ij,2}L^2) f_{jt}, \quad (4.4)$$

where the dynamic factors are given by $f_{jt} = (1 - 0.5L)^{-1} u_{jt}$ for $j = 1, \dots, q$. In matrix form, the DFM admits a static representation by stacking current and lagged factors, i.e. $\mathbf{F}_t = [\mathbf{f}'_t, \mathbf{f}'_{t-1}, \mathbf{f}'_{t-2}]'$, which is an $r = 3q$ -dimensional VAR(1) process with nonsingular covariance matrix.

4.2.3. Static factor model (SFM)

Finally, the static factor model is obtained from the DFM by further imposing $c_{ij,1} = c_{ij,2} = 0$ in (44), for all i, j . The common component then reduces to

$$\chi_{it}^* = \chi_{it}^{\text{stat}} = \sum_{j=1}^q b_{ij} F_{jt}, \quad (4.5)$$

where $b_{ij} = \varphi_i(a, c) c_{0,ij}$, with $c_{ij,0} \sim \mathcal{N}(0, 1)$, and $F_{jt} = f_{jt}$.

Thus, the three models (GDFM, DFM, and SFM) are nested within one another, with the GDFM representing the most flexible specification and the SFM the most restrictive.

4.3. Factor strength

We investigate how the strength of the factors affects the performance of the methodologies under examination. Factor strength is defined in terms of the limiting behavior of the sums of the corresponding loadings, either in absolute value or in squares. Following the taxonomy in Bailey, Kapetanios, and Pesaran (2021) and Chudik, Pesaran, and Tosetti (2011), and for simplicity considering the static factor model (45), we define the notion of factor strength as follows. Let α be a positive constant in the range $0 \leq \alpha \leq 1$, and consider the condition

$$\lim_{n \rightarrow \infty} n^{-\alpha} \sum_{i=1}^n |b_{ij}| = C < \infty. \quad (4.6)$$

A factor is said to be *strong* if $\alpha = 1$ and *weak* if $\alpha = 0$. For intermediate values $\alpha \in (0, 1)$, the factor can be classified as either *semi-strong* or *semi-weak*. Specifically, factors with $0 < \alpha < 1/2$ are regarded as semi-weak, while those with $1/2 \leq \alpha < 1$ are regarded as semi-strong.¹⁶ To simulate a SFM where the k -th factor is not strong, we define $\mathcal{A}$ as a random subset of the first n natural number with cardinality $n - \lfloor n^\alpha \rfloor$, for $\alpha \in [0, 1)$, where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to $x \in \mathbb{R}$. Then we set $b_{ik} = 0$ for all $i \in \mathcal{A}$.

Noting that $\Sigma_n^X(\theta) = \mathbf{B}(\theta)\mathbf{B}(\theta)'$, we extend the definition of factor strength to the GDFM by replacing b_{ij} with the transfer function $\sup_\theta b_{ij}(\theta)$ in (46). Similarly to the static case, in our simulations we will consider the case $b_{ik}(L) = 0$ for all $i \in \mathcal{A}$ to generate a dataset where only a subset of units are affected by the k -th common shock.

Finally, while most of the estimators considered in Section 3 require specific restrictions on α for their validity, it is worth noting that the ED estimator only requires $\alpha > 0$.

¹⁶These definitions can equivalently be expressed in terms of the sum of the *squared* loadings (see Onatski, 2012; Lettau and Pelger, 2020a, 2020b, among others). When one assumes uniform boundedness of the loadings, i.e. $\sup_i \sup_j |b_{ij}| \leq C < \infty$, (46) can be sharpened to show that the same rates are obtained whether one considers the loadings in levels or in squares (see the first formula on p. 932 of Bailey, Kapetanios, and Pesaran, 2016).

4.4. Simulation settings

In all experiments, across the three different DGPs, we set $T, n \in \{60, 120, 180, 240, 300\}$, $\alpha \in \{1, 0.8, 0.65, 0.51, 0.3\}$. When analyzing the GDFM, we set $q = 2$ (noting that the number of static factors r is not defined). Accordingly, in the SFM we set $r = q = 2$, while in the static re DFM we set $q = 2$ and r is equal to six in the static representation. The first factor (or common shock) is always strong, whereas the strength of the second factor will vary according to α . For all estimators that require the pre-specification of $k_{\max}$, we set $k_{\max} = 8$. The results of these experiments are reported in Section 5.

As further robustness checks, we also consider three complementary experiments, whose results are presented in the Appendix for the sake of space.

In the first case, we assume that both factors are non-strong and share the same factor strength parameter, $\alpha < 1$. Under this specification, the observed variable X_{it} may be affected by both factors, by only one of them, or none. The results for this case are reported in Appendix A1.5.

In the last two experiments, we focus on the strong-factor case ($\alpha = 1$) for simplicity and examine the sensitivity of the estimates to the choice of $k_{\max}$ and to a larger number of factors. In particular, in the former experiment we consider $k_{\max} \in \{6, 9, 12, 15, 18\}$ with $r = q = 2$ and $T = n \in \{120, 240\}$ (see Appendix A1.6). In the latter experiment, we set $r = q = 5$ while keeping $k_{\max} = 8$ (Appendix A1.7). Each experiment is based on 500 Monte Carlo replications.

5. Results

In this section, we present the results obtained from the three simulation experiments described in Section 4. The results are organized in subsections by experiment: Section 5.1 reports the findings for the experiment where the underlying DGP is an SFM; it is a DFM in Section 5.2 and a GDFM in Section 5.3. Each experiment is analyzed across all classes of estimators (designed for a static, dynamic, and generalized dynamic common component) and over a range of factor strengths.

Supplementary material with detailed numerical results is reported in the Appendix: Tables A1–A2 for the static experiments, Tables A3–A4 for the dynamic experiments, and Tables A5–A6 for the GDFM experiments. Additional plots showing the evolution of each estimator across values of α and experimental settings are provided in Figures A7–A17.

Throughout the following sections, for ease of exposition and with a slight abuse of terminology, we use the shorthand *static estimators*, *dynamic estimators*, and *GDFM estimators* to refer, respectively, to the estimators designed for the SFM, the DFM, and the GDFM. Similarly, the static, dynamic, and GDFM experiments refer to the underlying DGP.

5.1. Experiment 1: Static factor model

This section discusses the results of the static experiment (see Section 4, Equation (45)). The main results are summarized in Figure 1. For each estimator on the horizontal axis, the bar reports the frequency with which the selected number of factors equals the true number ($r = 2$, *Correct*, green), is under-selected (*Under*, blue), or over-selected (*Over*, red). The plot first summarizes the strong-factor case ($\alpha = 1$) in panel (a) and then tracks how performance evolves as α declines ($\alpha = 0.8, 0.65, 0.51, 0.3$) in panels (b)–(e).¹⁷ All figures are reported for $n = T = 60, 120, 240, 300$. Complementary evidence for $n \neq T$ cases are shown in Appendix in Figures A1–A2, which display performance as n increases with $T = 300$ held fixed, and as T increases with $n = 300$ held fixed, respectively.

¹⁷Our choice of α reflects our desire to show the performance of the various estimators under a variety of cases, in particular showing how the degree of (semi-) weakness affects the results. Case $\alpha = 0.3$ represents a case of extreme weakness, where all the estimators are not supposed to work in view of their theoretical analysis. We then explore an increasing degree of strength from 0.51 and 0.65 (two different degrees of semi-weak) to 0.8 (semi-strong), until one obtains the traditional case of strong factors ($\alpha = 1$).

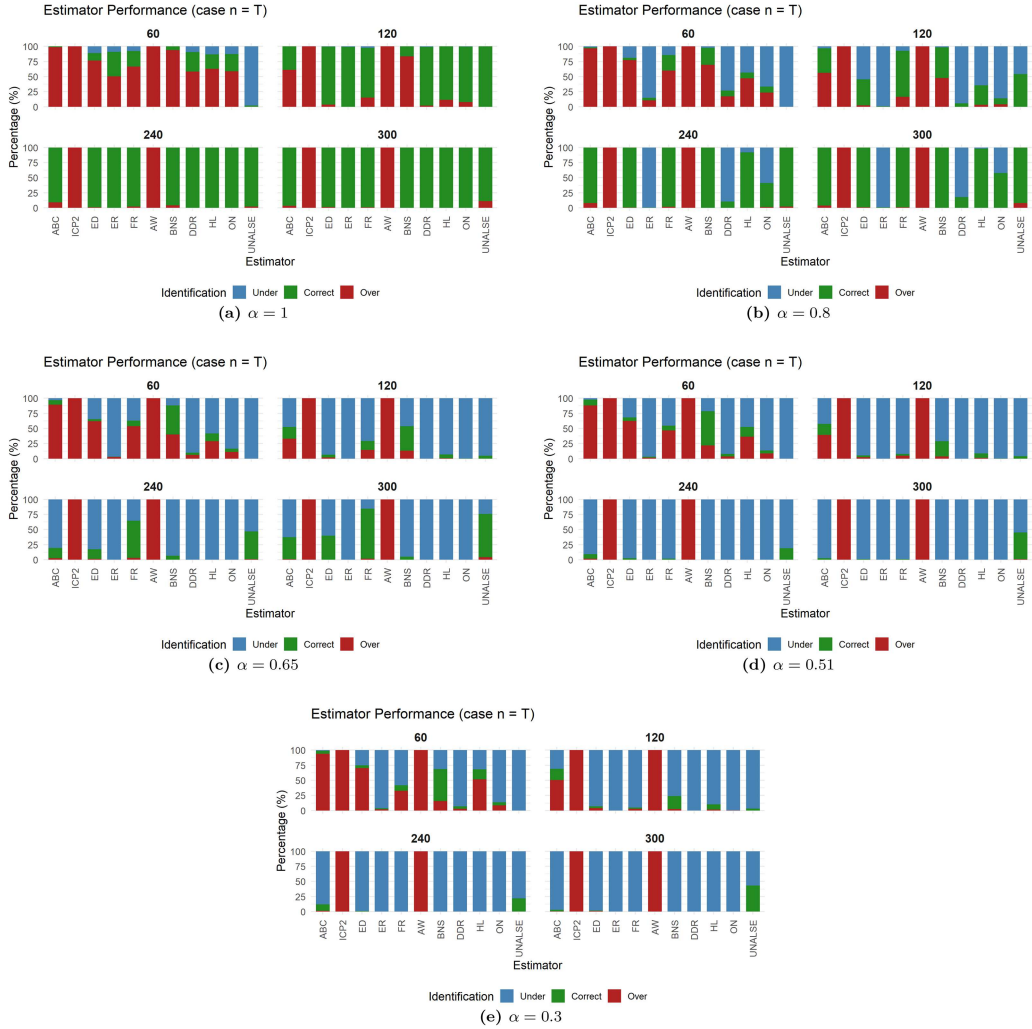


Figure 1. Static experiment: selection frequencies by estimator, sample size, and factor strength. Each panel reports the distribution of outcomes across estimators when $n = T \in \{60, 120, 240, 300\}$, under the DGP defined in Section 4, Equation (45), and $r = 2$. Bars sum to 100% and are decomposed into *Correct* identification (green), *Under* (blue), and *Over* (red). Panels (a)–(e) correspond to $\alpha = 1, 0.8, 0.65, 0.51, 0.3$, respectively, ordered from strong to semi-weak factors. The vertical axis reports percentages, while the horizontal axis lists the estimators: ABC (Alessi, Barigozzi, and Capasso (2010)), IC_{p_2} (Bai and Ng (2002)), ED (Onatski (2010)), ER (Ahn and Horenstein (2013)), FR (Freyaldenhoven (2022)), AW (Amengual and Watson (2007)), BNS (Bai and Ng (2007)), DDR (Avarucci et al. (2026)), HL (Hallin and Liška (2007)), ON (Onatski (2009)), UNALSE (Barigozzi and Farnè (2024)).

When the factor is strong ($\alpha = 1$), several GDFM estimators display perfect or near-perfect selection. Among static estimators, ER delivers 100% correct selection and performs very similarly to FR, indicating that its robustness does not come at a cost, except when n or T are very small. ED and ABC also perform well, reaching 99.2% and 98.8%, respectively, when both n and T are large. By contrast, IC_{p_2} fails (it systematically over-selects, identifying $\hat{r} = k_{\max}$) due to the non-trivial correlation structure of the idiosyncratic component. This behavior has already been documented by Alessi, Barigozzi, and Capasso (2010) and Onatski (2010).

Among dynamic estimators, AW does not perform well, as expected, because it employs IC_{p_2} in both steps and thus inherits its bias under autocorrelation and cross-sectional correlation in the idiosyncratic term.¹⁸ For BNS, the effect of an overestimated r in the first step is less severe, because the second stage

¹⁸In simulations not shown here, performance improves markedly when the first-step choice is replaced with $\hat{r} = \hat{r}_{ABC}$.

relies on an estimator different from IC; its performance improves and becomes very strong as the sample size increases.

All the GDFM estimators perform very well, delivering results comparable to the best static estimators. In our experiment, the HL estimator achieves its best performance with a stability interval larger than 0.05 when subsamples are only considered for the cross-sectional dimension, whereas performance deteriorates under the parameter choices recommended in the original paper.¹⁹ The UNALSE estimator performs well when both n and T exceed 60, although in that region it tends to under-reject. Moreover, its performance does not always improve as n and T increase, which is likely due to the choice of tuning parameters.

As the factor strength declines to $\alpha = 0.8$, the GDFM class remains the most reliable overall, especially when n and T are large: HL stays close to 100% correct, although it requires larger sample sizes, and UNALSE also remains highly accurate (about 97% when $n = T = 300$). By contrast, ON and DDR begin to under-select.²⁰ Within the dynamic class, BNS remains at 100%, while AW continues to systematically over-select. Among the static group, ED, FR, and ABC retain high accuracy (99.2%, 99.0%, 96.2%), while ER starts to deteriorate. These results are expected for FR and ED, which do not rely on a strong factor structure. The performance of ABC is consistent with recent theoretical results in Onatski (2024).²¹

When $\alpha = 0.65$, correct identification starts to decrease across classes. Among the GDFM estimators, only UNALSE retains some power when both n and T are large, while HL drops to notable under-selection. Within the static class, only FR exhibits reasonable performance. When $\alpha < 0.65$, performance deteriorates rapidly across all methods, with UNALSE still retaining some correct identification, particularly when T is large.

To complement the above discussion and to provide practical guidance to practitioners, Table 1 offers a compact comparison across alternative n – T configurations using the dot summaries. For each estimator, the table reports the correct-selection rate summarized as dot ratings by factor strength (columns) and by the n – T relation: $n = T$ (300,300), $n > T$ (300,180), and $n < T$ (180,300). A rating of $\bullet \bullet \bullet \bullet$ corresponds to $\geq 75\%$ correct, $\bullet \bullet \bullet \circ$ to 50.1–74.9%, $\bullet \bullet \circ \circ$ to 25.1–50.0%, $\bullet \circ \circ \circ$ to 10.1–25.0%, and $\circ \circ \circ \circ$ to $\leq 10\%$. The factor strength buckets are defined as “Strong” ($\alpha = 1$), “Semi-strong” ($\alpha = 0.8$), and “Semi-weak” ($\alpha = 0.30$). The final column highlights the main advantages and disadvantages of each estimator.

Taken together, Table 1 shows that, relative to the $n = T$ baseline, the performance of static and dynamic criteria remains unchanged when $n < T$ or $n > T$. Among the GDFM estimators, ON exhibits some improvement under the semi-strong setting when $n = T$, whereas UNALSE improves under the semi-weak factor setting when the time dimension dominates ($n < T$).

5.2. Experiment 2: Dynamic factor model

This section discusses the results of the dynamic experiment (see Section 4, Equation (44)). The main selection results are in Figure 2, following the graphical convention of Figure 1 in the static case, with the true number of factors now given by $q = 2$ (corresponding to $r = 6$ when evaluating the static estimators under this DGP). Supplementary evidence for $n \neq T$ is provided in Appendix Figures A3–A4, which report performance as n varies with $T = 300$ fixed, and as T varies with $n = 300$ fixed, respectively. Table 2 also provides a compact comparison via dot summaries.

With a strong factor ($\alpha = 1$), the GDFM class is clearly dominant: ON, HL, and DDR achieve 100% correct identification, while UNALSE is near-perfect. Among the static rules, ABC performs well when

¹⁹Details are available upon request.

²⁰DDR improves considerably when increasing the bandwidth m_T : using the same m_T as in Onatski (2009), DDR performs much better. However, improvements occur only in this specific case, and increasing the bandwidth is generally not advisable.

²¹Onatski (2024) uses large random matrix theory to analyze the behavior of Hallin and Liška (2007)’s criterion in a static setting. The author shows that HL remains consistent even when the factors are semi-weak, and these results extend to ABC.

Table 1. Static experiment: correct-selection summary by estimator class, factor strength, and n – T relationship, under the static DGP defined in Section 4, Equation (45), and $r = 2$.

Estimator	n vs T	Strong	Semi-strong	Semi-weak	Advantages (+)/ Disadvantages (–)
IC _{P2}	$n = T$	oooo	oooo	oooo	(–) It tolerates little to no cross-sectional dependence in the idiosyncratic error.
	$n > T$	oooo	oooo	oooo	(+) n, T can diverge at different rates.
	$n < T$	oooo	oooo	oooo	(–) The penalty term is defined up to a free parameter.
ABC	$n = T$	●●●●	●●●●	oooo	(+) Performance improves when $n \geq 240$ across different T .
	$n > T$	●●●●	●●●●	oooo	(–) When factor strength $\alpha < 0.8$, it tends to under-estimate, especially in large samples.
	$n < T$	●●●●	●●●●	oooo	(+) The tuning parameter improves the performance when idiosyncratic errors are cross-correlated.
ED	$n = T$	●●●●	●●●●	oooo	(–) User-chosen settings: subsample construction and stability interval width.
	$n > T$	●●●●	●●●●	oooo	(+) Performance improves when $n \geq 120$ across different T .
	$n < T$	●●●●	●●●●	oooo	(+) Factors are not assumed to be strong
ER	$n = T$	●●●●	oooo	oooo	(–) When factor strength $\alpha < 0.8$, it tends to under-estimate.
	$n > T$	●●●●	oooo	oooo	(+) Performance improves when $n \geq 120$
	$n < T$	●●●●	oooo	oooo	(–) Performance deteriorates rapidly when factors are not strong ($\alpha \leq 0.8$).
FR	$n = T$	●●●●	●●●●	oooo	(+) No user-chosen parameters.
	$n > T$	●●●●	●●●●	oooo	(+) Factors need not be strong: use eigenvector information, not only eigenvalues.
	$n < T$	●●●●	●●●●	oooo	(–) User-chosen parameters: “concentration” measure of eigenvectors
AW	$n = T$	oooo	oooo	oooo	(+) Reasonable performance also when $\alpha = 0.65$, especially in large samples.
	$n > T$	oooo	oooo	oooo	(–) Bias arises under autocorrelation and cross-sectional correlation in the idiosyncratic term.
	$n < T$	oooo	oooo	oooo	(+) Performance improves markedly when the first-step choice is fixed at $\hat{r} = \hat{r}_{ABC}$.
BNS	$n = T$	●●●●	●●●●	oooo	(+) Time domain analysis.
	$n > T$	●●●●	●●●●	oooo	(+) Second step estimate is more robust to cross-sectional dependence, the idiosyncratic component.
	$n < T$	●●●●	●●●●	oooo	(±) Performance requires n sufficiently large ($n \geq 240$), otherwise severe over-estimation occurs.
HL	$n = T$	●●●●	●●●●	oooo	(+) No need to compute the spectral density.
	$n > T$	●●●●	●●●●	oooo	(–) When factor strength $\alpha < 0.8$, it tends to under-estimate, especially in large samples.
	$n < T$	●●●●	●●●●	oooo	(+) Performs well when $n \geq 120$ across different T . With a small n , it leads to some over-estimation.
ON	$n = T$	●●●●	●●●○	oooo	(+) n, T can diverge at different rates
	$n > T$	●●●●	●●●○	oooo	(–) User-chosen settings: subsample construction and stability interval width
	$n < T$	●●●●	●●●○	oooo	(–) When factor strength $\alpha \leq 0.65$, it tends to under-estimate, especially in large samples.
UNALSE	$n = T$	●●●●	●●●●	●●●○	(+) Performs well when $n \geq 120$ across different T . With a small n , it leads to some over-estimation.
	$n > T$	●●●●	●●●●	oooo	(±) T is assumed to grow much faster than n but performance remains good when they are of comparable size
	$n < T$	●●●●	●●●○	●●●○	(±) Estimates may differ across frequencies.
DDR	$n = T$	●●●●	●○○○	oooo	(±) The eigenvalues are required to diverge faster than $n^{2/3}$. Performance deteriorates when $\alpha \leq 0.8$
	$n > T$	●●●●	oooo	oooo	(+) Retains some correct identification even when factors are semi-weak
	$n < T$	●●●●	oooo	oooo	(+) No need to specify $k_{\max}$.
DDR	$n = T$	●●●●	●○○○	oooo	(±) Computed at a single frequency
	$n > T$	●●●●	oooo	oooo	(–) Threshold parameters need to be properly chosen at each frequency.
	$n < T$	●●●●	oooo	oooo	(+) Can be applied over a user-chosen frequency range, assuming q is constant within the range.
					(+) Performs well when $n \geq 120$ across different T . With small n , it leads to some over-estimation.
					(–) When factor strength $\alpha \leq 0.8$, it tends to under-estimate.
					(+) Does not require user-chosen parameters, apart from the bandwidth to compute the smoothed periodogram.

The $n = T$ case corresponds to $(n, T) = (300, 300)$, the $n > T$ case to $(n, T) = (300, 180)$, and the $n < T$ case to $(n, T) = (180, 300)$. Dots indicate the average correct-selection rate within each strength bucket (4 = $\geq 75\%$, 3 = 50.1–74.9%, 2 = 25.1–50.0%, 1 = 10.1–25.0%, 0 = $\leq 10\%$). Strength buckets are defined as: Strong ($\alpha = 1$), Semi-strong ($\alpha = 0.8$), and Semi-weak ($\alpha = 0.3$). For references, we recall each estimator by citing the original paper in which it was proposed: IC_{P2} (Bai and Ng (2002)), ABC (Alessi, Barigozzi, and Capasso (2010)), ED (Onatski (2010)), ER (Ahn and Horenstein (2013)), FR (Freyaldenhoven (2022)), AW (Amengual and Watson (2007)), BNS (Bai and Ng (2007)), HL (Hallin and Liška (2007)), ON (Onatski (2009)), UNALSE (Barigozzi and Farnè (2024)), DDR (Avarucci et al. (2026)).

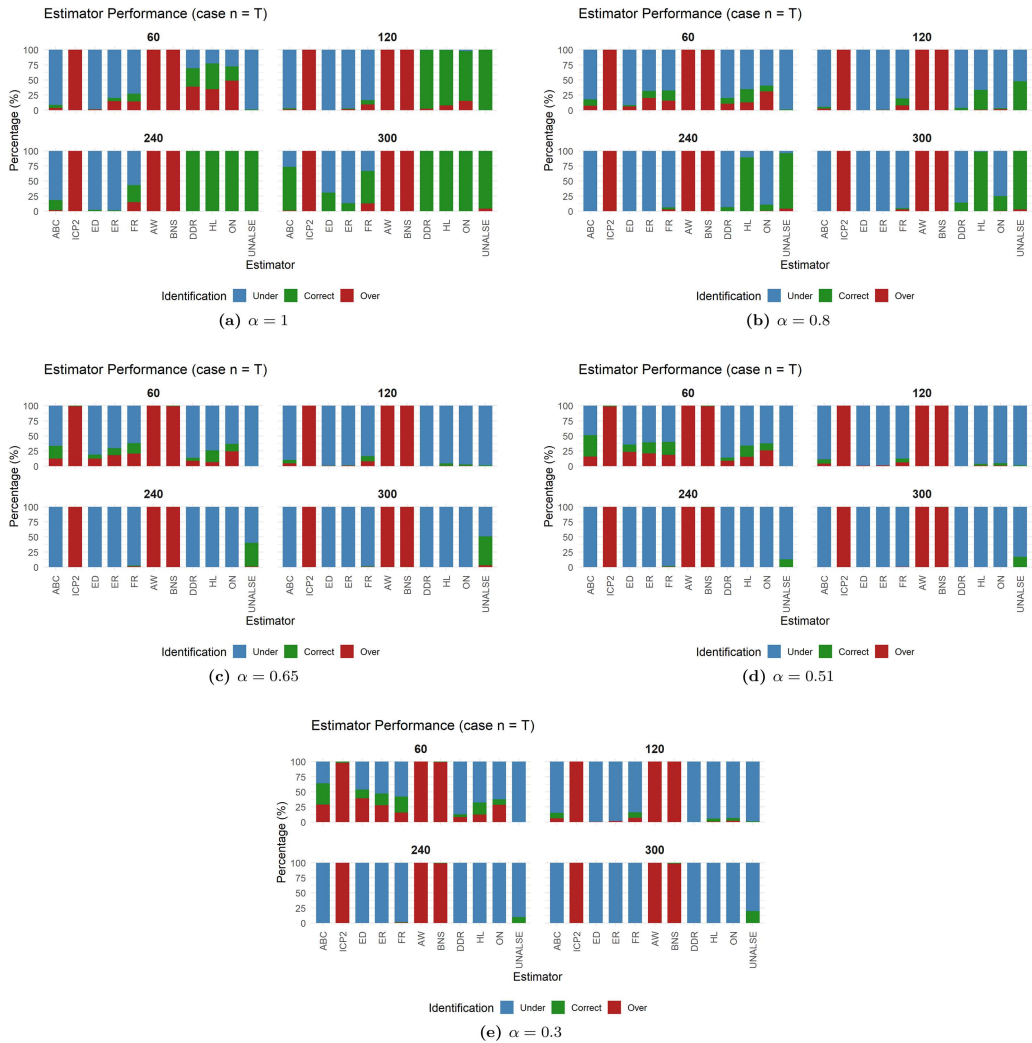


Figure 2. Dynamic experiment: selection frequencies by estimator, sample size, and factor strength. Each panel reports the distribution of outcomes across estimators when $n = T \in \{60, 120, 240, 300\}$, under the DGP defined in Section 4, Equation (44) and $q = 2$ (implying $r = 6$). Bars sum to 100% and are decomposed into *Correct* identification (green), *Under* (blue), and *Over* (red). Panels (a)–(e) correspond to $\alpha = 1, 0.8, 0.65, 0.51, 0.3$, respectively, ordered from strong to semi-weak factors. The vertical axis reports percentages, while the horizontal axis lists the estimators: ABC (Alessi, Barigozzi, and Capasso (2010)), IC_{p2} (Bai and Ng (2002)), ED (Onatski (2010)), ER (Ahn and Horenstein (2013)), FR (Freyaldenhoven (2022)), AW (Amengual and Watson (2007)), BNS (Bai and Ng (2007)), DDR (Avarucci et al. (2026)), HL (Hallin and Liška (2007)), ON (Onatski (2009)), UNALSE (Barigozzi and Farnè (2024)).

the sample size is large, whereas ED and ER tend to under-estimate.²² FR performs only marginally better than them. By contrast, IC_{p2} continues to over-select, reflecting the cross-sectional dependence in the idiosyncratic component used in our experiment. The dynamic class as a whole over-selects: both BNS and AW over-select in nearly all replications, again due to correlation in the idiosyncratic term.²³

As the factor strength declines to $\alpha = 0.8$, the GDFM estimators remain the most reliable: HL and UNALSE stay highly accurate, especially when n and T are both larger than or equal to 240. By contrast,

²²This is possibly due to correlation across the idiosyncratic terms. Indeed, when they were generated as white noise processes, the estimator attains high accuracy. As the autocorrelation increases, the estimator tends to severely under-select. Details are available upon request.

²³In simulations not reported here, these estimators performed well when the idiosyncratic component is i.i.d.

Table 2. Dynamic experiment: correct-selection summary by estimator class, factor strength, and n – T relationship, under the dynamic DGP defined in Section 4, Equation (44), and $q = 2$ (implying $r = 6$).

Estimator	n vs T	Strong	Semi-strong	Semi-weak	Advantages (+)/ Disadvantages (–)
IC _{p₂}	$n = T$	oooo	oooo	oooo	(–) It tolerates only weak temporal dependence and little to no cross-sectional dependence.
	$n > T$	oooo	oooo	oooo	(+) n, T can diverge at different rates
	$n < T$	oooo	oooo	oooo	(–) The penalty term is defined up to a free parameter.
ABC	$n = T$	●●●○	oooo	oooo	(–) Reliable performance only when n, T are large and factors are strong—under-selection in small samples.
	$n > T$	●●●○	oooo	oooo	(+) The tuning parameter improves the performance when idiosyncratic errors are cross-correlated.
	$n < T$	oooo	oooo	oooo	(–) User-chosen settings: subsample construction and stability interval width.
ED	$n = T$	●●●○	oooo	oooo	(+) Factors are not assumed to be strong
	$n > T$	oooo	oooo	oooo	(–) Under-selects even with strong factors.
	$n < T$	oooo	oooo	oooo	(+) High accuracy when factors are white noise.
ER	$n = T$	●○○○	oooo	oooo	(–) Under-selects even with strong factors
	$n > T$	○○○○	oooo	oooo	(+) Higher accuracy when factors are white noise.
	$n < T$	○○○○	oooo	oooo	(+) No user-chosen parameters
FR	$n = T$	●●●○	oooo	oooo	(–) Eigenvalue gap condition (requires strong factors)
	$n > T$	●●●○	oooo	oooo	(+) Reliable performance only when n, T are large and factors are strong—otherwise, under-selection.
	$n < T$	●○○○	oooo	oooo	(+) Factors need not be strong: use eigenvector information, not only eigenvalues.
AW	$n = T$	○○○○	oooo	oooo	(–) User-chosen parameters: “concentration” measure of eigenvectors
	$n > T$	○○○○	oooo	oooo	(–) Severe over-selection due to auto- and cross-sectional correlation in the idiosyncratic term.
	$n < T$	○○○○	oooo	oooo	(±) Performance improves slightly when the first-step choice is fixed at $\hat{\tau} = \hat{\tau}_{ABC}$.
BNS	$n = T$	○○○○	oooo	oooo	(+) Time domain analysis.
	$n > T$	○○○○	oooo	●○○○	(–) Severe over-selection, even with strong factors, due to idiosyncratic correlation.
	$n < T$	○○○○	oooo	oooo	(+) Estimators perform well with i.i.d. idiosyncratic components.
HL	$n = T$	●●●●	●●●●	oooo	(+) Performance improves markedly when the first-step choice is fixed at $\hat{\tau} = \hat{\tau}_{ABC}$.
	$n > T$	●●●●	●●●●	oooo	(+) Very accurate even when $\alpha = 0.8$ with large n, T . Under-selection as the factor strength decreases.
	$n < T$	●●●●	●●●●	oooo	(+) n, T can diverge at different rates
ON	$n = T$	●●●●	●○○○	oooo	(–) User-chosen settings: subsample construction and stability interval width
	$n > T$	●●●●	○○○○	oooo	(+) Accurate with strong factors and large n, T . Under-selection when $\alpha \leq 0.8$.
	$n < T$	●●●●	●○○○	oooo	(±) The eigenvalues are required to diverge faster than $n^{2/3}$. Performance deteriorates when $\alpha \leq 0.8$
UNALSE	$n = T$	●●●●	●●●●	●○○○	(±) Estimates may depend on the chosen frequency
	$n > T$	●●●●	●●●●	oooo	(+) Retains some correct identification even when factors are weak
	$n < T$	●●●●	●●●●	●○○○	(+) No need to specify $k_{\max}$.
DDR	$n = T$	●●●●	●○○○	oooo	(±) Computed at a single frequency
	$n > T$	●●●●	○○○○	oooo	(+) Accurate with strong factors. Under-selection as the factor strength decreases.
	$n < T$	●●●●	○○○○	oooo	(–) Under-selection as factor strength decreases.
					(+) Can be applied over a user-chosen frequency range, assuming q is constant within the range.

The $n = T$ case corresponds to $(n, T) = (300, 300)$, the $n > T$ case to $(n, T) = (300, 180)$, and the $n < T$ case to $(n, T) = (180, 300)$. Dots indicate the average correct-selection rate within each strength bucket ($4 = \geq 75\%$, $3 = 50.1$ – 74.9% , $2 = 25.1$ – 50.0% , $1 = 10.1$ – 25.0% , $0 = \leq 10\%$). Strength buckets are defined as: Strong ($\alpha = 1$), Semi-strong ($\alpha = 0.8$), and Semi-weak ($\alpha = 0.3$). For references, we recall each estimator by citing the original paper in which it was proposed: IC_{p₂} (Bai and Ng (2002)), ABC (Alessi, Barigozzi, and Capasso (2010)), ED (Onatski (2010)), ER (Ahn and Horenstein (2013)), FR (Freyaldenhoven (2022)), AW (Amengual and Watson (2007)), BNS (Bai and Ng (2007)), HL (Hallin and Liška (2007)), ON (Onatski (2009)), UNALSE (Barigozzi and Farnè (2024)), DDR (Avarucci et al. (2026)).

ON and DDR begin to under-select, with correct rates of roughly 24% and 14%. Within the dynamic class, BNS and AW still tend to over-select in most replications. All static estimators deteriorate sharply at $\alpha = 0.8$: ABC and FR fall close to zero, and the remaining static rules show similarly poor performance. The higher number of static factors ($r = 6$) may contribute to this behavior.

With semi-strong and semi-weak factors ($\alpha = 0.65, 0.51, 0.3$), correct identification becomes increasingly difficult. Among the GDFM estimators, only UNALSE estimates q correctly about 20% of the time. All static procedures tend to under-estimate (except IC_{p_2} , which continues to over-estimate), while dynamic estimators mostly continue to over-select.

To extend the analysis and provide a practical benchmark, Table 1 reports a concise comparison across n - T configurations using dot summaries, analogous to the static experiment. For instance, under strong factors, ABC and FR tend to perform better when n is large and equal to T , but their performance deteriorates rapidly when $n < T$; under semi-weak factors, by contrast, UNALSE appears to favor settings where $n < T$.

5.3. Experiment 3: Generalized dynamic factor model

This section turns to the results of the GDFM experiment (see Section 4, Equation (43)). The main results are summarized in Figure 3 and Table 3, which follow the same graphical convention as in the previous sections.²⁴ Under this experiment, the true number of dynamic factors is set to $q = 2$, corresponding to a representation with an infinite number of static factors.

When the factor is strong ($\alpha = 1$), the GDFM estimators lead the ranking: ON, HL, and DDR achieve 100% correct selection, while UNALSE is near-perfect. Within the dynamic class, BNS and AW continue to over-select in almost all replications, mainly due to the dependence structure in the idiosyncratic component used in the DGP.

Regarding the performance of the static estimators under this DGP, extreme caution is needed when interpreting the results in Figure 3 and Table 3. By construction, a GDFM does not, in general, admit a finite static representation, as the number of static factors is infinite. For reporting purposes, in Figure 3 we instead compare the estimated number of static factors with $r = 6$, corresponding to the number implied by a dynamic representation with $q = 2$ dynamic factors. In Table 3, correct selection means that the estimator identifies $r \geq 6$ factors. This convention is admittedly arbitrary and should be interpreted with caution, since the notion of a “true” number of static factors is not well defined in this context.²⁵ For a more reliable assessment, we refer the reader to the average number of estimated factors reported in the Appendix (see Figures A7–A11).

At $\alpha = 0.8$, the GDFM group continues to dominate, though performance becomes more heterogeneous: HL and UNALSE remain highly reliable (99.6% and 88.0% correct, respectively), while ON and DDR display pronounced under-selection.

As factor strength decreases further ($\alpha = 0.65, 0.51, 0.3$), correct identification becomes rare and under-selection starts to dominate across most estimators.

5.4. Additional robustness checks

We now discuss the results of the three complementary robustness experiments described in Section 4.4.

When no factor is strong (see Appendix A1.5), the estimators based on the eigenvalue gap (ER, DDR, ON) perform considerably better than in the case where only one factor is strong. In the latter case, the second-largest eigenvalue diverges at a slower rate, making the scree-plot curvature smoother and the separation between diverging and bounded eigenvalues less clear.

²⁴ Additional evidence for $n \neq T$ is reported in Appendix Figures A5–A6, which display performance as n increases with $T = 300$ fixed, and as T increases with $n = 300$ fixed, respectively.

²⁵ Notice that, for IC_{p_2} , Table 3 displays only one dot, even though the estimator always selects $r = 8$. This outcome is purely fortuitous, as the procedure systematically tends to estimate $k_{\max}$, regardless of the DGP. See Section 5.4 for further results.

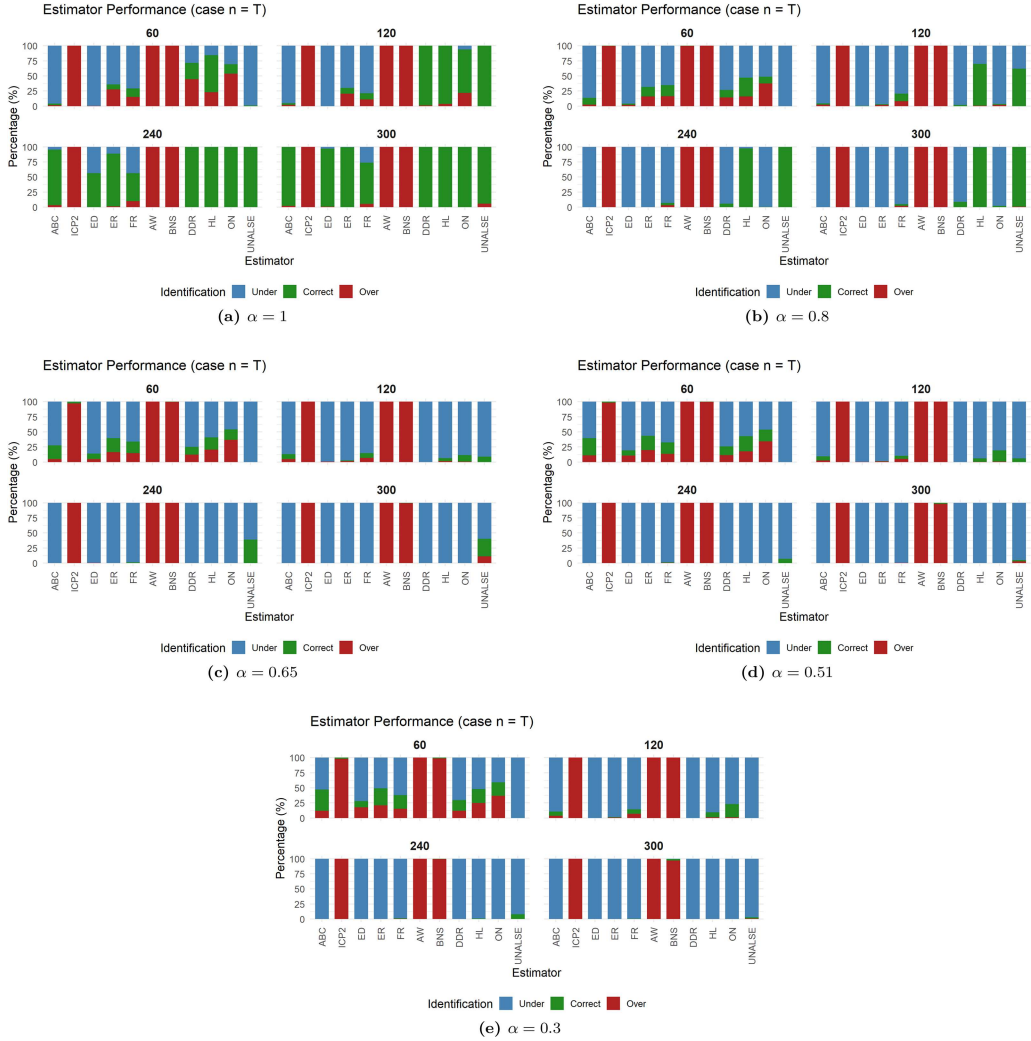


Figure 3. GDFM experiment: selection frequencies by estimator, sample size, and factor strength. Each panel reports the distribution of outcomes across estimators when $n = T \in \{60, 120, 240, 300\}$, under the DGP defined in Section 4, Equation (43), and $q = 2$ (the true r is not finite). Bars sum to 100% and are decomposed into *Correct* identification (green), *Under* (blue), and *Over* (red). Panels (a)–(e) correspond to $\alpha = 1, 0.8, 0.65, 0.51, 0.3$, respectively, ordered from strong to semi-weak factors. The vertical axis reports percentages, while the horizontal axis lists the estimators: ABC (Alessi, Barigozzi, and Capasso (2010)), ICP₂ (Bai and Ng (2002)), ED (Onatski (2010)), ER (Ahn and Horenstein (2013)), FR (Freyaldenhoven (2022)), AW (Amengual and Watson (2007)), BNS (Bai and Ng (2007)), DDR (Avarucci et al. (2026)), HL (Hallin and Liška (2007)), ON (Onatski (2009)), UNALSE (Barigozzi and Farnè (2024)). *Note:* For a GDFM, the number of static factors is not finite. Here we benchmark static estimators against $r = 6$ (from a dynamic representation with $q = 2$), an arbitrary choice that should be interpreted with caution. For a more reliable assessment of the static estimators, we refer the reader to the average number of estimated factors reported in the Appendix (see Figures A7–A11).

When $\alpha = 0.8$, these estimators perform comparably to the strong-factor case, and their performance remains acceptable for $\alpha = 0.65$, although the results depend on the DGP. For the same factor strength, under the GDFM design, ER, ABC, and ED estimate approximately three factors when only one common shock is weak, and zero when none of the factors is pervasive (for larger sample sizes). FR exhibits similar behavior when $\alpha = 0.65$, while HL also suggests no factor structure for larger values of n and T . The UNALSE estimator performs well only for some combinations of n and T . For smaller values of α , most estimators underestimate the number of factors as the sample size increases, with mean estimates close to zero. The estimators based on ICP₂, and to a less extent ON and UNALSE overestimate the number of factors and common shocks.

Table 3. GDFM experiment: correct-selection summary by estimator class, factor strength, and n – T relationship, under the generalized dynamic DGP defined in Section 4, Equation (43), and $q = 2$ (the true r is not finite).

Estimator	n vs T	Strong	Semi-strong	Semi-weak	Advantages (+)/ Disadvantages (–)
IC _{p₂}	$n = T$	●○○○	○○○○	○○○○	(–) Large estimated value is likely to be due to the non-trivially correlated idiosyncratic component.
	$n > T$	●○○○	○○○○	○○○○	(+) n, T can diverge at different rates
	$n < T$	●○○○	○○○○	○○○○	(–) The penalty term is defined up to a free parameter.
ABC	$n = T$	●●●○	○○○○	○○○○	(+) The estimated number of factors increases with the sample size.
	$n > T$	●●●○	○○○○	○○○○	(–) User-chosen settings: subsample construction and stability interval width.
	$n < T$	●●●○	○○○○	○○○○	(+) The tuning parameter helps when idiosyncratic errors are cross-correlated.
ED	$n = T$	●●●○	○○○○	○○○○	(+) The estimated number of factors increases with the sample size.
	$n > T$	●●●○	○○○○	○○○○	(–) Under-selection as factor strength decreases.
	$n < T$	●○○○	○○○○	○○○○	
ER	$n = T$	●●●○	○○○○	○○○○	(+) The estimated number of factors increases with the sample size.
	$n > T$	●●●○	○○○○	○○○○	(–) Rapid under-selection as factor strength decreases.
	$n < T$	●●●○	○○○○	○○○○	(+) No user-chosen parameters
FR	$n = T$	●●●○	○○○○	○○○○	(±) The estimated number of factors slowly increases with the sample size.
	$n > T$	●●●○	○○○○	○○○○	(+) Factors need not be strong: use eigenvector information, not only eigenvalues.
	$n < T$	●●○○	○○○○	○○○○	(–) User-chosen parameters: “concentration” measure of eigenvectors
AW	$n = T$	○○○○	○○○○	○○○○	(–) Large estimated value is likely to be due to the non-trivially correlated idiosyncratic component.
	$n > T$	○○○○	○○○○	○○○○	
	$n < T$	○○○○	○○○○	○○○○	
BNS	$n = T$	○○○○	○○○○	○○○○	(–) Estimates a number of factors close to the correctly specified dynamic case.
	$n > T$	○○○○	○○○○	●○○○	
	$n < T$	○○○○	○○○○	○○○○	
HL	$n = T$	●●●●	●●●●	○○○○	(+) Very accurate even when $\alpha = 0.8$ with large n, T . Under-selection as the factor strength decreases.
	$n > T$	●●●●	●●●●	○○○○	(+) n, T can diverge at different rates
	$n < T$	●●●●	●●●●	○○○○	(–) User-chosen settings: subsample construction and stability interval width
ON	$n = T$	●●●●	○○○○	○○○○	(+) Accurate with strong factors and large n, T . Under-selection when $\alpha \leq 0.8$.
	$n > T$	●●●●	○○○○	○○○○	(±) T is assumed to grow much faster than n but performance remains good when they are of comparable size
	$n < T$	●●●●	○○○○	○○○○	(±) Estimates may differ across frequencies.
UNALSE	$n = T$	●●●●	●●●●	○○○○	(+) Theoretically, it only requires $\alpha > 0$ for its validity.
	$n > T$	●●●●	●●●●	○○○○	(+) Retains some correct identification even when factors are semi-weak.
	$n < T$	●●●●	●●●●	●○○○	(+) No need to specify $k_{\max}$.
DDR	$n = T$	●●●●	○○○○	○○○○	(±) No need to specify $k_{\max}$.
	$n > T$	●●●●	○○○○	○○○○	(+) Accurate with strong factors. Under-selection as the factor strength decreases.
	$n < T$	●●●●	○○○○	○○○○	(+) Can be applied over a user-chosen frequency range, assuming q is constant within the range.
					(–) When factor strength $\alpha \leq 0.8$, it tends to under-estimate.

Note: The $n = T$ case corresponds to $(n, T) = (300, 300)$, the $n > T$ case to $(n, T) = (300, 180)$, and the $n < T$ case to $(n, T) = (180, 300)$. Dots indicate the average correct-selection rate within each strength bucket (4 = $\geq 75\%$, 3 = 50.1–74.9%, 2 = 25.1–50.0%, 1 = 10.1–25.0%, 0 = $\leq 10\%$). Strength buckets are defined as: Strong ($\alpha = 1$), Semi-strong ($\alpha = 0.8$), and Semi-weak ($\alpha = 0.3$). For references, we recall each estimator by citing the original paper in which it was proposed: IC_{p₂} (Bai and Ng (2002)), ABC (Alessi, Barigozzi, and Capasso (2010)), ED (Onatski (2010)), ER (Ahn and Horenstein (2013)), FR (Freyaldenhoven (2022)), AW (Amengual and Watson (2007)), BNS (Bai and Ng (2007)), HL (Hallin and Liška (2007)), ON (Onatski (2009)), UNALSE (Barigozzi and Farnè (2024)), DDR (Avarucci et al. (2026))). For a GDFM, the number of static factors is not finite. Here we benchmark static estimators against $r = 6$ (from a dynamic representation with $q = 2$), an arbitrary choice that should be interpreted with caution. For a more reliable assessment of the static estimators, we refer the reader to the average number of estimated factors reported in the Appendix (see Figures A7–A11). Notice that IC_{p₂} displays only one dot, even though the estimator always selects $r = 8$. This outcome is purely fortuitous, as the procedure systematically tends to estimate $k_{\max} = 8$, regardless of the DGP.

We next examine the sensitivity of the estimators to the choice of $k_{\max}$. The results, reported in Appendix A1.6, indicate that estimator performance is generally not highly sensitive to moderate variations in the pre-specified upper bound. Across the different configurations considered, changes in $k_{\max}$ have only a limited impact on average factor selection, suggesting that moderate misspecification of the maximum number of factors does not materially affect inference.

Finally, we consider the higher-dimensional specification with $r = q = 5$. The corresponding results under the SFM, DFM, and GDFM designs are presented in Appendix A1.7. Overall, the main conclusions remain unchanged. Although larger cross-sectional dimensions are typically required to achieve accurate factor detection in this setting, the estimators display broadly consistent behavior relative to the baseline case. This confirms that the previous findings are not driven by the low-dimensional structure of the benchmark design.

6. Summary and conclusion

This paper evaluates several widely used methodologies for determining the number of latent factors in large-dimensional factor models. Our simulation design allowed us to assess each estimator not only in the environment for which it was originally developed, but also under alternative DGPs and varying degrees of factor strength, from strong to semi-weak.

Our results show that some estimators retain good performance even when applied outside their original DGP. For example, unless the sample size is very small, the frequency-domain estimators HL, UNALSE, ON, and DDR, originally designed for the GDFM, behave similarly to the best-performing static estimators based on the covariance matrix. When the DGP is a dynamic model with $\alpha = 1$, some static criteria (e.g., ER, ED) estimate a number of factors close to q , whereas ABC and FR perform better. This pattern can be explained by the correlation structure of the idiosyncratic term, which tends to reduce the eigenvalue gap. By contrast, when the idiosyncratic term is i.i.d., all static estimators correctly recover the number of static factors implied by the static representation of the dynamic model.

When factors are strong and both n and T are sufficiently large, most procedures deliver high accuracy under their underlying DGP. The ICp2 criterion performs poorly when the idiosyncratic component is cross- and auto-correlated, often yielding estimates close to $k_{\max}$. This issue particularly affects the performance of AW, and, to a lesser extent, BNS. The improvement proposed by Alessi, Barigozzi, and Capasso (2010) proves effective, and using ABC as an initial estimator in AW and BNS improves performance in some cases.

As the factor strength weakens, performance diverges across estimator classes. Among static estimators, FR, and to some extent ED and ABC, remain comparatively robust, whereas ER deteriorates sharply. Within the GDFM class, UNALSE stands out for retaining correct identification even under semi-weak factors, especially when T is large, while HL performs well when factors are at most semi-strong.

No single estimator dominates across all scenarios. HL and UNALSE perform best overall, but rely on user-chosen parameters and, for UNALSE, on more restrictive assumptions. Scree-plot methods such as ER and DDR are not robust to weak factors, yet have the advantage of being free from tuning parameters. All estimators are broadly robust to the choice of $k_{\max}$ (when required) and to changes in the number of factors in the true DGP, although the latter case requires larger sample sizes for good performance.

Finally, while our analysis has focused on direct evaluation of performance, the evidence also suggests possible diagnostic tools for applied work. Estimators such as HL, UNALSE, ON, and DDR tend to yield similar estimates when the factors are strong, whereas discrepancies may signal weak factors. A similar comparison can be drawn in the static case by contrasting ER with FR. These considerations are beyond the scope of this paper, but could provide useful directions for empirical practice.

Acknowledgments

We are grateful to the two anonymous referees and the Associate Editor for their detailed and constructive comments, which substantially improved the exposition and overall quality of the paper. We also thank Dante Amengual, Matteo Barigozzi, and Alexei Onatski for sharing codes with us.

Author contributions

CRedit: **Marco Avarucci**: Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing; **Valentina Raponi**: Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing; **Paolo Zaffaroni**: Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Ahn, S. C., Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* 81:1203–1227.
- Alessi, L., Barigozzi, M., Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters* 80(23–24):1806–1813. doi: [10.1016/j.spl.2010.08.005](https://doi.org/10.1016/j.spl.2010.08.005)
- Amengual, D., Watson, M. W. (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics* 25(1):91–96. doi: [10.1198/073500106000000585](https://doi.org/10.1198/073500106000000585)
- Anatolyev, S., Mikusheva, A. (2022). Factor models with many assets: Strong factors, weak factors, and the two-pass procedure. *Journal of Econometrics* 229(1):103–126. doi: [10.1016/j.jeconom.2021.01.002](https://doi.org/10.1016/j.jeconom.2021.01.002)
- Angeletos, G.-M., Collard, F., Dellas, H. (2020). Business-cycle anatomy. *American Economic Review* 110(10):3030–3070. doi: [10.1257/aer.20181174](https://doi.org/10.1257/aer.20181174)
- Avarucci, M., Cavicchioli, M., Forni, M., Zaffaroni, P. (2026). The main business cycle shock (s): Frequency-band estimation of the number of dynamic factors. *Journal of the American Statistical Association* 1–13. doi: [10.1080/01621459.2025.2571246](https://doi.org/10.1080/01621459.2025.2571246)
- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* 71(1):135–171. doi: [10.1111/1468-0262.00392](https://doi.org/10.1111/1468-0262.00392)
- Bai, J., Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* 70(1):191–221. doi: [10.1111/1468-0262.00273](https://doi.org/10.1111/1468-0262.00273)
- Bai, J., Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics* 25(1):52–60. doi: [10.1198/073500106000000413](https://doi.org/10.1198/073500106000000413)
- Bailey, N., Kapetanios, G., Pesaran, M. H. (2016). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics* 31(6):929–960. doi: [10.1002/jae.2476](https://doi.org/10.1002/jae.2476)
- Bailey, N., Kapetanios, G., Pesaran, M. H. (2021). Measurement of factor strength: Theory and practice. *Journal of Applied Econometrics* 36(5):587–613. doi: [10.1002/jae.2830](https://doi.org/10.1002/jae.2830)
- Barigozzi, M., Farnè, M. (2024). An algebraic estimator for large spectral density matrices. *Journal of the American Statistical Association* 119(545):498–510. doi: [10.1080/01621459.2022.2126780](https://doi.org/10.1080/01621459.2022.2126780)
- Barigozzi, M., Hallin, M., Luciani, M., Zaffaroni, P. (2024). Inferential theory for generalized dynamic factor models. *Journal of Econometrics* 239(2):105422. doi: [10.1016/j.jeconom.2023.02.003](https://doi.org/10.1016/j.jeconom.2023.02.003)
- Caro, A., Peña, D. (2025). Selecting the number of factors in multi-variate time series. *Journal of Time Series Analysis* 46(1):113–136. doi: [10.1111/jtsa.12760](https://doi.org/10.1111/jtsa.12760)
- Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica* 51(5):1305–1323. doi: [10.2307/1912276](https://doi.org/10.2307/1912276)
- Chamberlain, G., Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* 51(5):1281–1304. doi: [10.2307/1912275](https://doi.org/10.2307/1912275)
- Chudik, A., Pesaran, M. H., Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal* 14(1):C45–C90. doi: [10.1111/j.1368-423X.2010.00330.x](https://doi.org/10.1111/j.1368-423X.2010.00330.x)
- Forni, M., Gambetti, L., Granese, A., Sala, L., Soccorsi, S., et al. (2023). An American Macroeconomic Picture: Supply and Demand Shocks in the Frequency Domain. *American Economic Journal: Macroeconomics*, 17(3), 311–341.
- Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics* 82(4):540–554. doi: [10.1162/003465300559037](https://doi.org/10.1162/003465300559037)

- Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association* 100(471):830–840. doi: [10.1198/016214504000002050](https://doi.org/10.1198/016214504000002050)
- Forni, M., Lippi, M. (2001). The generalized dynamic factor model: Representation theory. *Econometric Theory* 17(6):1113–1141. doi: [10.1017/S0266466601176048](https://doi.org/10.1017/S0266466601176048)
- Freyaldenhoven, S. (2022). Factor models with local factors—determining the number of relevant factors. *Journal of Econometrics* 229(1):80–102. doi: [10.1016/j.jeconom.2021.04.006](https://doi.org/10.1016/j.jeconom.2021.04.006)
- Geweke, J. (1977). The dynamic factor analysis of economic time series. In: D. J. Aigner and A. Goldberger (Eds.), *Latent Variables in Socio-Economic Models*, Amsterdam: North Holland, pp. 365–383.
- Giannone, D., Reichlin, L., Sala, L. (2004). Monetary policy in real time. *NBER Macroeconomics Annual* 19:161–200. doi: [10.1086/ma.19.3585335](https://doi.org/10.1086/ma.19.3585335)
- Giglio, S., Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy* 129(7):1947–1990. doi: [10.1086/714090](https://doi.org/10.1086/714090)
- Hallin, M., Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association* 102(478):603–617. doi: [10.1198/016214506000001275](https://doi.org/10.1198/016214506000001275)
- Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics* 28(3):397–409. doi: [10.1198/jbes.2009.07239](https://doi.org/10.1198/jbes.2009.07239)
- Kozak, S., Nagel, S., Santosh, S. (2018). Interpreting factor models. *The Journal of Finance* 73(3):1183–1223. doi: [10.1111/jofi.12612](https://doi.org/10.1111/jofi.12612)
- Kydland, F. E., Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica* 50(6):1345–1370. doi: [10.2307/1913386](https://doi.org/10.2307/1913386)
- Lam, C., Yao, Q., Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika* 98(4):901–918. doi: [10.1093/biomet/asr048](https://doi.org/10.1093/biomet/asr048)
- Lettau, M., Pelger, M. (2020a). Estimating latent asset-pricing factors. *Journal of Econometrics* 218(1):1–31. doi: [10.1016/j.jeconom.2019.08.012](https://doi.org/10.1016/j.jeconom.2019.08.012)
- Lettau, M., Pelger, M. (2020b). Factors that fit the time-series and cross-section of stock returns. *The Review of Financial Studies* 33(5):2274–2325. doi: [10.1093/rfs/hhaa020](https://doi.org/10.1093/rfs/hhaa020)
- Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica* 77(5):1447–1479.
- Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics* 92(4):1004–1016. doi: [10.1162/REST_a_00043](https://doi.org/10.1162/REST_a_00043)
- Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics* 168(2):244–258. doi: [10.1016/j.jeconom.2012.01.034](https://doi.org/10.1016/j.jeconom.2012.01.034)
- Onatski, A. (2024). The Hallin-Liška criterion through the lens of the random matrix theory. In: *Recent Advances in Econometrics and Statistics: Festschrift in Honour of Marc Hallin*, Cham: Springer, pp. 385–405.
- Pesaran, M. H., Smith, R. (2021). Factor strengths, pricing errors, and estimation of risk premia (No. 8947). CESifo Working Paper.
- Sargent, T. J., Sims, C. A., et al. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New Methods in Business Cycle research* 1:145–168.
- Stock, J. H., Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97(460):1167–1179. doi: [10.1198/016214502388618960](https://doi.org/10.1198/016214502388618960)
- Stock, J. H., Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20(2):147–162. doi: [10.1198/073500102317351921](https://doi.org/10.1198/073500102317351921)
- Stock, J. H., Watson, M. W. (2005). Implications of dynamic factor models for var analysis, NBER Working Paper 11467. doi: [10.3386/w11467](https://doi.org/10.3386/w11467)
- Trapani, L. (2018). A randomized sequential procedure to determine the number of factors. *Journal of the American Statistical Association* 113(523):1341–1349. doi: [10.1080/01621459.2017.1328359](https://doi.org/10.1080/01621459.2017.1328359)