

Determining the Structure of Dynamic Factor Models

Sangmyung Ha*

August 24, 2026

Abstract

We propose two procedures for determining the number of dynamic factors, extending Bai and Ng (2002) and Ahn and Horenstein (2013) to dynamic factor models where lagged factors may directly influence the observed variables. As an intermediate step, we develop a simple and computationally efficient alternating least squares algorithm that directly estimates the dynamic factors, rather than their static representations. By working with these direct estimates, our approach enables joint determination of the number of factors and the filter length. Our approach does not require the exact finite-order VAR specification maintained by Bai and Ng (2007) and Amengual and Watson (2007). We apply our procedures to estimate the number of primitive shocks in a large panel of U.S. macroeconomic time series.

JEL classification: C10, C38, C51, C55

Key words and phrases: Dynamic Factor Model, Number of Factors, Filter Length, Alternating Least Squares

*Department of Economics, Indiana University (Email: ha1@iu.edu). I am grateful to Joon Y. Park, Yoosoon Chang, Keli Xu, and Laura Liu for excellent advice and suggestions. I also thank the participants of the 2023 MEG Meeting, especially Bin Chen, who gave numerous useful suggestions, and the participants of 2024 SETA. All errors are my own.

1 Introduction

Factor models are a central tool in empirical macroeconomics and finance for summarizing co-movements in large-dimensional time series. Their appeal lies in the idea that a small number of latent factors can explain the dynamics of many observed variables.

A fundamental distinction is between *static* and *dynamic* factor models. In a static factor model, the observables x_t depend only contemporaneously on a low-dimensional vector of factors f_t , e.g., $x_t = \lambda f_t + \varepsilon_t$, whereas a dynamic factor model allows lags of the factors to affect the observables, e.g., $x_t = \sum_{k=0}^{m-1} \lambda_k f_{t-k} + \varepsilon_t$. The dynamic specification is more general and economically informative, as it captures delayed responses and propagation mechanisms that are central to macroeconomic and financial dynamics.¹

While a large body of work has developed methods for estimating the number of static factors—see, for example, Bai and Ng (2002), Onatski (2010), and Ahn and Horenstein (2013)—the literature on estimating the number of dynamic factors remains relatively limited. Notable exceptions are Bai and Ng (2007) and Amengual and Watson (2007). These papers approach the problem by constructing a static representation of the dynamic factor model. Under the assumption that the static representations of the dynamic factors follow a VAR process with a lower-dimensional innovation, Bai and Ng (2007) and Amengual and Watson (2007) develop consistent procedures for estimating the number of dynamic factors.

Although VAR models provide a flexible and widely used framework for capturing time series dynamics, the literature on the study of the static factor models shows that a VAR specification for the static factors is not necessary for either estimating the factors or determining their number. This raises the question of whether it is possible to consistently estimate the number of dynamic factors without imposing a VAR model on the factors.

This paper proposes a new approach to estimating the number of dynamic factors that does not rely on the assumption that the factors follow a VAR process. Instead, our method requires that the second moment matrix of the static representations of the dynamic factors be positive definite, a substantially weaker condition than

¹The dynamic factor model studied in this paper allows the factor process (f_t) to exhibit arbitrary serial dynamics, provided that it satisfies the Assumption 1 to be introduced in Section 2. This is in contrast to the generalized dynamic factor model of Forni et al. (2000).

imposing a parametric VAR structure. Our approach is enabled by directly estimating the dynamic factors using an alternating least squares algorithm. The procedure iteratively updates factors and loadings via simple least squares steps, making it computationally efficient and scalable to large datasets. When applied to static factor models, the alternating least squares estimator coincides with the principal components estimator commonly used in static factor models. We establish consistency under double asymptotics when both the time-series dimension and the dimension of the observed variable diverge, obtaining the standard convergence rates typically associated with the estimation of static factors. By leveraging consistent estimators of the dynamic factors, we propose two procedures that jointly estimate the number of dynamic factors and the filter length, defined as the length of the horizon over which current and lagged factors exert a direct effect on the observed variables.

Our approach extends the criteria of Ahn and Horenstein (2013) and Bai and Ng (2002) to dynamic factor models. Under general Assumptions to be introduced in Sections 2 and 4, the number of dynamic factors and the filter length of the dynamic factor model are jointly identified asymptotically independently of any parametric specification for the stochastic process governing the dynamic factors. In particular, if the dynamic factors follow a VAR(p) process, the identification of the filter length is invariant to the choice of the lag order p . We also address the issue of VAR lag order selection using estimated factors when the dynamic factors follow a VAR process.

The eigenvalue ratio test of Ahn and Horenstein (2013) identifies the number of factors by detecting the point at which the ratio of consecutive singular values of the data matrix diverges. We extend this idea to dynamic factor models by introducing a notion of 'dynamic' singular values, defined as the spectral norm of the residual matrix after removing a given number of estimated dynamic factors. Building on this concept, we propose the dynamic singular value ratio test, which generalizes the eigenvalue ratio test of Ahn and Horenstein (2013) and coincides with it when applied to a static factor model. Moreover, the proposed framework can be used to determine the filter length of the dynamic factor model by examining the behavior of ratios of consecutive dynamic singular values as the filter length varies for a fixed number of dynamic factors.

The information criterion of Bai and Ng (2002) selects the number of static factors by augmenting the residual variance with a penalty term that increases with the number of static factors. Our second procedure extends this approach to dynamic

factor models by employing a penalty term that depends on both the number of dynamic factors and the filter length. When we restrict the filter length to one, the proposed criterion coincides with that of Bai and Ng (2002), up to a constant factor of two in the penalty term.

1.1 Related literature and contributions of this paper

There is a large literature on dynamic factor models, which can be broadly divided into two classes. The first class comprises generalized dynamic factor models with infinite filter length, which do not admit a finite-dimensional static representation. The second class consists of models with a finite filter length, for which a static representation exists. Our work falls into the second category: the dynamic factor models with a finite filter length. What distinguishes our work from many existing ones is that we work directly with the dynamic factors themselves, rather than the static representations.

The generalized dynamic factor model with infinite filter length was introduced by Forni et al. (2000), who proposed estimation based on frequency-domain principal components. Subsequent work has developed one-sided representations, estimation methods, and asymptotic theory for these models: Forni et al. (2015), Forni et al. (2017), and Barigozzi et al. (2024). Methods for determining the number of dynamic factors in the generalized dynamic factor framework have also been proposed by Hallin and Liška (2007) and Onatski (2010). However, the greater generality afforded by infinite-length filters typically requires the dynamic factors to be serially uncorrelated processes for identification. We focus our attention on dynamic factor models with finite-length filter in this paper.

A separate strand of the literature studies the static representations of dynamic factor models. Bai (2003) and Bai and Li (2016) provide inferential theory for static factors estimated by principal components and quasi-maximum likelihood, respectively. Procedures for determining the number of static factors include information criteria (Bai and Ng, 2002) and eigenvalue-ratio tests (Ahn and Horenstein, 2013), among others. Despite this extensive literature, comparatively little attention has been paid to the direct estimation of dynamic factors. An exception is Bai and Wang (2015), who studies Bayesian estimation of dynamic factor models. This paper is most closely related to Bai and Ng (2007) and Amengual and Watson (2007), who

study the estimation of the number of dynamic factors in models with finite filter length.

To summarize, our contributions are as follows.

First, we generalize the procedures of Bai and Ng (2002) and Ahn and Horenstein (2013) to estimate the number of dynamic factors in dynamic factor models. Our approach is more robust than those of Bai and Ng (2007) and Amengual and Watson (2007), as it does not require imposing a VAR structure on the static factors.

Second, our procedure provides a consistent estimator of the filter length, jointly with the number of dynamic factors, which, to the best of our knowledge, is the first in the literature. Identifying the extent to which primitive shocks directly affect the economy is of independent interest. For instance, using the FRED–MD dataset of McCracken and Ng (2015) over the sample period March 1973 to March 2025, the filter length is estimated to be two.²

Third, we examine the finite-sample performance of our selection procedures and show that they outperform those of Bai and Ng (2007) and Amengual and Watson (2007) across a range of designs in determining q .

Finally, we establish the consistency of the proposed alternating least squares estimator in estimating the dynamic factors. Furthermore, when the dynamic factors follow a VAR(p) process, we show that the lag order selection can also be done using the information criteria methods, with unobserved factors replaced with their estimates.

The remainder of the paper is organized as follows. Section 2 presents the model and discusses representations of the dynamic factors. Section 3 describes the alternating least squares estimation procedure and the consistency of the proposed method. Section 4 develops two methods for determining the number of dynamic factors and the filter length. Section 5 reports simulation evidence on the finite-sample performance of the proposed procedures. Section 6 provides an empirical application. Section 7 concludes. All technical proofs are collected in the Appendix.

Notations. We denote by $\|\cdot\|$ the Euclidean norm when applied to vectors and the associated spectral (operator) norm when applied to matrices. The Frobenius norm is denoted by $\|\cdot\|_F$. For a symmetric matrix A , the notation $A > 0$ and $A \geq 0$

²The associated number of dynamic factors is found to be four, which increases to seven after the onset of COVID.

indicates that A is positive definite and positive semi-definite, respectively. We denote $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$ for full column matrix A . We write $\lceil x \rceil$ for the smallest integer greater than or equal to x and $\lfloor x \rfloor$ for the largest integer less than or equal to x . Throughout, q denotes the number of dynamic factors, m the filter length, and $r = qm$ the corresponding number of static factors. Finally, we denote $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{N}_0$, and $\mathbb{R}$ the set of integers, natural numbers, nonnegative integers, and real numbers, respectively.

2 Model, Assumptions, and Representations

2.1 Model

We denote the pair (q, m) representing the number of dynamic factors and the filter length as the structure of a dynamic factor model. For a given structure (q, m) , we consider the finite-order dynamic factor model

$$x_{it} = \chi_{it} + \varepsilon_{it}, \quad \chi_{it} = \sum_{k=0}^{m-1} \lambda'_{ik} f_{t-k}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where $f_t \in \mathbb{R}^q$ is the vector of dynamic factors, $\lambda_{ik} \in \mathbb{R}^q$ is the loading vector associated with lag k , and m is the filter length. We assume that the factor process is defined for all $t \in \mathbb{Z}$, so the required presample lags are well defined.

When $m = 1$, (1) coincides with a static factor model. If the finite-order restriction is relaxed and the loading filter is allowed to be infinite under appropriate summability conditions, the model is related to the generalized dynamic factor framework of Forni et al. (2000). We restrict attention to the finite-filter case $m < \infty$, which can be viewed as a finite-order dynamic extension of the approximate static factor models studied in Bai (2003) and Stock and Watson (2002).

Let $r = qm$ and define the stacked static factor and loading vectors by

$$g_t = \begin{pmatrix} f'_t & f'_{t-1} & \cdots & f'_{t-m+1} \end{pmatrix}' \in \mathbb{R}^r, \quad \gamma_i = \begin{pmatrix} \lambda'_{i0} & \lambda'_{i1} & \cdots & \lambda'_{i,m-1} \end{pmatrix}' \in \mathbb{R}^r.$$

Then (1) has the algebraically equivalent static representation

$$x_{it} = \gamma'_i g_t + \varepsilon_{it}. \quad (2)$$

For each $k = 0, \dots, m-1$, define the $N \times q$ loading matrix

$$\lambda_k = \begin{pmatrix} \lambda'_{1k} \\ \vdots \\ \lambda'_{Nk} \end{pmatrix},$$

and let

$$x_t = \begin{pmatrix} x_{1t} \\ \vdots \\ x_{Nt} \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Nt} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \lambda_0 & \lambda_1 & \cdots & \lambda_{m-1} \end{pmatrix} \in \mathbb{R}^{N \times r}.$$

Equivalently, the i -th row of Γ is γ'_i . The vector form of the model is therefore

$$x_t = \Gamma g_t + \varepsilon_t. \tag{3}$$

Stacking the observations over time, define

$$X = \begin{pmatrix} x'_1 \\ \vdots \\ x'_T \end{pmatrix} \in \mathbb{R}^{T \times N}, \quad E = \begin{pmatrix} \varepsilon'_1 \\ \vdots \\ \varepsilon'_T \end{pmatrix} \in \mathbb{R}^{T \times N}, \quad G = \begin{pmatrix} g'_1 \\ \vdots \\ g'_T \end{pmatrix} \in \mathbb{R}^{T \times r}.$$

The model can then be written in static matrix form as

$$X = G\Gamma' + E. \tag{4}$$

For $k = 0, \dots, m-1$, let

$$F_{-k} = \begin{pmatrix} f'_{1-k} \\ \vdots \\ f'_{T-k} \end{pmatrix} \in \mathbb{R}^{T \times q}$$

collect the k -th lag of the dynamic factors. It follows that

$$G = \begin{pmatrix} F_0 & F_{-1} & \cdots & F_{-(m-1)} \end{pmatrix},$$

and hence the dynamic matrix representation is

$$X = \sum_{k=0}^{m-1} F_{-k} \lambda'_k + E. \quad (5)$$

For convenience, we write $F = F_0$ for the $T \times q$ matrix collecting the contemporaneous dynamic factors.

Although (4) has the algebraic form of a static factor model with $r = qm$ factor coordinates, the columns of G are not unrestricted. In particular, the m blocks of g_t are lagged copies of the same q -dimensional process. If g_{jt} denotes the j th q -dimensional block of g_t , then

$$g_{1t} = f_t, \quad g_{j+1,t} = g_{j,t-1}, \quad j = 1, \dots, m-1.$$

Relative to an unrestricted static factor model with r factors, the dynamic parameterization therefore uses only q underlying factor series and imposes cross-time shift restrictions on their stacked representation. This restriction is the source of its parsimony; the number of loading coefficients remains Nqm in both representations.

The common component may admit multiple dynamic representations. To distinguish the true structure from generic candidate structures, we let q_0 denote the smallest factor dimension among all exact representations of the common component. Among representations attaining q_0 , let m_0 denote the smallest filter length, and define $r_0 = q_0 m_0$. We refer to (q_0, m_0) as the true structure. We use (q, m) and $r = qm$ to denote a generic candidate structure. The existence and characterization of alternative observationally equivalent representations are discussed in Section 2.3.

2.2 Assumptions

Throughout, q_0 and m_0 are fixed, $r_0 = q_0 m_0$, and $N, T \rightarrow \infty$ jointly. Recall that $G \in \mathbb{R}^{T \times r_0}$ and $\Gamma \in \mathbb{R}^{N \times r_0}$ denote the stacked factor and loading matrices associated with the true structure (q_0, m_0) . The following two assumptions are used to establish consistency of the estimated factor and loading spaces. Additional assumptions used to establish consistency of the procedures for determining (q_0, m_0) are introduced in Section 4.

Assumption 1 (Nondegeneracy and pervasiveness). *The stacked factor and loading*

matrices associated with the true structure satisfy

$$\frac{1}{T}G'G \rightarrow_p \Sigma_g, \quad \frac{1}{N}\Gamma'\Gamma \rightarrow_p \Sigma_\gamma,$$

as $N, T \rightarrow \infty$, where Σ_g and Σ_γ are finite $r_0 \times r_0$ positive-definite matrices.

Assumption 2 (Weak idiosyncratic component). *The idiosyncratic-error matrix satisfies*

$$\|E\| = O_p\left(\sqrt{\max\{N, T\}}\right).$$

The two parts of Assumption 1 have distinct roles. The first requires the lag-stacked factor vector

$$g_t = \begin{pmatrix} f'_t & f'_{t-1} & \cdots & f'_{t-m_0+1} \end{pmatrix}'$$

to have a nonsingular asymptotic second-moment matrix. The second is the usual pervasiveness condition, requiring each true static-factor direction to have nonvanishing cross-sectional strength.

A convenient sufficient condition for the factor-side restriction is that (f_t) is a covariance-stationary and ergodic process with finite second moments whose spectral density matrix is positive definite on a set of frequencies with positive Lebesgue measure. Under this condition,

$$\mathbb{E}(g_t g'_t) > 0,$$

and an ergodic law of large numbers yields

$$\frac{1}{T}G'G \rightarrow_p \mathbb{E}(g_t g'_t).$$

This condition includes, for example, stable and invertible VARMA processes driven by innovations with positive-definite covariance. In particular, a stable VAR driven by full-rank innovations has a spectral density matrix that is positive definite at every frequency; see, for example, (Brockwell and Davis, 1991, Sections 11.3 and 11.8). The factor-side condition in Assumption 1 is therefore weaker than requiring a finite-order VAR specification.³ Analogous nondegeneracy and pervasiveness conditions are maintained in Bai and Ng (2007) and Amengual and Watson (2007) when estimating the number of static factors. Unlike their fixed-order VAR formulations, however,

³The factor-side condition in Assumption 1 is discussed in greater detail in Supplemental Appendix B.

Assumption 1 does not require the factors to possess an exact finite-order autoregressive representation. For example, a generic invertible VMA(1) process has a VAR(∞) representation and can satisfy Assumption 1, even though it does not generally admit an exact finite-order VAR representation. Bai and Ng (2007) discuss finite-order VAR approximations to such processes but note that their fixed-order theory does not directly cover the MA case.

Assumption 2 places a global bound on the strength of the idiosyncratic component. It permits weak cross-sectional and serial dependence in the errors but rules out an additional idiosyncratic component with singular values of the same order as those generated by the strong common component. Spectral-norm and related eigenvalue restrictions of this type are standard in the factor-model literature; see, for example, Onatski (2010), Ahn and Horenstein (2013), Moon and Weidner (2017), and Bai and Ng (2023).

2.3 Equivalent Dynamic Representations

Recall that the true common component is

$$\chi_t = \sum_{k=0}^{m_0-1} \lambda_k f_{t-k}, \quad f_t \in \mathbb{R}^{q_0}.$$

Throughout this subsection, we restrict attention to exact finite-order causal linear reparameterizations of the true factor process. Specifically, an alternative representation with structure (q, m) is of the form

$$h_t = H(L)f_t = \sum_{\ell=0}^d H_\ell f_{t-\ell}, \quad \chi_t = \sum_{k=0}^{m-1} \tilde{\Gamma}_k h_{t-k},$$

for some finite d , where $H_\ell \in \mathbb{R}^{q \times q_0}$ and $\tilde{\Gamma}_k \in \mathbb{R}^{N \times q}$. We interpret m as the maximum candidate filter length, so some $\tilde{\Gamma}_k$ may be zero, and the coordinates of h_t may be redundant.

The true representation is minimal within this class: no exact representation exists with factor dimension $q < q_0$, and, among representations with $q = q_0$, no exact representation exists with filter length $m < m_0$. Two dynamic factor representations are called equivalent if they generate the same common component χ_t for every t .

When combined with the same idiosyncratic component ε_t , the resulting models for x_t are observationally equivalent.

The true representation (q_0, m_0) always admits the static structure $(r_0, 1)$ obtained by stacking the current and lagged values of f_t . There may exist other admissible dynamic representations that are observationally equivalent. To illustrate, consider a model with structure $(q_0, m_0) = (2, 3)$:

$$\chi_t = \lambda_0 f_t + \lambda_1 f_{t-1} + \lambda_2 f_{t-2}, \quad f_t \in \mathbb{R}^2.$$

Its static representation is

$$\chi_t = \begin{pmatrix} \lambda_0 & \lambda_1 & \lambda_2 \end{pmatrix} g_t, \quad g_t = \begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \end{pmatrix} \in \mathbb{R}^6. \quad (6)$$

Thus, $(q, m) = (6, 1)$ is an admissible structure. The same common component also admits a representation with structure $(q, m) = (3, 2)$. Define

$$h_t = \begin{pmatrix} f_{t,1} \\ f_{t,2} + f_{t-1,1} \\ f_{t-1,2} \end{pmatrix} \in \mathbb{R}^3$$

and let

$$\tilde{\Gamma}_0 = \begin{pmatrix} \lambda_{0,1} & \lambda_{0,2} & \lambda_{1,2} - \lambda_{2,1} \end{pmatrix}, \quad \tilde{\Gamma}_1 = \begin{pmatrix} \lambda_{1,1} - \lambda_{0,2} & \lambda_{2,1} & \lambda_{2,2} \end{pmatrix},$$

where $\lambda_{k,j}$ denotes the j -th column of λ_k . Then

$$\chi_t = \tilde{\Gamma}_0 h_t + \tilde{\Gamma}_1 h_{t-1}. \quad (7)$$

Hence, $(q, m) = (3, 2)$ is also admissible.

The example illustrates the tradeoff between the number of factor coordinates and the filter length. Reducing the filter length below m_0 requires additional factor coordinates, while increasing the factor dimension weakly reduces the minimum admissible filter length. The following proposition shows that admissibility is governed by the factor dimension q and the comparison between $r = qm$ and the true static

rank $r_0 = q_0 m_0$.

Proposition 1 (Characterization of admissible dynamic structures). *Let $r_0 = q_0 m_0$. Every pair $(q, m) \in \mathbb{N}^2$ satisfying*

$$q \geq q_0 \quad \text{and} \quad qm \geq r_0$$

is an admissible dynamic structure.

Conversely, every admissible dynamic structure satisfies $q \geq q_0$. Moreover, under Assumption 1, an exact representation with $qm < r_0$ is ruled out with probability approaching one as $N, T \rightarrow \infty$. Thus, admissible dynamic structures are asymptotically characterized by

$$q \geq q_0 \quad \text{and} \quad qm \geq q_0 m_0.$$

Proposition 1 implies that structures satisfying $qm = r_0$ lie on the boundary of the admissible region and may provide nonredundant representations of the common component. However, the true structure (q_0, m_0) is the most parsimonious representation because its factor dimension is the smallest. Structures satisfying $qm > r_0$ are also admissible, but necessarily contain redundant static factor coordinates or loading directions under Assumption 1.

Corollary 1 (Lower bounds on admissible structures). *Suppose that $q_0, m_0 \geq 1$ and let $r_0 = q_0 m_0$. Under Assumption 1, the minimum admissible factor dimension for a fixed $m \geq 1$ is*

$$q_{\min}(m) = \max \left\{ q_0, \left\lceil \frac{r_0}{m} \right\rceil \right\} = \begin{cases} \left\lceil \frac{q_0 m_0}{m} \right\rceil, & 1 \leq m < m_0, \\ q_0, & m \geq m_0. \end{cases}$$

For a fixed $q \geq q_0$, the minimum admissible filter length is

$$m_{\min}(q) = \left\lceil \frac{r_0}{q} \right\rceil = \left\lceil \frac{q_0 m_0}{q} \right\rceil.$$

These lower bounds are asymptotically sharp.

For $1 \leq m < m_0$, the minimum number of additional factor coordinates required

to reduce the candidate filter length to m is

$$q_{\min}(m) - q_0 = \left\lceil \frac{q_0 m_0}{m} \right\rceil - q_0.$$

Moreover,

$$m_{\min}(q_0) = m_0,$$

and $m_{\min}(q)$ is weakly decreasing in q . Because of the ceiling operator, increasing q need not strictly reduce the minimum filter length at every value of q .

Let $q_{\max} \geq q_0$ and $m_{\max} \geq m_0$ be given, and let

$$S = \{0, \dots, q_{\max}\} \times \{0, \dots, m_{\max}\}.$$

Consider the disjoint partition of S given by

$$\begin{aligned} S_1 &= \{(q, m) \in S : q \geq q_0, \quad qm \geq r_0\}, \\ S_2 &= \{(q, m) \in S : qm < r_0\}, \\ S_3 &= \{(q, m) \in S : q < q_0, \quad qm \geq r_0\}. \end{aligned} \tag{8}$$

Proposition 1 shows that every structure in S_1 admits an exact representation of the true common component. Structures in S_2 have static dimension smaller than the true static rank $r_0 = q_0 m_0$. Under Assumption 1, the matrix of true common components has rank r_0 with probability approaching one, whereas a representation with structure (q, m) has rank at most qm . Hence, structures in S_2 cannot reproduce the true common component with probability approaching one. Structures in S_3 have sufficient static dimension, $qm \geq q_0 m_0$, but fewer than q_0 dynamic factor coordinates. The static-rank argument therefore does not distinguish these structures from the true model. Their misspecification instead follows from the minimality of q_0 .

3 Estimation

In the dynamic factor model,

$$x_t = \sum_{k=0}^{m-1} \lambda_k f_{t-k} + \varepsilon_t, \quad t = 1, \dots, T,$$

the dynamic factors (f_t) are indexed from $t = 2 - m$ to $t = T$ to account for the lagged factors entering the equation. We jointly estimate $(f_t)_{t=2-m}^T$ and $(\lambda_k)_{k=0}^{m-1}$ by minimizing the least squares objective

$$\frac{1}{NT} \left\| X - \sum_{k=0}^{m-1} F_{-k} \lambda'_k \right\|_F^2 = \frac{1}{NT} \sum_{t=1}^T \left\| x_t - \sum_{k=0}^{m-1} \lambda_k f_{t-k} \right\|^2 \quad (9)$$

over (f_t) and (λ_k) . The first-order conditions for the factor loadings (λ_j) are

$$\sum_{k=0}^{m-1} \hat{\lambda}_k \hat{F}'_{-k} \hat{F}_{-j} = X' \hat{F}_{-j}, \quad j = 0, \dots, m-1. \quad (10)$$

The first-order conditions for (f_t) are

$$\sum_{j=\max(0,1-t)}^{\min(m-1,T-t)} \lambda'_j \left(x_{t+j} - \sum_{k=0}^{m-1} \lambda_k f_{t+j-k} \right) = 0 \quad (11)$$

for $t = 2 - m, \dots, T$.

We minimize (9) using an alternating least squares procedure. We start with an initial guess for $(\hat{\lambda}_k)$. Given $(\hat{\lambda}_k)$, we may update (f_t) by solving (11). Given $(\hat{f}_t)$, we may update (λ_k) by solving (10). We repeat the updating steps until convergence. The first-order conditions (10) define a multivariate regression with rank qm , which has a closed-form solution. The first-order condition (11) can be efficiently solved using the discrete Fourier transform, requiring inversion of only $q \times q$ matrices.⁴ Since each step minimizes a least squares objective, the algorithm guarantees a monotone decrease of the objective function. The minimization problem is not convex, and the alternating least squares algorithm generally does not converge to the global minimum. We try multiple initial values and choose the solution that obtains the smallest objective function value in practice.

3.1 Consistency

Before introducing the selection criteria for the structure (q, m) of the dynamic factor models, we first show that the estimated dynamic factors are consistent.

⁴Derivation of the discrete Fourier transform needed to solve the first order condition is given in the Supplemental Materials C.

Proposition 2 (Factor-space consistency under admissible structures). *Let Assumptions 1 and 2 hold. For any fixed $(q, m) \in S_1$, let $(\hat{G}, \hat{\Gamma})$ be a global minimizer of the least-squares objective (9) over representations with structure (q, m) . Then*

$$\frac{1}{\sqrt{T}} \|M_{\hat{G}} G\|_F = O_p \left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}} \right),$$

where G is the true stacked factor matrix.

Remark Proposition 2 establishes one-sided factor-space consistency for every fixed admissible structure $(q, m) \in S_1$. Thus, the estimated stacked factor space asymptotically contains the true factor space. When $qm = r_0 = q_0 m_0$, this result yields the usual consistency of the estimated factor space. When $qm > r_0$, the estimated space may contain additional redundant directions.

Under the restriction $m = 1$, the admissible structures are

$$S_1 \cap \{(q, m) : m = 1\} = \{(q, 1) : q \geq r_0\}.$$

For this unrestricted static-factor problem, Theorem 1 of Bai and Ng (2002) establishes average consistency, up to a possibly rank-deficient transformation, of the rescaled principal-component factor estimator for any fixed number of estimated factors. Proposition 2 provides a dynamic analogue of this static result by allowing every admissible structure $(q, m) \in S_1$, including structures with $m > 1$.

For $m > 1$, the dynamically constrained least-squares estimator imposes the lag-stacking restriction

$$\hat{g}_t = \begin{pmatrix} \hat{f}_t \\ \hat{f}_{t-1} \\ \vdots \\ \hat{f}_{t-m+1} \end{pmatrix}.$$

Equivalently, the m blocks of $\hat{g}_t$ are shifted copies of the same estimated q -dimensional factor process. An unrestricted principal-component estimator of the static factors does not impose this temporal restriction, so its estimated factor coordinates need not satisfy the displayed relation.

This distinction is relevant for determining the dynamic structure of the model. Unrestricted principal components identifies the static factor space and its dimension

$r_0 = q_0 m_0$, but it does not separately impose a decomposition into the dynamic factor dimension q_0 and the filter length m_0 . The dynamically constrained estimator preserves the temporal structure used by the subsequent procedures for determining these two dimensions.

4 Determining the Number of Dynamic Factors and Filter Length

4.1 Selection Based on the Dynamic Singular Value Ratio

Throughout this section, $\sigma_j(X)$ denotes the j -th largest singular value of X . We define a dynamic analogue of the singular-value approximation error by imposing the lag-stacking restrictions of a dynamic factor model on the low-rank approximating matrix. Formally, let $\delta_{q+1,m}(X)$ denote the $(q+1)$ -th dynamic singular value of a matrix X with filter length m , defined as

$$\delta_{q+1,m}(X) = \inf_{C \in \mathcal{C}_{q,m}} \|X - C\|,$$

where

$$\mathcal{C}_{q,m} = \left\{ \sum_{k=0}^{m-1} A_{-k} B'_k : B_k \in \mathbb{R}^{N \times q}, A_{-k} = \begin{pmatrix} a'_{1-k} \\ \vdots \\ a'_{T-k} \end{pmatrix}, a_{2-m}, \dots, a_T \in \mathbb{R}^q \right\}.$$

with $q \geq 0$ and $m \geq 0$. For $q = 0$ or $m = 0$, we define $\mathcal{C}_{q,m} = \{0\}$ so that $\delta_{1,m}(X)$ or $\delta_{q+1,0}(X)$ denotes spectral norm of X .

When $m = 1$, $\delta_{q+1,1}(X)$ coincides with the usual $(q+1)$ -th singular value of X because $\mathcal{C}_{q,1} = \{C : \text{rank}(C) \leq q\}$. For general m , every $C \in \mathcal{C}_{q,m}$ has rank at most qm . Therefore,

$$\delta_{q+1,m}(X) \geq \delta_{qm+1,1}(X) \tag{12}$$

The inequality reflects the additional temporal restrictions imposed by the dynamic approximation relative to an unrestricted rank- qm static approximation. More gen-

erally, given (q_0, m_0) and its associated set S_1 , we have

$$\delta_{q+1,m}(X) \leq \delta_{q_0+1,m_0}(X)$$

for any integer $(q, m) \in S_1$ because $\mathcal{C}_{q_0,m_0} \subseteq \mathcal{C}_{q,m}$ due to Proposition 1.

Let

$$C_0 = \sum_{k=0}^{m_0-1} F_{-k} \lambda'_k = G\Gamma'$$

denote the $T \times N$ matrix of true common components.

Assumption 3. *For any $\epsilon > 0$, there exists a constant $c_0 > 0$ such that*

$$\mathbb{P}\left(\frac{1}{\sqrt{N+T}} \delta_{k_{\max}+1,1}(E) < c_0\right) \leq \epsilon$$

for all sufficiently large N and T , where $k_{\max}$ is fixed.

Assumption 4. *For any $\epsilon > 0$, there exists a constant $c_2 > 0$ such that*

$$\mathbb{P}\left(\min_{\substack{0 \leq q < q_0 \\ 1 \leq m \leq m_{\max}}} \frac{1}{\sqrt{NT}} \delta_{q+1,m}(C_0) < c_2\right) \leq \epsilon$$

and for all sufficiently large N and T , where $m_{\max} \geq m_0$ is fixed.

Assumption 3 requires that the leading singular values of $\frac{1}{\sqrt{\max(N,T)}}E$ be bounded away from zero in probability. Under conditions such as those in Ahn and Horenstein (2013), a positive fraction of the singular values of $\frac{E}{\sqrt{\max\{N,T\}}}$ remain bounded away from zero. Assumption 3, which requires this property only for a fixed number of leading singular values, is correspondingly weaker.

Assumption 4 imposes a corresponding lower bound on the dynamic approximation error of the true common component under every underfactored structure. Part of this condition follows directly from Assumption 1. For every (q, m) satisfying $0 \leq q < q_0$ and $1 \leq m \leq m_0$, we have

$$qm \leq (q_0 - 1)m_0 < r_0.$$

Therefore, by (12),

$$\delta_{q+1,m}(C_0) \geq \sigma_{qm+1}(C_0) \geq \sigma_{r_0}(C_0).$$

The smallest nonzero singular value of $C_0 = G\Gamma'$ satisfies

$$\frac{\sigma_{r_0}^2(C_0)}{NT} \geq \lambda_{\min}\left(\frac{G'G}{T}\right) \lambda_{\min}\left(\frac{\Gamma\Gamma'}{N}\right).$$

Assumption 1 implies that the right-hand side is bounded away from zero in probability. It follows that

$$\min_{\substack{0 \leq q < q_0 \\ 1 \leq m \leq m_0}} \frac{1}{\sqrt{NT}} \delta_{q+1,m}(C_0)$$

is bounded away from zero in probability.

Thus, Assumption 1 already provides the required separation for underfactored structures whose candidate filter length does not exceed m_0 . For $m > m_0$, however, it is possible to have

$$qm \geq r_0$$

even when $q < q_0$. In this case, static rank alone does not provide a lower bound on the approximation error. Assumption 4 strengthens the minimality of q_0 by requiring underfactored representations to remain uniformly separated from the true common component at the $\sqrt{NT}$ scale over the fixed candidate range

$$1 \leq m \leq m_{\max}.$$

Supplemental Appendix B provides sufficient conditions for Assumption 4. In particular, under the assumption that the optimized sample mean-squared approximation criterion converges to its population counterpart, a stable and invertible VARMA factor process driven by innovations with positive-definite covariance satisfies the spectral component of these sufficient conditions. The resulting frequency-domain condition is related to the dynamic-eigenvalue restrictions used in the generalized dynamic factor literature; see, for example, Forni et al. (2000), Forni et al. (2015), and Forni et al. (2017).

We construct ratio-based selection criteria constructed with residuals obtained from the Frobenius norm least-squares estimator. For each (q, m) with $q, m \geq 1$, let $(\hat{F}_{-k}^{(q,m)}, \hat{\lambda}_k^{(q,m)})_{k=0}^{m-1}$ form any global minimizer of the least-squares objective (9).

Define the fitted common component by

$$\hat{C}_{q,m} = \sum_{k=0}^{m-1} \hat{F}_{-k}^{(q,m)} \hat{\lambda}_k^{(q,m)'}$$

For $q = 0$ or $m = 0$, set $\hat{C}_{q,m} = 0$. We define $\hat{\delta}_{q+1,m}(X)$ as

$$\hat{\delta}_{q+1,m}(X) = \left\| X - \hat{C}_{q,m} \right\|, \quad q, m \in \mathbb{N}_0.$$

It is important to note that the minimizer of the Frobenius norm generally differs from the minimizer of the spectral norm. When $m = 1$, the Frobenius norm minimizer also minimizes the spectral norm, i.e., gives a best rank- q approximation solution with $\hat{\delta}_{q+1,1}(X) = \delta_{q+1,1}(X)$. For $m > 1$, $\hat{\delta}_{q+1,m}(X) \neq \delta_{q+1,m}(X)$ in general. While dynamic singular values can in principle be computed by directly minimizing the spectral norm, we rely on the Frobenius-norm solution for computational tractability.

It can be shown that under Assumptions 1, 2, 3, and 4, there exists $q_{\max}$ and $m_{\max}$ such that for all $0 \leq q \leq q_{\max}$ and $0 \leq m \leq m_{\max}$, $\frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{N+T}}$ is bounded above and away from zero in probability for $(q, m) \in S_1$, whereas $\frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{NT}}$ is bounded above and away from zero in probability for $(q, m) \in S_2 \cup S_3$. This behavior of $\hat{\delta}_{q+1,m}(X)$ motivates the dynamic singular value ratio tests given in Theorem 1 and 2 below.

In Theorem 1, we fix m and estimate the number of dynamic factors using a ratio-type statistic based on adjacent estimated dynamic singular values.

Theorem 1 (Consistency of the factor-dimension estimator). *Suppose that $q_0, m_0 \geq 1$ and that Assumptions 1, 2, 3, and 4 hold. Let $q_{\max}$, $m_{\max}$, and $k_{\max}$ be fixed positive integers satisfying*

$$q_0 \leq q_{\max}, \quad m_0 \leq m_{\max}, \quad q_0 m_0 + q_{\max} m_{\max} \leq k_{\max}.$$

Fix $m \in \{1, \dots, m_{\max}\}$ and define $q_m^* = \max \{q_0, \lceil \frac{q_0 m_0}{m} \rceil\}$. Suppose that $q_m^* \leq q_{\max}$. For $q = 1, \dots, q_{\max}$, define

$$DR_{q,m}(X) = \frac{\hat{\delta}_{q,m}(X)}{\hat{\delta}_{q+1,m}(X)}.$$

If a denominator is zero, set the corresponding ratio equal to $+\infty$. Using the smallest

maximizer as a deterministic tie-breaking rule, define

$$\hat{q}(m) = \min \arg \max_{1 \leq q \leq q_{\max}} DR_{q,m}(X).$$

Then

$$\mathbb{P} \{ \hat{q}(m) = q_m^* \} \rightarrow 1$$

as $N, T \rightarrow \infty$.

For a fixed m , the quantity q_m^* is the smallest factor dimension for which (q, m) belongs to S_1 . Hence, at $q = q_m^*$, the numerator of $DR_{q,m}(X)$ corresponds to the inadmissible structure $(q_m^* - 1, m)$ and is of order $\sqrt{NT}$, whereas the denominator corresponds to the admissible structure (q_m^*, m) and is of order $\sqrt{N + T}$. The ratio therefore diverges at q_m^* and remains stochastically bounded at every other candidate value.

When $m \geq m_0$, the admissibility boundary satisfies $q_m^* = q_0$, so the estimator consistently identifies the true number of dynamic factors. When $m < m_0$, it selects the minimum number of factor coordinates required to represent the true common component with candidate filter length m , as characterized by Proposition 1.

When $m = 1$,

$$DR_{q,1}(X) = \frac{\sigma_q(X)}{\sigma_{q+1}(X)}.$$

The eigenvalue-ratio statistic of Ahn and Horenstein (2013) is based on

$$\frac{\sigma_q^2(X)}{\sigma_{q+1}^2(X)} = DR_{q,1}^2(X).$$

The two criteria therefore have the same maximizer. In this case, Theorem 1 consistently estimates the true static dimension

$$r_0 = q_0 m_0.$$

The theorem extends this static ratio criterion by allowing the candidate filter length to exceed one and by imposing the lag-stacking restrictions of a dynamic factor model.

We next fix the factor dimension and estimate the minimum admissible filter length.

Theorem 2 (Consistency of the filter-length estimator). *Suppose that $q_0, m_0 \geq 1$ and that Assumptions 1, 2, and 3 hold. Let $q_{\max}$, $m_{\max}$, and $k_{\max}$ be fixed positive integers satisfying*

$$q_0 \leq q_{\max}, \quad m_0 \leq m_{\max}, \quad q_0 m_0 + q_{\max} m_{\max} \leq k_{\max}.$$

Fix $q \in \{q_0, \dots, q_{\max}\}$ and define $m_q^* = \left\lceil \frac{q_0 m_0}{q} \right\rceil$. For $m = 1, \dots, m_{\max}$, define

$$MR_{q,m}(X) = \frac{\hat{\delta}_{q+1,m-1}(X)}{\hat{\delta}_{q+1,m}(X)}.$$

If a denominator is zero, set the corresponding ratio equal to $+\infty$. Define

$$\hat{m}(q) = \min_{1 \leq m \leq m_{\max}} \arg \max MR_{q,m}(X).$$

Then

$$\mathbb{P} \{ \hat{m}(q) = m_q^* \} \rightarrow 1$$

as $N, T \rightarrow \infty$.

For a fixed $q \geq q_0$, the quantity m_q^* is the smallest filter length for which (q, m) belongs to S_1 . At $m = m_q^*$, the numerator of $MR_{q,m}(X)$ corresponds to the inadmissible structure $(q, m_q^* - 1)$, while the denominator corresponds to the admissible structure (q, m_q^*) . The ratio therefore diverges only at m_q^* .

Theorem 2 consistently estimates the admissibility boundary m_q^* for a fixed factor dimension $q \geq q_0$. In particular, it consistently estimates the true filter length m_0 when $q = q_0$. When $q > q_0$, it selects the smallest candidate filter length required to represent the true common component using q factor coordinates.

To recover (q_0, m_0) using Theorem 1 and 2, we may start with any m sufficiently large so that $m \geq m_0$. Using $m \geq m_0$, we may recover q_0 using Theorem 1. Then, using Theorem 2, we may recover m_0 using $q = q_0$.

4.2 Selection Based on Information Criteria

Let $V(q, m)$ denote the mean squared errors when the structure (q, m) is used to estimate a dynamic factor model, i.e.,

$$V(q, m) = \frac{1}{NT} \left\| X - \hat{C}_{q,m} \right\|_F^2,$$

where $\hat{C}_{q,m} = \sum_{k=0}^{m-1} \hat{F}_{-k}^{(q,m)} \hat{\lambda}_k^{(q,m)'} is the least squares estimator obtained under the structure (q, m) .$

Following Bai and Ng (2002), we augment the residual variance with a penalty term. The key is to construct a penalty sequence that vanishes asymptotically, yet dominates the difference in the residual sum of squares between the true model and any overparameterized alternative. In this spirit, we define the following two information criteria:

$$\begin{aligned} PC(q, m) &= V(q, m) + (r + q)g(N, T), \\ DC(q, m) &= \frac{1}{NT} \hat{\delta}_{q+1,m}^2(X) + (r + q)g(N, T), \end{aligned}$$

where $r = qm$ and $g(N, T)$ is a vanishing penalty function. Compared to Bai and Ng (2002), the main differences are: (i) model fit is evaluated after estimating the dynamic factor model, not its static representation, and (ii) the penalty accounts for both the number of static factors $r = qm$, and the number of dynamic factors q .

Theorem 3. *Suppose Assumptions 1, 2, and 4 hold. Let $q_{max} \geq q_0$ be a fixed integer and let $m_{max} \geq m_0$ be given as in Assumption 4. Let $\hat{q}$ and $\hat{m}$ be given by*

$$(\hat{q}, \hat{m}) = \underset{0 \leq q \leq q_{max}, 0 \leq m \leq m_{max}}{\operatorname{argmin}} PC(q, m)$$

or

$$(\hat{q}, \hat{m}) = \underset{0 \leq q \leq q_{max}, 0 \leq m \leq m_{max}}{\operatorname{argmin}} DC(q, m).$$

If $g(N, T) \rightarrow 0$ and $\left(\frac{NT}{N+T}\right) g(N, T) \rightarrow \infty$ as $N, T \rightarrow \infty$, we have

$$\mathbb{P}\{(\hat{q}, \hat{m}) = (q_0, m_0)\} \rightarrow 1$$

as $N, T \rightarrow \infty$.

The key to determining the true structure (q_0, m_0) is penalizing both the number of static factors $r = qm$ and the number of dynamic factors q . Penalizing only the number of static factors would select the static representation of the dynamic factor model, since among the structures (q, m) with the same static factors, the structure $(qm, 1)$ has the smallest squared error. Including q as a separate term in the penalty ensures recovery of the most parsimonious dynamic structure. In fact, we may use any penalty function of the form $\varphi(q, m)g(N, T)$ as long as $g(N, T) \rightarrow 0$, $(\frac{NT}{N+T})g(N, T) \rightarrow \infty$ as $N, T \rightarrow \infty$, and

$$\varphi(q_0, m_0) \leq \varphi(q, m)$$

for all $(q, m) \in S_1$ with strict inequality when $(q, m) \neq (q_0, m_0)$. For example, any function of the form $\varphi(q, m) = \alpha qm + \beta q$ with $\alpha, \beta > 0$ can also be used to consistently recover (q_0, m_0) .

Theorem 3 shows that Assumptions 1, 2, and 4 suffice to asymptotically identify (q, m) . The result does not require a particular finite-dimensional parametric model, such as a finite-order VAR, for the factor process. Assumptions 1 and 2 are commonly used to identify the number of static factors, e.g., Chamberlain and Rothschild (1983), Bai and Ng (2002), Bai and Ng (2023). To recover the number of dynamic factors, along with the corresponding filter length, the only additional assumption required is the Assumption 4.

Note that when $q_0 = 0$ or $m_0 = 0$, so that there does not exist a common component with factor structure, the information criterion consistently recovers $(q_0, m_0) = (0, 0)$.

Assumption 5 (Nondegenerate average idiosyncratic variance). *There exists a constant*

$$\sigma_\varepsilon^2 \in (0, \infty)$$

such that

$$\frac{1}{NT} \|E\|_F^2 \rightarrow_p \sigma_\varepsilon^2.$$

Corollary 2. *Suppose Assumptions 1, 2, 4, and 5 hold. Let $q_{\max}$ be a fixed integer and let $m_{\max}$ be given as in Assumption 4. Let*

$$(\hat{q}, \hat{m}) = \underset{0 \leq q \leq q_{\max}, 0 \leq m \leq m_{\max}}{\operatorname{argmin}} IC(q, m),$$

where

$$IC(q, m) = \log V(q, m) + (r + q)g(N, T).$$

If $g(N, T) \rightarrow 0$ and $\left(\frac{NT}{N+T}\right)g(N, T) \rightarrow \infty$, then

$$\mathbb{P}\{(\hat{q}, \hat{m}) = (q_0, m_0)\} \rightarrow 1.$$

As in Bai and Ng (2002), we may use three candidate penalty terms:

$$\begin{aligned} g_1(N, T) &= \left(\frac{N+T}{NT}\right) \log \left(\frac{NT}{N+T}\right) \\ g_2(N, T) &= \left(\frac{N+T}{NT}\right) \log (\min(N, T)) \\ g_3(N, T) &= \left(\frac{1}{\min(N, T)}\right) \log (\min(N, T)). \end{aligned}$$

We also scale the penalty term with $\hat{\sigma}_{PC}^2 = V(q_{max}, m_{max})$ and $\hat{d}_{DC}^2 = \frac{1}{NT}\hat{\delta}_{1,0}^2(X)$ for PC and DC respectively. Under Assumptions 2 and 5, we may show that

$$\hat{\sigma}_{PC}^2 \rightarrow_p \sigma_\varepsilon^2$$

as $N, T \rightarrow \infty$. Also, we have

$$\left| \frac{\|X\|}{\sqrt{NT}} - \frac{\|C_0\|}{\sqrt{NT}} \right| \leq \frac{\|E\|}{\sqrt{NT}}$$

so that

$$\hat{d}_{DC} \rightarrow_p \lambda_{max}(\Sigma_\gamma^{1/2} \Sigma_g \Sigma_\gamma^{1/2})$$

as $N, T \rightarrow \infty$ when $q_0, m_0 \geq 1$. The DC criterion with the scale $\hat{d}_{DC}^2 = \frac{1}{NT}\hat{\delta}_{1,0}^2(X)$ does not depend on (q_{max}, m_{max}) through the scale term. However, it may not recover the true structure under $(q_0, m_0) = (0, 0)$. Under the null structure, this scale is only of order $\frac{N+T}{NT}$, making the effective penalty too small to dominate the noise-scale reduction from fitting spurious factors.

Both PC and IC can be viewed as dynamic extensions of the original criteria in Bai and Ng (2002): if $m = 1$, they reduce numerically to the criteria in Bai and Ng (2002), up to a multiplicative factor of two in the penalty term.

4.3 Determining the Lag Order of the VAR Model

Although our method does not require the dynamic factors (f_t) to follow a VAR process, it may still be useful in practice to model their dynamics using a VAR model. For instance, one may be interested in estimating the dynamic responses of macroeconomic variables to structural shocks identified using residuals of a VAR model for (f_t) , as in Forni et al. (2009).

In this context, modeling (f_t) directly using a VAR is substantially more parsimonious than working with the static representation (g_t) . As shown by Bai and Ng (2007) and Forni et al. (2009), if the dynamic factors (f_t) follow a VAR(p) process, then the associated static factors (g_t) follow a VAR($\tilde{p}$) process, where $\tilde{p} = 1$ if $m \geq p$ and $\tilde{p} = p - m + 1$ if $m < p$.

When $m \geq p$, estimation based on the static representation requires fitting a VAR(1) in dimension qm , involving q^2m^2 parameters. In contrast, direct estimation of (f_t) requires fitting a VAR(p) in dimension q , involving only q^2p parameters. Since $m \geq p$, and $m > 1$ in genuine dynamic factor models, the reduction in dimensionality can be substantial. When $m < p$, the estimation of the static representation requires $(p - m + 1)q^2m^2$ parameters, which again strictly exceed pq^2 whenever $m > 1$. This highlights substantial reductions in dimension and parameter count by directly modeling the dynamic factors.

In this section, we assume (f_t) follows a VAR(p) process, and show that the lag order p can be selected consistently using the estimated factors.

Assumption 6 (Nondegeneracy of the augmented factor stack). *Let*

$$J_0 = \begin{pmatrix} I_{T-1} & 0_{(T-1) \times 1} \end{pmatrix}, \quad J_1 = \begin{pmatrix} 0_{(T-1) \times 1} & I_{T-1} \end{pmatrix},$$

so that J_0 and J_1 select the first and last $T - 1$ rows, respectively. Define

$$\overline{G}^{(m+1)} = J_1 \begin{pmatrix} F_0 & F_{-1} & \cdots & F_{-m} \end{pmatrix} \in \mathbb{R}^{(T-1) \times q(m+1)}.$$

Then

$$\frac{1}{T-1} \overline{G}^{(m+1)'} \overline{G}^{(m+1)} \rightarrow_p \Sigma_g^{(m+1)} > 0.$$

Assumption 7 (VAR lag-order regularity). *The dynamic factor process follows a*

stable VAR(p_0) model,

$$f_t = \sum_{j=1}^{p_0} A_j f_{t-j} + u_t, \quad 0 \leq p_0 \leq p_{\max},$$

where q , p_0 , and $p_{\max}$ are fixed. The order p_0 is minimal, and the innovation sequence satisfies

$$\mathbb{E}(u_t \mid \mathcal{F}_{t-1}) = 0, \quad \mathbb{E}(u_t u_t') = \Sigma_u > 0,$$

together with uniformly bounded fourth moments and the weak-dependence conditions needed for the usual fixed-dimensional VAR laws of large numbers and central limit theorems.

For every $\ell = 1, \dots, p_{\max}$, the population covariance matrix of

$$z_{\ell,t} = \begin{pmatrix} f'_{t-1} & \cdots & f'_{t-\ell} \end{pmatrix}'$$

is positive definite. Moreover, for every $\ell = p_0 + 1, \dots, p_{\max}$, the population covariance matrix of the additional lag regressors after projecting them on $f_{t-1}, \dots, f_{t-p_0}$ is positive definite.

Assumption 6 rules out an exact linear relation among $m + 1$ consecutive factor vectors. It is used to identify a common $q \times q$ rotation across the m lag blocks of the static factor representation.

Proposition 2 only gives consistency of static representations g_t of dynamic factors f_t , i.e., consistency up to a $qm \times qm$ dimensional invertible matrix. Assumption 6 is used to identify and consistently estimate the dynamic factors f_t up to a $q \times q$ dimensional matrix. Bai and Wang (2014) uses a similar assumption for identification. After establishing consistency of the dynamic factors f_t in Proposition 3, we use the result to establish consistency of the VAR lag order selection in Theorem 4.

Proposition 3 (Consistency of the dynamic factors and lag loadings). *Let $(q, m) = (q_0, m_0)$ be the true structure, with q and m fixed. Let*

$$\hat{G} = \begin{pmatrix} \hat{F}_0 & \hat{F}_{-1} & \cdots & \hat{F}_{-(m-1)} \end{pmatrix}, \quad \hat{\Gamma} = \begin{pmatrix} \hat{\lambda}_0 & \hat{\lambda}_1 & \cdots & \hat{\lambda}_{m-1} \end{pmatrix}$$

be constructed from a global minimizer of (9). Choose a normalized representative

satisfying

$$\frac{1}{N} \sum_{k=0}^{m-1} \hat{\lambda}'_k \hat{\lambda}_k = I_q, \quad \frac{1}{N} \sum_{k=0}^{m-1} \lambda'_k \lambda_k = I_q$$

Suppose that Assumptions 1, 2, and 6 hold and

$$\lambda_{\max} \left(\frac{1}{T} \hat{G}' \hat{G} \right) = O_p(1).$$

Define

$$H'_f = \left(\sum_{k=0}^{m-1} \lambda'_k \hat{\lambda}_k \right) \left(\sum_{k=0}^{m-1} \hat{\lambda}'_k \hat{\lambda}_k \right)^{-1}.$$

Then H_f is nonsingular with probability approaching one,

$$\|H_f\| + \|H_f^{-1}\| = O_p(1),$$

and

$$\frac{1}{\sqrt{T}} \left\| \hat{F} - F H'_f \right\|_F = O_p \left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}} \right).$$

Moreover, for each $k = 0, \dots, m-1$,

$$\frac{1}{\sqrt{N}} \left\| \hat{\lambda}_k - \lambda_k H_f^{-1} \right\|_F = O_p \left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}} \right).$$

Suppose we use the common effective sample

$$t = p_{\max} + 1, \dots, T$$

for all candidate lag orders in the VAR regression. Define

$$Y = \begin{pmatrix} f'_{p_{\max}+1} \\ \vdots \\ f'_T \end{pmatrix} \in \mathbb{R}^{n \times q}, \quad \hat{Y} = \begin{pmatrix} \hat{f}'_{p_{\max}+1} \\ \vdots \\ \hat{f}'_T \end{pmatrix}.$$

For $\ell \geq 1$, define

$$Z_\ell = \begin{pmatrix} F_{-1} & \cdots & F_{-\ell} \end{pmatrix} \in \mathbb{R}^{n \times q\ell}, \quad \hat{Z}_\ell = \begin{pmatrix} \hat{F}_{-1} & \cdots & \hat{F}_{-\ell} \end{pmatrix},$$

where all matrices are restricted to the common effective sample. For $\ell = 0$, the regressor matrix is empty. For $\ell \geq 1$, let

$$\hat{B}'_\ell = (\hat{Z}'_\ell \hat{Z}_\ell)^{-1} \hat{Z}'_\ell \hat{Y},$$

and define

$$\hat{\Sigma}_\ell = \frac{1}{T - p_{\max}} \left(\hat{Y} - \hat{Z}_\ell \hat{B}'_\ell \right)' \left(\hat{Y} - \hat{Z}_\ell \hat{B}'_\ell \right) = \frac{1}{T - p_{\max}} \hat{Y}' M_{\hat{Z}_\ell} \hat{Y}.$$

For $\ell = 0$, set

$$\hat{\Sigma}_0 = \frac{1}{T - p_{\max}} \hat{Y}' \hat{Y}.$$

Define

$$\text{PIC}_{\text{VAR}}(\ell) = \log(\det(\hat{\Sigma}_\ell)) + q^2 \ell g(N, T), \quad \ell = 0, \dots, p_{\max}.$$

Theorem 4 (Consistency of log-determinant VAR lag-order selection using estimated factors). *Suppose that the true dynamic structure (q_0, m_0) is used to estimate the factors and that Assumptions 1, 2, 6, and 7 hold. Let $g(N, T)$ satisfy*

$$g(N, T) \rightarrow 0, \quad \frac{NT}{N + T} g(N, T) \rightarrow \infty$$

and define

$$\hat{p} = \min \arg \min_{0 \leq \ell \leq p_{\max}} \text{PIC}_{\text{VAR}}(\ell).$$

Then

$$\mathbb{P}\{\hat{p} = p_0\} \rightarrow 1$$

as $N, T \rightarrow \infty$. No restriction on the relative rates at which N and T diverge is required.

Remark. Existing structural-factor and FAVAR applications generally select the VAR lag order using conventional AIC or BIC applied to estimated factors. Forni and Gambetti (2010), for example, choose two lags as the midpoint between the orders selected by AIC and BIC applied to estimated factors.

The dependence of the penalty on N arises because the VAR is estimated using generated regressors. If the factors were observed, the reduction in the log-determinant criterion from adding finitely many irrelevant lags would be $O_p(T^{-1})$, and an ordinary

time-series BIC penalty would suffice. With estimated factors, additional lags may also fit factor-estimation error. Under Proposition 3, the resulting contribution to the nested criterion comparison is bounded by

$$O_p\left(\frac{1}{N} + \frac{1}{T}\right).$$

The panel-adaptive penalty in Theorem 4 dominates this rate without requiring a restriction on the relative growth of N and T .

For the conventional time-series BIC penalty $g_{\text{BIC}}(N, T) = \frac{\log T}{T}$ to yield consistent model selection in Theorem 4, we need an additional rate condition

$$\frac{T}{N \log T} \rightarrow 0$$

to hold so that the squared factor estimation error $O_p\left(\frac{1}{N} + \frac{1}{T}\right)$ is negligible compared to the penalty term.

5 Simulations

We conduct Monte Carlo simulations of dynamic factor models to evaluate the finite-sample performance of estimators for the number of dynamic factors and the filter length. Specifically, we simulate the model

$$x_t = \sum_{k=0}^{m-1} \lambda_k f_{t-k} + \varepsilon_t,$$

where the idiosyncratic errors satisfy

$$\varepsilon_{it} = \sqrt{\frac{\theta(1-\rho^2)}{1+2J\beta^2}} e_{it}, \quad e_{it} = \rho e_{i,t-1} + v_{it} + \sum_{1 \leq |j| \leq J} \beta v_{i-j,t}.$$

The innovations v_{it} and the factor loadings λ_k are drawn independently from a standard normal distribution. This data-generating process is similar to those considered in Bai and Ng (2002) and Ahn and Horenstein (2013), except that we allow the filter

length m to exceed one. The common factors f_t follow a VARMA(1, 1) process,

$$f_t = Af_{t-1} + u_t + \Theta u_{t-1},$$

where u_t follows independent standard normal under the following four specifications:

1. $\rho = \beta = J = 0$ and $A = \Theta = 0$.
2. $\rho = 0.3$, $\beta = 0.1$, $J = 10$, and $A = \Theta = 0$.
3. $\rho = \beta = J = 0$, $A = \text{Diag}(0.7, 0.5, 0.3)$, and $\Theta = 0$.
4. $\rho = \beta = J = 0$, $A = 0$, and $\Theta = \text{Diag}(0.7, 0.5, 0.3)$.

By construction, the variance of ε_{it} equals θ . The variance of the common component $\sum_{k=0}^{m-1} \lambda_k f_{t-k}$ is given by $m \cdot \text{tr}(\Sigma_f)$, where $\Sigma_f = \mathbb{E}[f_t f_t']$. We therefore set $\theta = m \cdot \text{tr}(\Sigma_f)$, ensuring that the expected signal-to-noise ratio equals one in all specifications.

The first specification corresponds to an exact dynamic factor model, in which both the dynamic factors and idiosyncratic errors are serially uncorrelated. The second specification represents an approximate dynamic factor model, allowing for serial and cross-sectional dependence in the idiosyncratic errors. The third and fourth specifications examine settings in which the dynamic factors follow VAR(1) and VMA(1) processes, respectively.

5.1 Selection Based on the Dynamic Singular Value Ratio

We use the four specifications described in the previous section with $(q_0, m_0) = (3, 3)$.

		DGP1			DGP2			DGP3			DGP4		
N	T	$m = 2$	$m = 3$	$m = 4$	$m = 2$	$m = 3$	$m = 4$	$m = 2$	$m = 3$	$m = 4$	$m = 2$	$m = 3$	$m = 4$
100	100	0.92	0.99	0.96	0.17	0.28	0.23	0.00	0.79	0.47	0.12	0.93	0.81
100	200	0.99	0.98	1.00	0.17	0.40	0.24	0.04	0.98	0.86	0.36	1.00	1.00
100	300	1.00	1.00	1.00	0.20	0.55	0.29	0.09	1.00	0.98	0.58	1.00	1.00
200	100	1.00	1.00	1.00	0.52	0.88	0.61	0.10	0.88	0.66	0.43	1.00	0.88
200	200	1.00	1.00	1.00	0.89	0.99	0.93	0.43	1.00	0.99	0.90	1.00	1.00
200	300	1.00	1.00	1.00	0.93	1.00	0.95	0.82	1.00	1.00	0.98	1.00	1.00
300	100	1.00	0.99	1.00	0.85	0.99	0.86	0.18	1.00	0.89	0.72	1.00	0.98
300	200	1.00	1.00	1.00	0.99	0.99	1.00	0.81	1.00	1.00	0.99	1.00	1.00
300	300	1.00	1.00	1.00	1.00	1.00	1.00	0.93	1.00	1.00	1.00	1.00	1.00

Table 1: Monte Carlo frequency with which $\hat{q}(m) = q_m^*$, for each fixed candidate filter length m .

Table 1 reports the Monte Carlo frequency with which $\hat{q}(m) = q_m^*$. For $m = 2$, we have $q_m^* = \lceil \frac{q_0 m_0}{m} \rceil = 5$. For $m = 3, 4$, we have $q_m^* = q_0 = 3$. The estimator performs very well in DGP1. DGP3 and DGP4 also perform strongly when $m \geq m_0$ as the panel dimensions increase. DGP2 is more challenging, particularly when N is small, reflecting the serial and cross-sectional dependence in the idiosyncratic component. The short-filter case $m = 2$ is also difficult under persistent factor dynamics.

		DGP1			DGP2			DGP3			DGP4		
N	T	$q = 3$	$q = 4$	$q = 5$	$q = 3$	$q = 4$	$q = 5$	$q = 3$	$q = 4$	$q = 5$	$q = 3$	$q = 4$	$q = 5$
100	100	0.99	0.54	0.97	0.16	0.05	0.30	0.02	0.00	0.02	0.49	0.00	0.41
100	200	0.98	0.94	1.00	0.17	0.07	0.28	0.10	0.00	0.04	0.88	0.01	0.76
100	300	1.00	0.99	1.00	0.29	0.06	0.38	0.26	0.00	0.07	0.97	0.14	0.95
200	100	1.00	0.98	1.00	0.73	0.20	0.81	0.21	0.00	0.13	0.91	0.05	0.79
200	200	1.00	1.00	1.00	1.00	0.77	0.99	0.49	0.00	0.26	1.00	0.39	0.98
200	300	1.00	1.00	1.00	1.00	0.89	1.00	0.76	0.00	0.56	1.00	0.79	1.00
300	100	0.98	1.00	1.00	0.97	0.69	0.95	0.29	0.00	0.15	0.97	0.15	0.94
300	200	1.00	1.00	1.00	0.99	0.98	1.00	0.83	0.02	0.61	1.00	0.70	1.00
300	300	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.12	0.90	1.00	0.91	1.00

Table 2: Monte Carlo frequency with which $\hat{m}(q) = m_q^*$, for each fixed factor dimension q .

In Table 2, we use DR test to estimate the filter length m given q . For $q = 3, 4$, we have $m_q^* = \lceil \frac{q_0 m_0}{q} \rceil = 3$, whereas for $q = 5$, we have $m_q^* = \lceil \frac{q_0 m_0}{q} \rceil = 2$.

The particularly difficult case is $q = 4$ in DGP3. The candidate structure $(4, 2)$ has static dimension eight, only one below the true static rank $q_0 m_0 = 9$. When the factors follow a VAR(1) process, factor persistence can make the final required static direction relatively weak. Consequently, the dynamic approximation error under $(4, 2)$

can be small in finite samples, limiting the separation between $m = 2$ and $m = 3$.

5.2 Selection Based on Information Criteria

We see finite sample performance of the PC, DC, IC criteria in selecting (q, m) . We compare the results with the two existing methods given by Bai and Ng (2007) and Amengual and Watson (2007)⁵.

In Table 3 and 4, PC, DC, and IC denote the information criterion methods proposed in this paper. BN and AW denote methods proposed in Bai and Ng (2007) and Amengual and Watson (2007), respectively. To illustrate the attainable performance of each procedure, the tables report the best-performing penalty variant among g_1, g_2, g_3 for each method. In particular, PC and AW use $g_1(N, T) = (\frac{N+T}{NT}) \log(\frac{NT}{N+T})$, BN uses $g_2(N, T) = (\frac{N+T}{NT}) \log(\min(N, T))$, and DC, IC use $g_3(N, T) = (\frac{1}{\min(N, T)}) \log(\min(N, T))$. Results for selecting filter length are missing in BN and AW, as they only select the number of dynamic factors.

		<i>DGP1</i>								<i>DGP2</i>							
		<i>PC</i>		<i>DC</i>		<i>IC</i>		<i>BN</i>	<i>AW</i>	<i>PC</i>		<i>DC</i>		<i>IC</i>		<i>BN</i>	<i>AW</i>
N	T	<i>q</i>	<i>m</i>	<i>q</i>	<i>m</i>	<i>q</i>	<i>m</i>	<i>q</i>	<i>q</i>	<i>q</i>	<i>m</i>	<i>q</i>	<i>m</i>	<i>q</i>	<i>m</i>	<i>q</i>	<i>q</i>
100	100	0.97	0.99	0.98	0.98	0.97	0.98	0.34	0.23	0.47	0.50	0.15	0.18	0.26	0.26	0.04	0.45
100	200	0.99	0.99	0.98	0.99	0.98	0.99	0.55	0.89	0.91	0.91	0.18	0.18	0.93	0.93	0.00	0.01
100	300	1.00	1.00	1.00	1.00	1.00	1.00	0.89	0.99	0.98	0.98	0.20	0.22	0.99	0.99	0.00	0.00
200	100	1.00	1.00	1.00	1.00	1.00	1.00	0.82	0.98	0.97	0.99	0.57	0.66	0.98	0.99	0.45	0.94
200	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.98	0.75	0.75	0.33	0.99
200	300	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.01	0.68
300	100	0.99	1.00	0.98	0.98	0.99	1.00	0.97	0.98	1.00	0.99	0.91	0.94	1.00	1.00	0.98	1.00
300	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.99	0.93	1.00
300	300	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.79	1.00

Table 3: Marginal Monte Carlo selection frequencies under DGP1 and DGP2. For PC, DC, and IC, the columns labeled q and m report $\Pr(\hat{q} = q_0)$ and $\Pr(\hat{m} = m_0)$, respectively. BN and AW estimate only q_0 .

Table 3 reports marginal selection frequencies under DGP1 and DGP2. All three proposed selectors perform very well in DGP1, even in the smallest panel, and their

⁵In both Bai and Ng (2007) and Amengual and Watson (2007), we used IC test whenever it was needed to test for the number of static factors. When the lag order of the VAR(p) model needed to be specified, we used $p = 1$ for the first three specifications and $p = 2$ for the fourth specification. For Bai and Ng (2007), we used the default parameters $(\delta, m) = (0.1, 1)$ with covariance matrix method. For Amengual and Watson (2007), we used the residuals from direct regression, i.e., $\hat{Y}_t^B$ in Amengual and Watson (2007).

factor-dimension selection frequencies are substantially higher than those of BN and AW in the smaller samples. DGP2 is more challenging because the idiosyncratic component is serially and locally cross-sectionally dependent. PC is the most robust criterion in this design and improves rapidly with the panel dimensions. DC and IC are more conservative in small panels but also become accurate as both N and T increase.

		<i>DGP3</i>								<i>DGP4</i>							
		<i>PC</i>		<i>DC</i>		<i>IC</i>		<i>BN</i>	<i>AW</i>	<i>PC</i>		<i>DC</i>		<i>IC</i>		<i>BN</i>	<i>AW</i>
N	T	q	m	q	m	q	m	q	q	q	m	q	m	q	m	q	q
100	100	0.90	0.34	0.31	0.08	0.91	0.49	0.51	0.01	0.94	0.82	0.72	0.55	0.99	0.92	0.62	0.02
100	200	0.96	0.18	0.39	0.11	0.92	0.21	0.10	0.58	0.98	0.66	0.94	0.86	0.97	0.71	0.67	0.80
100	300	0.99	0.07	0.51	0.22	0.93	0.02	0.08	0.88	1.00	0.70	1.00	0.98	0.99	0.63	0.84	0.99
200	100	0.96	0.82	0.41	0.29	0.82	0.61	0.60	0.44	1.00	0.97	0.87	0.82	0.97	0.94	0.96	0.68
200	200	1.00	0.97	0.99	0.98	1.00	1.00	0.32	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99
200	300	1.00	0.99	1.00	1.00	1.00	1.00	0.54	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
300	100	0.99	0.98	0.34	0.18	0.86	0.70	0.86	0.88	1.00	1.00	0.95	0.93	0.98	0.99	1.00	0.94
300	200	1.00	1.00	1.00	1.00	1.00	1.00	0.87	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
300	300	1.00	1.00	1.00	1.00	1.00	1.00	0.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4: Marginal Monte Carlo selection frequencies under DGP3 and DGP4. For PC, DC, and IC, the columns labeled q and m report $\Pr(\hat{q} = q_0)$ and $\Pr(\hat{m} = m_0)$, respectively. BN and AW estimate only q_0 .

Table 4 reports results under serially dependent factor processes. For the common task of estimating q_0 , PC has a higher or equal selection frequency than BN and AW in all reported designs. DC and IC also perform strongly in many moderately large panels, although DC does not uniformly dominate the benchmark procedures in unbalanced panels. Filter-length selection is generally more difficult than factor-dimension selection, particularly for PC under DGP3 when N is small relative to T . Because BN and AW do not estimate m_0 , comparisons with these methods concern only the selection of q_0 .

6 Empirical Analysis: Shocks in the United States

In this section, we estimate the factor structure (q, m) of dynamic factor models estimated using a large panel of U.S. macroeconomic variables. The Supplemental Materials D additionally reports the estimated impulse responses of these variables to monetary policy shocks identified as in Eichenbaum and Evans (1995) within the

dynamic factor model framework.

The empirical analysis is based on the FRED–MD dataset of McCracken and Ng (2015), covering the period from March 1959 to March 2025. The sample comprises $T = 793$ observations on $N = 126$ macroeconomic time series. All variables are transformed in accordance with the procedures recommended by McCracken and Ng (2015).

To determine the factor structure, we implement the PC2, DC2, and IC2 criteria using a rolling window of fixed length equal to 10 years. The evaluation period begins in March 1969 and extends through March 2025. For the March of each year in this period, the criteria are computed using the information set consisting of the preceding 10 years of observations.

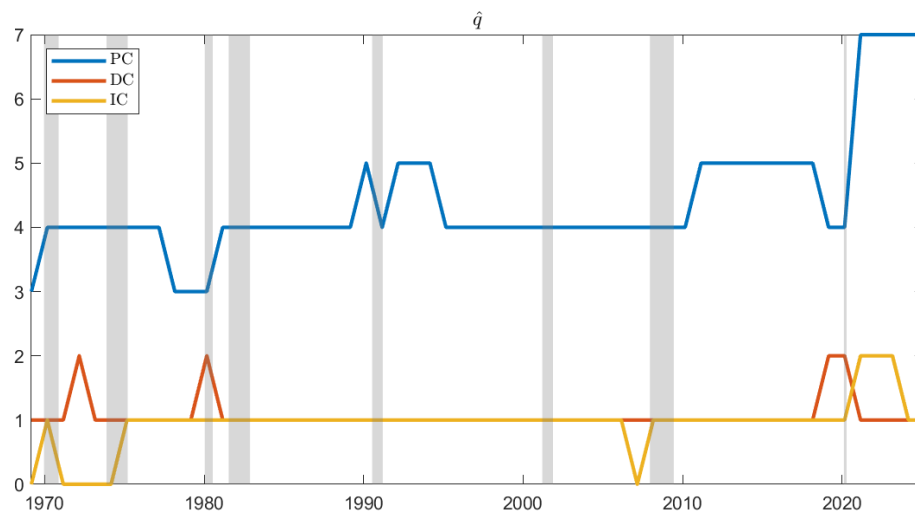


Figure 1: Estimated q using PC2, DC2, and IC2.

The PC2 criterion tends to select a larger number of factors, whereas the DC2 and IC2 criteria typically yield a more parsimonious specification. Based on the PC2 results, we find evidence in favor of four dynamic factors prior to the COVID-19 period. In contrast, the post-COVID-19 period is characterized by an increase in the estimated number of factors, with the criterion indicating the presence of seven dynamic factors.

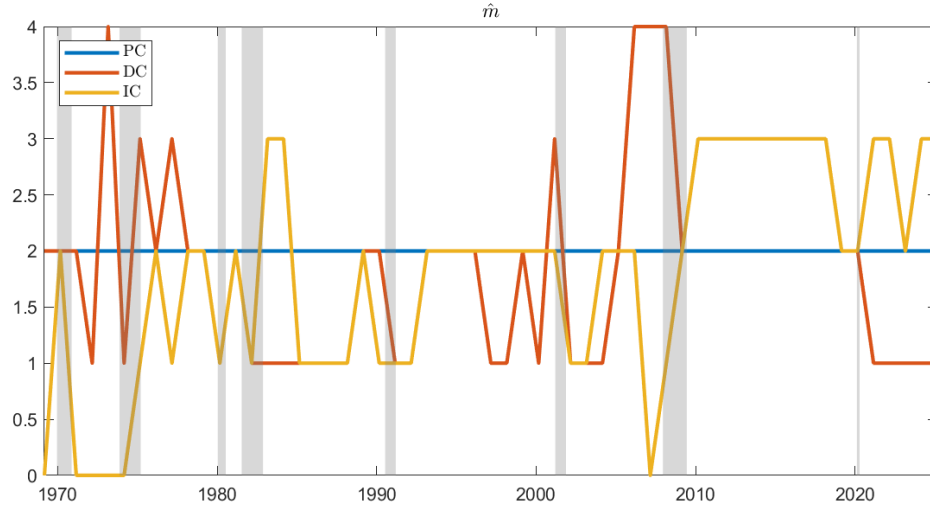


Figure 2: Estimated m Using PC2, DC2, and IC2.

The PC2 criterion consistently selects $m = 2$ across the sample. The DC2 and IC2 criteria fluctuates around selecting $m = 1$ to $m = 3$.

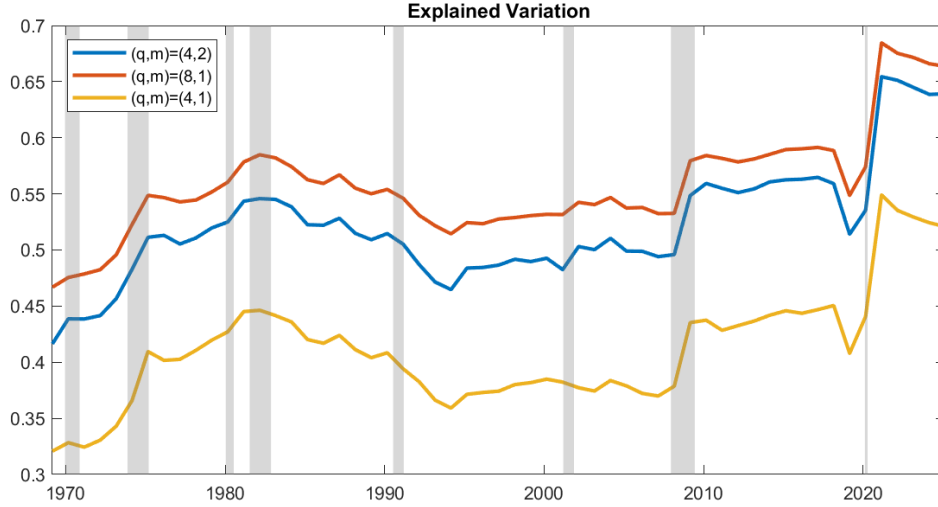


Figure 3: Explained variation of the common component.

Based on the PC2 criterion, we find evidence supporting the specification $(q, m) = (4, 2)$ for the majority of the sample period. The corresponding common component explains approximately 52% of the total variation in average, with this share increas-

ing to about 65% in the post-COVID-19 period. A similar, but less drastic, increase in explained variation is also observed during the 2008 global financial crisis.

When adopting the static specification, $(q, m) = (8, 1)$, the explained variation increases by approximately 3.6 percentage points. This gain can be interpreted as reflecting improved in-sample fit due to the greater flexibility of the model.

In contrast, holding the number of factors fixed while reducing the filter length to $(q, m) = (4, 1)$ leads to a substantial decline in explained variation, on the order of 11.1 percentage points. This deterioration can be interpreted as evidence of misspecification.

7 Conclusion

This paper develops methods for determining the structure of finite-filter dynamic factor models. Because a common component may admit multiple observationally equivalent representations, the number of dynamic factors and the filter length cannot be considered separately. We characterize the admissible structures by showing that a candidate pair (q, m) can represent the true common component when

$$q \geq q_0 \quad \text{and} \quad qm \geq q_0 m_0.$$

The true structure (q_0, m_0) is identified as the representation with the smallest factor dimension and, among such representations, the shortest filter length.

We propose an alternating least-squares estimator that directly imposes the lag-stacking restrictions of the dynamic factor model. Under the maintained nondegeneracy, pervasiveness, and weak-idiosyncratic-component conditions, the estimated factor space is consistent for every admissible structure. When the true structure is used, the underlying dynamic factors and lag-loading matrices are consistently estimated up to a common nonsingular rotation.

Using these estimates, we develop dynamic singular-value-ratio and information-criterion procedures for selecting (q_0, m_0) . The ratio procedures extend the eigenvalue-ratio method of Ahn and Horenstein (2013), while the information criteria extend Bai and Ng (2002) by penalizing both the dynamic factor dimension q and the associated static dimension qm . Both classes of procedures are consistent over a fixed candidate grid. Importantly, the selection results do not require the factors to follow an exact

finite-order VAR process. When a VAR specification is subsequently imposed, its lag order can also be consistently selected using the estimated dynamic factors.

The simulations show that the proposed procedures improve as both panel dimensions increase, although serial and cross-sectional dependence and weak separation near the admissibility boundary can make filter-length selection difficult in finite samples. Empirical results using the FRED-MD dataset indicate evidence of filter length $m = 2$ with four dynamic factors prior to the COVID-19 period, which increases to seven post-COVID-19.

References

- Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.
- Amengual, D. and 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.
- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.
- Bai, J. and Li, K. (2016). Maximum likelihood estimation and inference for approximate factor models of high dimension. *The Review of Economics and Statistics*, 98(2):298–309.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.
- Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, 25(1):52–60.
- Bai, J. and Ng, S. (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*, 235(2):1893–1916.
- Bai, J. and Wang, P. (2014). Identification theory for high dimensional static and dynamic factor models. *Journal of Econometrics*, 178(2):794–804.
- Bai, J. and Wang, P. (2015). Identification and bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics*, 33(2):221–240.
- Barigozzi, M., Hallin, M., Luciani, M., and Zaffaroni, P. (2024). Inferential theory for generalized dynamic factor models. *Journal of Econometrics*, 239(2):105422.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York, 2 edition.
- Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304.

- Eichenbaum, M. and Evans, C. L. (1995). Some empirical evidence on the effects of shocks to monetary policy on exchange rates*. *The Quarterly Journal of Economics*, 110(4):975–1009.
- Forni, Reichlin, L., Hallin, M., and Lippi, M. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82:540–554.
- Forni, M. and Gambetti, L. (2010). The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics*, 57(2):203–216.
- Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25(5):1319–1347.
- Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, 185(2):359–371.
- Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2017). Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. *Journal of Econometrics*, 199(1):74–92.
- Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.
- McCracken, M. W. and Ng, S. (2015). Fred-md: A monthly database for macroeconomic research. Working Papers 2015-12, Federal Reserve Bank of St. Louis.
- Moon, H. R. and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195.
- Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016.
- Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

8 Appendix

Proof of Proposition 1 and Proposition 4 are in Supplemental Materials.

8.1 Proof of Proposition 2

Proof of Proposition 2. Let

$$C = G\Gamma'$$

denote the true common-component matrix, so that $X = C + E$, and set

$$a_{NT} = \frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}, \quad x_{NT} = \frac{1}{\sqrt{T}} \|M_{\hat{G}}G\|_F.$$

Note that the minimization problem can be concentrated as

$$\min_{(G,\Gamma)} \frac{1}{NT} \|X - G\Gamma'\|_F^2 = \min_G \frac{1}{NT} \|M_G X\|_F^2.$$

Because $(q, m) \in S_1$, Proposition 1 implies that C has an exact representation with structure (q, m) . Hence, by the optimality of $(\hat{G}, \hat{\Gamma})$ and the definition of $P_{\hat{G}}$,

$$\frac{1}{NT} \|M_{\hat{G}}X\|_F^2 \leq \frac{1}{NT} \|X - \hat{G}\hat{\Gamma}'\|_F^2 \leq \frac{1}{NT} \|E\|_F^2.$$

Using $X = G\Gamma' + E$ and the orthogonal decomposition $E = P_{\hat{G}}E + M_{\hat{G}}E$, we obtain

$$\begin{aligned} 0 \leq & -\frac{1}{NT} \|M_{\hat{G}}G\Gamma'\|_F^2 + \frac{1}{NT} \|P_{\hat{G}}E\|_F^2 \\ & - \frac{2}{NT} \text{tr}[(M_{\hat{G}}G)'E\Gamma]. \end{aligned} \tag{13}$$

For the first term,

$$\begin{aligned} \frac{1}{NT} \|M_{\hat{G}}G\Gamma'\|_F^2 &= \text{tr} \left[\left(\frac{\Gamma'\Gamma}{N} \right) \left(\frac{G' M_{\hat{G}} G}{T} \right) \right] \\ &\geq \lambda_{\min} \left(\frac{\Gamma'\Gamma}{N} \right) x_{NT}^2. \end{aligned}$$

By Assumption 1, there exists a constant $c > 0$ such that

$$\lambda_{\min} \left(\frac{\Gamma' \Gamma}{N} \right) \geq c$$

with probability approaching one. Next, letting $r = qm$, we have $\text{rank}(P_{\hat{G}}) \leq r$. Because (q, m) is fixed,

$$\begin{aligned} 0 \leq \frac{1}{NT} \|P_{\hat{G}} E\|_F^2 &\leq \frac{r}{NT} \|E\|^2 \\ &= O_p \left(\frac{1}{N} + \frac{1}{T} \right) \\ &= O_p(a_{NT}^2), \end{aligned}$$

where $\|\cdot\|$ denotes the spectral norm and the second equality follows from Assumption 2.

For the cross term, the Frobenius Cauchy–Schwarz inequality gives

$$\begin{aligned} \left| \frac{2}{NT} \text{tr} [(M_{\hat{G}} G)' E \Gamma] \right| &\leq \frac{2}{NT} \|M_{\hat{G}} G\|_F \|E \Gamma\|_F \\ &\leq \frac{2}{NT} \|M_{\hat{G}} G\|_F \|E\| \|\Gamma\|_F. \end{aligned}$$

Assumption 1 implies $\|\Gamma\|_F = O_p(\sqrt{N})$, while Assumption 2 implies $\|E\| = O_p(\sqrt{\max\{N, T\}})$. Therefore,

$$\begin{aligned} \left| \frac{2}{NT} \text{tr} [(M_{\hat{G}} G)' E \Gamma] \right| &= O_p \left(\frac{\sqrt{\max\{N, T\}}}{\sqrt{NT}} \right) x_{NT} \\ &= O_p(a_{NT}) x_{NT}. \end{aligned}$$

Combining these bounds with (13), there exist nonnegative random variables $A_{NT} = O_p(1)$ and $B_{NT} = O_p(1)$ such that, with probability approaching one,

$$c x_{NT}^2 \leq A_{NT} a_{NT}^2 + B_{NT} a_{NT} x_{NT}.$$

By Young's inequality,

$$B_{NT} a_{NT} x_{NT} \leq \frac{c}{2} x_{NT}^2 + \frac{B_{NT}^2}{2c} a_{NT}^2.$$

It follows that

$$x_{NT}^2 \leq \left(\frac{2A_{NT}}{c} + \frac{B_{NT}^2}{c^2} \right) a_{NT}^2 = O_p(a_{NT}^2).$$

Consequently,

$$\frac{1}{\sqrt{T}} \|M_{\hat{G}} G\|_F = x_{NT} = O_p \left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}} \right).$$

□

8.2 Proof of Theorem 1 and 2

Lemma 1 (Lower bound on the fitted residual). *For every $(q, m) \in S$,*

$$\hat{\delta}_{q+1,m}(X) \geq \delta_{r_0+qm+1,1}(E) = \sigma_{r_0+qm+1}(E).$$

Proof. Since

$$\text{rank}(C_0) \leq r_0 \quad \text{and} \quad \text{rank}(\hat{C}_{q,m}) \leq qm,$$

we have

$$\text{rank}(\hat{C}_{q,m} - C_0) \leq r_0 + qm.$$

Using $X = C_0 + E$, it follows that

$$\begin{aligned} \hat{\delta}_{q+1,m}(X) &= \left\| E - (\hat{C}_{q,m} - C_0) \right\| \\ &\geq \inf_{\text{rank}(A) \leq r_0 + qm} \|E - A\| \\ &= \sigma_{r_0+qm+1}(E) \\ &= \delta_{r_0+qm+1,1}(E), \end{aligned}$$

where the third equality follows from the Eckart–Young theorem. □

To apply Assumption 3 to the preceding bound uniformly over S , it is sufficient to choose

$$k_{\max} \geq r_0 + q_{\max} m_{\max}.$$

Lemma 2 (Consistency of the fitted common component). *Let Assumption 2 hold. Then*

$$\max_{(q,m) \in S_1} \frac{1}{\sqrt{NT}} \left\| \hat{C}_{q,m} - C_0 \right\|_F = O_p \left(\sqrt{\frac{N+T}{NT}} \right).$$

Proof. Fix $(q, m) \in S_1$ and define

$$\Delta_{q,m} = \hat{C}_{q,m} - C_0.$$

By Proposition 1, C_0 admits an exact representation with structure (q, m) and is therefore feasible in the least-squares problem defining $\hat{C}_{q,m}$. Global optimality gives

$$\left\| X - \hat{C}_{q,m} \right\|_F^2 \leq \left\| X - C_0 \right\|_F^2 = \|E\|_F^2.$$

Since

$$X - \hat{C}_{q,m} = E - \Delta_{q,m},$$

we obtain

$$\left\| \Delta_{q,m} \right\|_F^2 \leq 2 \left| \text{tr} (E' \Delta_{q,m}) \right|.$$

By spectral–nuclear norm duality,

$$\left| \text{tr} (E' \Delta_{q,m}) \right| \leq \|E\| \left\| \Delta_{q,m} \right\|_*,$$

where $\|\cdot\|_*$ denotes the nuclear norm. Moreover,

$$\left\| \Delta_{q,m} \right\|_* \leq \sqrt{\text{rank}(\Delta_{q,m})} \left\| \Delta_{q,m} \right\|_F,$$

and

$$\text{rank}(\Delta_{q,m}) \leq \text{rank}(\hat{C}_{q,m}) + \text{rank}(C_0) \leq qm + r_0.$$

Therefore, if $\Delta_{q,m} \neq 0$,

$$\left\| \Delta_{q,m} \right\|_F \leq 2\sqrt{qm + r_0} \|E\|.$$

The same inequality holds trivially when $\Delta_{q,m} = 0$.

Since

$$qm + r_0 \leq q_{\max} m_{\max} + r_0$$

uniformly over $(q, m) \in S_1$, it follows that

$$\max_{(q,m) \in S_1} \left\| \hat{C}_{q,m} - C_0 \right\|_F \leq 2\sqrt{q_{\max} m_{\max} + r_0} \|E\|.$$

Assumption 2 gives

$$\|E\| = O_p \left(\sqrt{\max\{N, T\}} \right).$$

Consequently,

$$\begin{aligned} \max_{(q,m) \in S_1} \frac{1}{\sqrt{NT}} \|\hat{C}_{q,m} - C_0\|_F &= O_p \left(\sqrt{\frac{\max\{N, T\}}{NT}} \right) \\ &= O_p \left(\sqrt{\frac{N+T}{NT}} \right), \end{aligned}$$

which proves the result. $\square$

Lemma 3 (Upper bound on the fitted residual). *Suppose Assumptions 1 and 2 hold. We have*

$$\max_{(q,m) \in S_1} \frac{1}{\sqrt{NT}} \hat{\delta}_{q+1,m}(X) = O_p \left(\sqrt{\frac{N+T}{NT}} \right).$$

Proof. Let

$$C_0 = \sum_{k=0}^{m_0-1} F_{-k} \lambda'_k.$$

For every $(q, m) \in S_1$,

$$\begin{aligned} \hat{\delta}_{q+1,m}(X) &= \|X - \hat{C}_{q,m}\| \\ &\leq \|E\| + \|\hat{C}_{q,m} - C_0\| \\ &\leq \|E\| + \|\hat{C}_{q,m} - C_0\|_F. \end{aligned}$$

Assumption 2 gives

$$\frac{\|E\|}{\sqrt{NT}} = O_p \left(\sqrt{\frac{N+T}{NT}} \right),$$

while Lemma 2 gives

$$\max_{(q,m) \in S_1} \frac{\|\hat{C}_{q,m} - C_0\|_F}{\sqrt{NT}} = O_p \left(\sqrt{\frac{N+T}{NT}} \right).$$

The result follows. $\square$

Lemma 4. *Suppose Assumptions 2 and Assumption 3 hold with $q_0 m_0 + q_{\max} m_{\max} \leq k_{\max}$. Let $(q, m) \in S_1$. Then, for all $\epsilon > 0$, there exists $c_0, c_1 > 0$ such that*

$$\mathbb{P} \left\{ c_0 \leq \frac{1}{\sqrt{N+T}} \hat{\delta}_{q+1,m}(X) \leq c_1 \text{ for all } (q, m) \in S_1 \right\} \geq 1 - \epsilon$$

for all N, T large.

Proof. By Lemma 3,

$$\max_{(q,m) \in S_1} \frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{N+T}} = O_p(1).$$

Hence, for every $\epsilon > 0$, there exists $c_U > 0$ such that

$$\mathbb{P} \left(\max_{(q,m) \in S_1} \frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{N+T}} > c_U \right) \leq \frac{\epsilon}{2}$$

for all sufficiently large N and T .

Moreover, Lemma 1 and the ordering of singular values give, uniformly over $(q, m) \in S_1$,

$$\hat{\delta}_{q+1,m}(X) \geq \sigma_{r_0+qm+1}(E) \geq \sigma_{k_{\max}+1}(E).$$

Assumption 3 therefore implies that there exists $c_L > 0$ such that

$$\mathbb{P} \left(\min_{(q,m) \in S_1} \frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{N+T}} < c_L \right) \leq \frac{\epsilon}{2}$$

for all sufficiently large N and T . Combining the two bounds proves the result. $\square$

Lemma 5. *Suppose Assumption 1, 2, and 4 hold. For every $\epsilon > 0$, there exist constants $c_L, c_U > 0$ such that*

$$\mathbb{P} \left(c_L \leq \frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{NT}} \leq c_U \text{ for every } (q, m) \in S_2 \cup S_3 \right) \geq 1 - \epsilon$$

for all sufficiently large N and T .

Proof. First, because 0 is feasible in every candidate model and $\hat{C}_{q,m}$ is a global least-squares minimizer,

$$\hat{\delta}_{q+1,m}(X) \leq \|X - \hat{C}_{q,m}\|_F \leq \|X\|_F.$$

Assumptions 1 and 2 imply

$$\frac{\|X\|_F}{\sqrt{NT}} = O_p(1).$$

Hence,

$$\max_{(q,m) \in S_2 \cup S_3} \frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{NT}} = O_p(1).$$

For the lower bound, for every (q, m) ,

$$\frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{NT}} \geq \frac{\delta_{q+1,m}(C_0)}{\sqrt{NT}} - \frac{\|E\|}{\sqrt{NT}}.$$

Assumption 2 gives

$$\frac{\|E\|}{\sqrt{NT}} = o_p(1).$$

If $(q, m) \in S_2$, then $qm < r_0$, and hence

$$\delta_{q+1,m}(C_0) \geq \sigma_{qm+1}(C_0) \geq \sigma_{r_0}(C_0).$$

Moreover,

$$\frac{\sigma_{r_0}^2(C_0)}{NT} \geq \lambda_{\min} \left(\frac{G'G}{T} \right) \lambda_{\min} \left(\frac{\Gamma\Gamma}{N} \right),$$

which is bounded away from zero in probability by Assumption 1.

If $(q, m) \in S_3$, Assumption 4 implies that

$$\frac{\delta_{q+1,m}(C_0)}{\sqrt{NT}}$$

is uniformly bounded away from zero in probability.

Since $S_2 \cup S_3$ is finite, the preceding bounds imply that, for every $\epsilon > 0$, there exist constants $c_L, c_U > 0$ such that

$$\mathbb{P} \left(c_L \leq \frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{NT}} \leq c_U \text{ for every } (q, m) \in S_2 \cup S_3 \right) \geq 1 - \epsilon$$

for all sufficiently large N and T . □

Proof of Theorem 1

Proof of Theorem 1. Fix $m \in \{1, \dots, m_{\max}\}$ and define

$$q_m^* = \max \left\{ q_0, \left\lceil \frac{q_0 m_0}{m} \right\rceil \right\}.$$

By the definition of S_1 ,

$$(q, m) \in S_1 \iff q \geq q_m^*.$$

Thus, for the fixed candidate filter length m , structures with $q < q_m^*$ belong to $S_2 \cup S_3$, whereas structures with $q \geq q_m^*$ belong to S_1 .

Recall that

$$DR_{q,m}(X) = \frac{\hat{\delta}_{q,m}(X)}{\hat{\delta}_{q+1,m}(X)}, \quad q = 1, \dots, q_{\max},$$

where $\hat{\delta}_{q,m}(X)$ is the residual associated with $q - 1$ fitted dynamic factors.

Because the candidate set is finite, Lemmas 4 and 5, together with a union bound, imply that, for every $\epsilon > 0$, there exist constants

$$0 < c_L < c_U < \infty, \quad 0 < d_L < d_U < \infty,$$

such that the event

$$\mathcal{E}_{NT} = \left\{ \bigcap_{\ell=q_m^*}^{q_{\max}} \left\{ c_L \leq \frac{\hat{\delta}_{\ell+1,m}(X)}{\sqrt{N+T}} \leq c_U \right\} \right\} \cap \left\{ \bigcap_{\ell=0}^{q_m^*-1} \left\{ d_L \leq \frac{\hat{\delta}_{\ell+1,m}(X)}{\sqrt{NT}} \leq d_U \right\} \right\}$$

satisfies

$$\mathbb{P}(\mathcal{E}_{NT}) \geq 1 - \epsilon$$

for all sufficiently large N and T .

We first consider the ratio at $q = q_m^*$. Its numerator, $\hat{\delta}_{q_m^*,m}(X)$, corresponds to the structure $(q_m^* - 1, m) \in S_2 \cup S_3$, whereas its denominator, $\hat{\delta}_{q_m^*+1,m}(X)$, corresponds to

$(q_m^*, m) \in S_1$. Hence, on $\mathcal{E}_{NT}$,

$$\begin{aligned} DR_{q_m^*, m}(X) &= \frac{\hat{\delta}_{q_m^*, m}(X)}{\hat{\delta}_{q_m^*+1, m}(X)} \\ &= \sqrt{\frac{NT}{N+T}} \frac{\hat{\delta}_{q_m^*, m}(X)/\sqrt{NT}}{\hat{\delta}_{q_m^*+1, m}(X)/\sqrt{N+T}} \\ &\geq \frac{d_L}{c_U} \sqrt{\frac{NT}{N+T}}. \end{aligned}$$

Because $N, T \rightarrow \infty$,

$$\frac{NT}{N+T} \geq \frac{1}{2} \min\{N, T\} \rightarrow \infty.$$

Therefore,

$$DR_{q_m^*, m}(X) \rightarrow_p \infty.$$

We next bound the ratios away from q_m^* . If $1 \leq q < q_m^*$, then both $(q-1, m)$ and (q, m) belong to $S_2 \cup S_3$. Thus, on $\mathcal{E}_{NT}$,

$$DR_{q, m}(X) = \frac{\hat{\delta}_{q, m}(X)/\sqrt{NT}}{\hat{\delta}_{q+1, m}(X)/\sqrt{NT}} \leq \frac{d_U}{d_L}.$$

If $q_m^* < q \leq q_{\max}$, then both $(q-1, m)$ and (q, m) belong to S_1 . Hence, on $\mathcal{E}_{NT}$,

$$DR_{q, m}(X) = \frac{\hat{\delta}_{q, m}(X)/\sqrt{N+T}}{\hat{\delta}_{q+1, m}(X)/\sqrt{N+T}} \leq \frac{c_U}{c_L}.$$

Consequently, on $\mathcal{E}_{NT}$,

$$\max_{\substack{1 \leq q \leq q_{\max} \\ q \neq q_m^*}} DR_{q, m}(X) \leq M, \quad M = \max \left\{ \frac{d_U}{d_L}, \frac{c_U}{c_L} \right\}.$$

Since

$$\frac{d_L}{c_U} \sqrt{\frac{NT}{N+T}} \rightarrow \infty,$$

for all sufficiently large N and T ,

$$\frac{d_L}{c_U} \sqrt{\frac{NT}{N+T}} > M.$$

Therefore, on $\mathcal{E}_{NT}$ and for all sufficiently large N, T ,

$$DR_{q_m^*, m}(X) > \max_{\substack{1 \leq q \leq q_{\max} \\ q \neq q_m^*}} DR_{q, m}(X).$$

It follows that the maximizer is uniquely attained at q_m^* on $\mathcal{E}_{NT}$. Hence,

$$\mathbb{P}\{\hat{q} = q_m^*\} \geq \mathbb{P}(\mathcal{E}_{NT}) \geq 1 - \epsilon$$

for all sufficiently large N and T . Since $\epsilon > 0$ is arbitrary,

$$\mathbb{P}\{\hat{q} = q_m^*\} \rightarrow 1.$$

Finally, if $m_0 \leq m \leq m_{\max}$, then

$$\frac{q_0 m_0}{m} \leq q_0,$$

and therefore

$$q_m^* = q_0.$$

Thus,

$$\mathbb{P}\{\hat{q} = q_0\} \rightarrow 1.$$

If instead $1 \leq m < m_0$, then

$$\frac{q_0 m_0}{m} > q_0,$$

so that

$$q_m^* = \left\lceil \frac{q_0 m_0}{m} \right\rceil.$$

Consequently,

$$\mathbb{P}\left\{\hat{q} = \left\lceil \frac{q_0 m_0}{m} \right\rceil\right\} \rightarrow 1.$$

□

Proof of Theorem 2

Proof of Theorem 2. Fix $q \in \{q_0, \dots, q_{\max}\}$ and let

$$m_q^* = \left\lceil \frac{q_0 m_0}{q} \right\rceil.$$

Because $q \geq q_0$, the definition of S_1 implies

$$(q, m) \in S_1 \iff m \geq m_q^*.$$

Hence, $(q, m) \in S_2$ for $m < m_q^*$ and $(q, m) \in S_1$ for $m \geq m_q^*$.

Since the candidate set is finite, Lemmas 4 and 5 imply that, for every $\epsilon > 0$, there exist constants

$$0 < c_L < c_U < \infty, \quad 0 < d_L < d_U < \infty,$$

such that, with probability at least $1 - \epsilon$ for all sufficiently large N and T ,

$$c_L \leq \frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{N+T}} \leq c_U \quad \text{for all } m = m_q^*, \dots, m_{\max},$$

and

$$d_L \leq \frac{\hat{\delta}_{q+1,m}(X)}{\sqrt{NT}} \leq d_U \quad \text{for all } m = 0, \dots, m_q^* - 1.$$

At $m = m_q^*$, the numerator of $MR_{q,m}(X)$ corresponds to the inadmissible structure $(q, m_q^* - 1)$, whereas the denominator corresponds to the admissible structure (q, m_q^*) . Therefore, on the preceding event,

$$\begin{aligned} MR_{q,m_q^*}(X) &= \frac{\hat{\delta}_{q+1,m_q^*-1}(X)}{\hat{\delta}_{q+1,m_q^*}(X)} \\ &= \sqrt{\frac{NT}{N+T}} \frac{\hat{\delta}_{q+1,m_q^*-1}(X)/\sqrt{NT}}{\hat{\delta}_{q+1,m_q^*}(X)/\sqrt{N+T}} \\ &\geq \frac{d_L}{c_U} \sqrt{\frac{NT}{N+T}}. \end{aligned}$$

Since

$$\frac{NT}{N+T} \geq \frac{1}{2} \min\{N, T\} \rightarrow \infty,$$

we have

$$MR_{q,m_q^*}(X) \rightarrow_p \infty.$$

For $m < m_q^*$, both $(q, m - 1)$ and (q, m) belong to S_2 , so

$$MR_{q,m}(X) \leq \frac{d_U}{d_L}.$$

For $m > m_q^*$, both $(q, m - 1)$ and (q, m) belong to S_1 , so

$$MR_{q,m}(X) \leq \frac{c_U}{c_L}.$$

Thus,

$$\max_{\substack{1 \leq m \leq m_{\max} \\ m \neq m_q^*}} MR_{q,m}(X) = O_p(1),$$

whereas

$$MR_{q,m_q^*}(X) \rightarrow_p \infty.$$

It follows that

$$\mathbb{P}\{\hat{m} = m_q^*\} \rightarrow 1.$$

Finally, when $q = q_0$,

$$m_q^* = \left\lceil \frac{q_0 m_0}{q_0} \right\rceil = m_0,$$

which gives

$$\mathbb{P}\{\hat{m} = m_0\} \rightarrow 1.$$

For $q > q_0$, the stated conclusion follows directly from the definition of m_q^* . $\square$

8.3 Proof of Theorem 3

Let $V(q, m)$ denote the mean squared errors when the structure (q, m) is used to estimate a dynamic factor model, i.e.,

$$V(q, m) = \frac{1}{NT} \left\| X - \sum_{k=0}^{m-1} \hat{F}_{-k}^{(q)} \hat{\lambda}_k^{(q)'} \right\|_F^2,$$

where $(\hat{F}_{-k}^{(q,m)}, \hat{\lambda}_k^{(q,m)})_{k=0}^{m-1}$ is the least squares estimator obtained under the structure (q, m) .

Lemma 6 (Rates of the minimized mean squared errors). *Suppose that $q_0, m_0 \geq 0$*

and that Assumptions 1, 2, and 4 hold. Then

$$\max_{(q,m) \in S_1} \{V(q_0, m_0) - V(q, m)\} = O_p\left(\frac{N+T}{NT}\right).$$

For every $\epsilon > 0$, there exists a constant $c > 0$ such that

$$\mathbb{P}\left\{\min_{(q,m) \in S_2 \cup S_3} [V(q, m) - V(q_0, m_0)] \geq c\right\} \geq 1 - \epsilon$$

for all sufficiently large N and T .

Proof. For any given (q, m) , we have

$$V(q, m) = \frac{1}{NT} \left\| \hat{C}_{q,m} - C_0 \right\|_F^2 + \frac{1}{NT} \|E\|_F^2 - \frac{2}{NT} \text{tr} \left(\left(\hat{C}_{q,m} - C_0 \right)' E \right),$$

so that

$$\begin{aligned} V(q_0, m_0) - V(q, m) &= \frac{1}{NT} \left\| \hat{C}_{q_0, m_0} - C_0 \right\|_F^2 - \frac{1}{NT} \left\| \hat{C}_{q,m} - C_0 \right\|_F^2 \\ &\quad - \frac{2}{NT} \text{tr} \left(\left(\hat{C}_{q_0, m_0} - C_0 \right)' E \right) + \frac{2}{NT} \text{tr} \left(\left(\hat{C}_{q,m} - C_0 \right)' E \right). \end{aligned} \quad (14)$$

Also, we have

$$\frac{1}{NT} \text{tr} \left(\left(\hat{C}_{q,m} - C_0 \right)' E \right) \leq \frac{\sqrt{mq + m_0 q_0}}{NT} \left\| \hat{C}_{q,m} - C_0 \right\|_F \|E\|. \quad (15)$$

If $(q, m) \in S_1$, the RHS of equation (14) is $O_p\left(\frac{N+T}{NT}\right)$ due to Lemma 2, equation (15), and Assumption 2.

Now, if $(q, m) \in S_2 \cup S_3$, we have $\frac{1}{NT} \left\| \hat{C}_{q,m} - C_0 \right\|_F^2$ strictly positive in the limit. To see this,

$$\frac{1}{\sqrt{NT}} \left\| \hat{C}_{q,m} - C_0 \right\|_F \geq \frac{1}{\sqrt{NT}} \left\| \hat{C}_{q,m} - C_0 \right\| \geq \frac{1}{\sqrt{NT}} \delta_{q+1, m} \left(\sum_{k=0}^{m_0-1} F_{-k} \lambda'_{-k} \right),$$

which is bounded away from zero in probability by Lemma 5. Thus, from (14), (15)

and Lemma 2, we have

$$V(q, m) - V(q_0, m_0) = \frac{1}{NT} \left\| \hat{C}_{q,m} - C_0 \right\|_F^2 + o_p(1)$$

bounded away from zero for all $(q, m) \in S_2 \cup S_3$. $\square$

Proof of Theorem 3

Proof of Theorem 3. We show that

$$\mathbb{P}\{PC(q, m) < PC(q_0, m_0)\} \rightarrow 0$$

as $N, T \rightarrow \infty$ for all $(q, m) \in S \setminus \{(q_0, m_0)\}$.

We first consider the case $(q, m) \in S_1 \setminus \{(q_0, m_0)\}$. Using Lemma 6, we have

$$\max_{(q,m) \in S_1} \{V(q_0, m_0) - V(q, m)\} = O_p\left(\frac{N+T}{NT}\right).$$

Note that we may write

$$\mathbb{P}\{PC(q, m) < PC(q_0, m_0)\} = \mathbb{P}\{V(q_0, m_0) - V(q, m) > (r + q - (r_0 + q_0))g(N, T)\},$$

where the $V(q_0, m_0) - V(q, m)$ vanishes at the order $O_p\left(\frac{N+T}{NT}\right)$ and the $g(N, T)$ vanishes at a slower rate. Note that any $(q, m) \in S_1 \setminus \{(q_0, m_0)\}$ implies $r + q > r_0 + q_0$ so that $\mathbb{P}\{PC(q, m) < PC(q_0, m_0)\} \rightarrow 0$.

Next, we consider the case $(q, m) \in S_2 \cup S_3$. Using Lemma 6, for every $\epsilon > 0$, there exists a constant $c > 0$ such that

$$\mathbb{P}\left\{\min_{(q,m) \in S_2 \cup S_3} [V(q, m) - V(q_0, m_0)] \geq c\right\} \geq 1 - \epsilon$$

for all sufficiently large N and T . Note that we have

$$PC(q, m) - PC(q_0, m_0) = V(q, m) - V(q_0, m_0) - (r_0 + q_0 - (r + q))g(N, T),$$

where the first term $V(q, m) - V(q_0, m_0)$ is bounded away from zero, whereas the second term $(r_0 + q_0 - (r + q))g(N, T)$ vanishes. Thus, we have $\mathbb{P}\{PC(q, m) < PC(q_0, m_0)\} \rightarrow 0$.

For $DC(q, m)$, note that we have

$$\frac{1}{NT} \hat{\delta}_{q+1, m}^2(X) = O_p\left(\frac{N+T}{NT}\right)$$

when $(q, m) \in S_1$ due to Lemma 3. Also, for all $\epsilon > 0$, there exists $c > 0$ such that

$$\mathbb{P}\left\{\frac{1}{NT} \hat{\delta}_{q+1, m}^2(X) \geq c\right\} \geq 1 - \epsilon$$

when $(q, m) \in S_2 \cup S_3$ due to Lemma 5. Since $\frac{1}{NT} \hat{\delta}_{q_0+1, m_0}^2(X) = O_p\left(\frac{N+T}{NT}\right)$, we may use the same argument as in the case of $PC(q, m)$ to finish. $\square$

Proof of Corollary 2. Let

$$a_{NT} = \frac{N+T}{NT}, \quad K(q, m) = qm + q.$$

By Lemma 6,

$$\max_{(q, m) \in S_1} \left| V(q, m) - \frac{\|E\|_F^2}{NT} \right| = O_p(a_{NT}).$$

Assumption 5 therefore implies

$$\max_{(q, m) \in S_1} |V(q, m) - \sigma_\epsilon^2| \rightarrow_p 0.$$

The mean-value theorem then gives

$$\max_{(q, m) \in S_1} |\log V(q, m) - \log V(q_0, m_0)| = O_p(a_{NT}).$$

For every $(q, m) \in S_1 \setminus \{(q_0, m_0)\}$,

$$K(q, m) - K(q_0, m_0) > 0.$$

Since $g(N, T)/a_{NT} \rightarrow \infty$,

$$\mathbb{P}\left\{\min_{(q, m) \in S_1 \setminus \{(q_0, m_0)\}} [IC(q, m) - IC(q_0, m_0)] > 0\right\} \rightarrow 1.$$

For $(q, m) \in S_2 \cup S_3$, Lemma 6 gives a constant $c > 0$ such that

$$\mathbb{P} \left\{ \min_{(q,m) \in S_2 \cup S_3} [V(q, m) - V(q_0, m_0)] \geq c \right\} \rightarrow 1.$$

Because $V(q_0, m_0) \rightarrow_p \sigma_\varepsilon^2 > 0$,

$$\min_{(q,m) \in S_2 \cup S_3} [\log V(q, m) - \log V(q_0, m_0)]$$

is bounded away from zero in probability. Since $g(N, T) \rightarrow 0$ and the candidate set is finite,

$$\mathbb{P} \left\{ \min_{(q,m) \in S_2 \cup S_3} [IC(q, m) - IC(q_0, m_0)] > 0 \right\} \rightarrow 1.$$

Combining the two cases proves

$$\mathbb{P}\{(\hat{q}, \hat{m}) = (q_0, m_0)\} \rightarrow 1.$$

□