# Estimation of Constrained Factor Models for High-Dimensional Time Series

Yitian Liu[a] Jiazhu Pan[b] Qiang Xia[a*]

[a] *College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China*

[b] *Department of Mathematics & Statistics, University of Strathclyde, Glasgow G1 1XH, UK*

**Abstract**

This article studies the estimation of the constrained factor models for high-dimensional time series. The approach is based on the eigenanalysis of a non-negative definite matrix constructed from the auto-covariance matrices. The convergence rate of the estimator for loading matrix and the asymptotic normality of the estimated factor score are explored under regularity conditions set for the proposed model. Our estimation for the constrained factor models can achieve the optimal rate of convergence even in the case of weak factors. The finite sample performance of our approach is examined and compared with the existing methods by Monte Carlo simulations. Our methodology is illustrated and supported by a real data example.

**Key words**: Constrained factor models; High-dimensional time series; Convergence rate; Asymptotic normality.

## 1 Introduction

In this era of big data, data often have high dimensions in many fields such as finance, economics, environmental science and medicineh. Modeling high-dimensional time series data is of paramount interest and increasing importance. However, when the dimension of time series is large, the standard multiple time series models such as vector ARMA models

[*]The authors are arranged in alphabetical order. Qiang Xia is the corresponding author, Email: xiaqiang@scau.edu.cn.



This is a peer-reviewed, accepted manuscript of the following paper: Liu, Y., Pan, J., & Xia, Q. (in press). Estimation of constrained factor models for high-dimensional time series. Journal of Forecasting.

can seldom be used directly in practice due to the problem of overparametrization or so-called "curse of dimensionality". A common approach to efficient analysis of high-dimensional time series is to construct a factor model, which achieves dimension reduction by expressing the high-dimensional time series as a low-dimensional process. Many attempts have been made to estimate the number of common factors and the factor loading matrix. Examples of this type of attempts include Chamberlain and Rothschild (1983), Bai and Ng (2002), Stock and Watson (2002), Bai (2003), Alessi et al. (2010), Onatski (2010), Ahn and Horenstein (2013), Fan et al. (2013), Xia et al. (2017) and Wu (2021) for the static factor model, where the observed data only exhibit a contemporaneous relationship with the factors. Forni et al. (2000), Forni et al. (2005), Stock and Watson (2005), Amengual and Watson (2007), Hallin and Liska (2007), Forni et al. (2009), Onatski (2009), Doz et al. (2011), and Stock and Watson (2011) investigated the generalized case where the factors are loaded via filters.

In the above literature, it is assumed that the common factors have a significant impact on most of the high-dimensional time series in terms of the cross-section dependence. Additionally, after removing the common factors associated with the factor loading matrix from the high-dimensional time series, the error process is referred to as the idiosyncratic noise component (Bai, 2003) and assumed to have weak cross-sectional and serial correlations. Because both of two components exhibit serial dependence, one drawback of the above assumptions is that the identification of common factors and idiosyncratic noise is only possible under an asymptotic sense.

Alternatively, relying on the eigen-analysis of the auto-covariance structure of the time series data, the different latent factor models were considered by Pan and Yao (2008), Lam et al. (2011), Lam and Yao (2012), Xia et al. (2015), Chan et al. (2017) and Zhang et al. (2023). In these models, the error process must be a white noise, and the serial dependence of the high-dimensional time series is entirely driven by the common factors. Since white noise does not exhibit a serial correlation, the decomposition is unique in that the number of factors and the factor loading space can be identified for any finite dimension. Meanwhile, the theoretical framework for these time series factor models has the following advantages: (i) The idiosyncratic component does not require the distributional assumption and can have strong cross-sectional correlation; (ii) it allows for the cross-correlation between the factors and idiosyncratic components; (iii) it introduces the strength of factors to characterize the influence of factors.

However, when the dimension of the observed variables is very large, both factors and noises will be identified as the factors in the factor model. To address this difficulty, Tsai and Tsay (2010) and Tsai et al. (2016) added a constraint matrix composed of known prior information to the time series factor model. Constraints come from the experience and prior knowledge system or the data itself exhibiting certain specific structure. This prior information should be incorporated into the factor model to help the researchers obtain more concise econometric models for prediction. The results of Tsai and Tsay (2010) showed that adding constraints could improve the accuracy of inference and produce more interpretive factors. To our best knowledge, for the factor models, there are different definitions of "factors" in the literature. Then the approaches for estimating procedure of the loading matrix as well as the number of factors are different. In this article, we investigate the estimation for the constrained factor model introduced by Tsai and Tsay (2010) under different conditions. So it is worthy to have further investigation on estimation for the constrained factor models.

It is well known that the principal component estimation (PCE) is a simple and stable estimator for the approximating factor model. The PCE is obtained by performing eigen-decomposition on a non-negative definite matrix, such as the sample covariance matrix. See Stock and Watson (2002), Bai and Ng (2002) and Bai (2003) among others. Tsai and Tsay (2010) suggested the maximum likelihood estimation (MLE) and the least square estimation (LSE) for the constrained time series factor model. Compared to MLE, PCE does not require assumption for the distribution of noise and is robust to cross-sectional and serial correlation of idiosyncratic noise. Lam et al. (2011) and Lam and Yao (2012) adopted the information from the auto-covariance matrices at nonzero lags, which was empirically at least as good as PCE when the data were serially correlated. Therefore, motivated by Lam et al. (2011) and Lam and Yao (2012), we propose to estimate the constrained time series factor model based on the eigen-analysis of the non-negative definite matrix constructed by the lagged auto-covariance matrices. It is proved that, compared with the results of Lam et al. (2011) and Lam and Yao (2012), adding the constraint information can achieve the optimal rate of convergence even in the weak factor case. Moreover, unlike the work for the time series factor model in Chan et al. (2017), we can establish the asymptotic normality of our estimated factors without distributional assumption for noises. Also our approach is different from new works on high-dimensional factor analysis by Gao and Tsay (2022), Gao and Tsay (2023) and Choi and Yuan (2024).

The contents of this article is organized as follows: In Section 2, we introduce the constrained factor model and discuss the issues of model identification. Section 3 outlines the procedure for estimating the loading matrix and factors of the constrained time series factor model. In Section 4, we impose some regularity conditions on the constrained time series factor model and derive the convergence rate and asymptotic normality of the proposed estimator. The simulation studies and a real data example are presented in Sections 5 and 6, respectively. Technical details are relegated to the appendix.

Before going to mention our methodology, we need to introduce some notations that will be used in this paper. For any matrix $A$, $\|A\|_2$ and $\|A\|_{\min}$, denote the positive square root of the maximum non-zero eigenvalue and the minimum non-zero eigenvalue of $A'A$, respectively. $\|A\|_F = \sqrt{\operatorname{tr}(A'A)}$ means the Frobenius norm of matrix $A$. Note that, for any vector $a$, $\|a\|_2 = \|a\|_{\min} = \|a\|_F = \sqrt{a'a}$. For any $a, b$, $a \asymp b$ denotes $a = O_P(b)$ and $b = O_P(a)$.

## 2 Model and Identification Issues

Let $y_t = (y_{1,t}, \cdots, y_{N,t})'$ be an $N$-dimensional vector of observed time series for $1 \le t \le T$, where $T$ is the time length. The time series factor model of Lam et al. (2011) assumes that $y_t$ can be decomposed into the sum of two parts: a common component driven by a low-dimensional process and a static idiosyncratic component which is a white noise process. Referring to the idea of Lam et al. (2011), we study the constrained time series factor model under this new line as follows:

$$y_t = HAx_t + \varepsilon_t, \tag{1}$$

where $H$ is a prespecified $N \times m$ known constraint matrix with rank $m$; $A$ is an $m \times r$ unknown factor loading matrix which reflects the importance of common factors and their interactions with observed variables; $x_t$ is an $r$-dimensional vector of unobserved common factor time series which is weakly stationary with finite first two moments; $r$ is the unknown number of factor and $r \le m < N$; $\varepsilon_t$ is an $N$-dimensional vector white-noise process with mean 0 and covariance matrix $\Sigma_\varepsilon$.

**Remark 2.1.** The constraint matrix $H$ can be constructed based on substantive information of the observed variables or some empirical procedures. It incorporates the inherent data structure in practice. For example, if $H$ is a block diagonal matrix, or its columns consist of binary vectors, it represents a classification of the observed variables into groups.

Similar to the other factor models, the identification issue also exists in the constrained factor models. Specifically, the factor loading matrix and common factors are not unique. Let $V$ be an invertible $r \times r$ matrix. We can use pair $\left(AV^{-1}, Vx_t\right)$ instead of $(A, x_t)$, and such a transformation is equivalent to model (1). However, the columns space of the loading matrix $A$ is uniquely determined. We denote the factor loading space by $\mathcal{M}(A)$, and the identifiability of model (1) can be achieved in terms of the factor loading space $\mathcal{M}(A)$.

# 3 Estimation Procedure

We discuss the estimation procedure separately in two cases as Tsai and Tsay (2010): the case when the constraint matrix is semi-orthogonal, and the case when it is not semi-orthogonal.

## 3.1 Semi-orthogonal Constraint Matrix

Note that the semi-orthogonal constraint matrix $H$ satisfies $H'H = I_m$. We multiply both sides of model (1) by the matrix $H'$ and define $z_t = H'y_t$ and $u_t = H'\varepsilon_t$. This leads the model to

$$z_t = Ax_t + u_t. \tag{2}$$

This transformation projects the observed time series into the constrained space and makes the analysis more efficient. Next we use the QR decomposition $A = QR$, where $Q$ is a $m \times r$ semi-orthogonal matrix with orthogonal columns, and $R$ is an $r \times r$ full-rank upper triangular matrix. Since $A$ and $Q$ share the same column spaces, estimation of $\widehat{A}$ is equivalent to estimation of $\widehat{Q}$. Let $f_t = Rx_t$, then $f_t$ can be regarded as the factor corresponding to $Q$ and an estimate $\widehat{f_t}$ can be regarded as an estimate $\widehat{x_t}$. Thus model (2) can be re-expressed as

$$z_t = Qf_t + u_t, \tag{3}$$

where $Q'Q = I_r$. From model (3), for any $k \geq 1$, we have

$$\begin{aligned}\Sigma_z(k) &= Q\Sigma_f(k)Q' + Q\Sigma_{f,u}(k) + \Sigma_{u,f}(k)Q' + \Sigma_u(k) \\ &= Q\Sigma_f(k)Q' + Q\Sigma_{f,u}(k).\end{aligned}$$

The last equality is due to $u_t$ is still a white noise and $\Sigma_{u,f}(k) = H'\Sigma_{\varepsilon,x}(k)R' = 0$ from Assumption 4.1-(f) in Section 4. Then, define a non-negative definite symmetric matrix $M$ as follows,

$$M = \sum_{k=1}^{k_0} \Sigma_z(k)\Sigma_z(k)' = Q\left\{\sum_{k=1}^{k_0}\left(\Sigma_f(k)Q' + \Sigma_{f,u}(k)\right)\left(\Sigma_f(k)Q' + \Sigma_{f,u}(k)\right)'\right\}Q', \quad (4)$$

where $k_0$ is a predescribed positive integer. In practice, the choice of $k_0$ is not sensitive for the estimation empirically. Let $B$ be the $m \times (m-r)$ orthogonal complement of $Q$ satisfying $Q'B = 0$, and then $MB = 0$. This leads to that the factor loading space $\mathcal{M}(Q)$ is spanned by the eigenvectors of $M$ corresponding to its non-zero eigenvalues, and the number of the non-zero eigenvalues is $r$. Write eigenvalues of $M$ in descending order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_m = 0$, and the eigenvector corresponding to the $i$-th largest eigenvalue $\lambda_i$ is denoted by $q_i$. Then $Q$ is uniquely defined by $Q = (q_1, q_2, \cdots, q_r)$ up to a sign change as a representative of $\mathcal{M}(Q)$.

Define the sample version of $M$ as

$$\widehat{M} = \sum_{k=1}^{k_0} \widehat{\Sigma}_z(k)\widehat{\Sigma}_z(k)', \quad (5)$$

where $\widehat{\Sigma}_z(k) = (T-k)^{-1}\sum_{t=1}^{T-k}(z_{t+k} - \bar{z})(z_t - \bar{z})'$, $\bar{z} = T^{-1}\sum_{t=1}^{T} z_t$. Then, the loading matrix can be estimated by $\widehat{Q} = (\hat{q}_1, \hat{q}_2, \cdots, \hat{q}_r)$, where $\hat{q}_i$ is an eigenvector corresponding to its $i$-th largest eigenvalue of $\widehat{M}$. Consequently, the estimated factors and residuals can be obtained by

$$\widehat{f}_t = \widehat{Q}'z_t = \widehat{Q}'H'y_t, \quad \widehat{\varepsilon}_t = y_t - H\widehat{Q}\widehat{f}_t = \left(I - H\widehat{Q}\widehat{Q}'H'\right)y_t. \quad (6)$$

Now we need to determine the factor number $\widehat{r}$. In practice, the last $m - r$ eigenvalues

$\widehat{\lambda}_{r+1}, \widehat{\lambda}_{r+2}, \cdots, \widehat{\lambda}_m$ of $\widehat{M}$ may not be exactly 0, we cannot directly determine $\widehat{r}$ by counting the number of zero eigenvalues. Here we can use the eigenvalue-based estimator to determine the number of factors. The factor number $r$ can be estimated as the minimizer of the following ratio:

$$\widehat{r} = \arg \min_{1 \le k \le r_{\max}} \frac{\widehat{\lambda}_{k+1}}{\widehat{\lambda}_k}, \tag{7}$$

where $r_{\max}$ is an integer with $r \le r_{\max} \le m$. There is a large body of literature that estimates the number of factors through different forms of the ratio-based method. See Lam and Yao (2012), Ahn and Horenstein (2013) and Xia et al. (2015) among others.

### 3.2 Non-semi-orthogonal Constraint Matrix

If the constraint matrix $H$ is not semi-orthogonal, we perform QR decomposition on $H$ and obtain $H = H_Q H_R$, where $H_Q$ is a $N \times m$ semi-orthogonal matrix, $H_R$ is a $m \times m$ non-singular matrix. Let $A^* = H_R A$. Non-orthogonal constrained factor model can be transformed into orthogonal constrained model as follows

$$y_t = H_Q A^* x_t + \varepsilon_t.$$

Then, we apply the same estimation method in Section 3.1 to obtain $\widehat{Q^*}$ as the representative of $\mathcal{M}(\widehat{A^*})$. Note that $H_R$ is an invertible matrix, and the estimator of $A$ is $\widehat{A} = H_R^{-1}\widehat{Q^*}$.

## 4 Asymptotic Theory

In this Section, we present the rate of convergence for the loading matrix estimator $\widehat{Q}$, and establish the asymptotic normality for the proposed estimator $\widehat{f}_t$ of factors, as $N$, $m$ and $T$ tend to infinity while $r$ is fixed and known. It goes without saying explicitly that we may replace some $\widehat{q}_j$ by $-\widehat{q}_j$ in order to match the direction of $q_j$. For vector time series $x_t$ and $\varepsilon_t$, denote the auto-covariance matrix of $x_t$ and the cross-covariance matrix between $x_t$ and $\varepsilon_t$ by $\Sigma_x(k) = \operatorname{cov}(x_{t+k}, x_t)$ and $\Sigma_{x,\varepsilon}(k) = \operatorname{cov}(x_{t+k}, \varepsilon_t)$ respectively, where $k \ge 0$. To facilitate statistical inference, we assume some technical conditions for model (1).

**Assumption 4.1.** (a) No linear combination of the components of $x_t$ is white noise.

(b) $HA = (\beta_1, \beta_2, \cdots, \beta_r)$, such that $\|\beta_i\|_2^2 \asymp N^{1-\delta}, i = 1, 2, \cdots, r, \delta \in [0, 1]$ is a constant.

(c) For each $\beta_i$ and $\delta$ given in (b), $\min_{\theta_j, j \neq i} ||\beta_i - \sum_{j \neq i} \theta_j \beta_j||_2^2 \asymp N^{1-\delta}$.

(d) The time series $\{y_t\}$ is strictly stationary and $\psi$-mixing with the mixing coefficients $\psi(\cdot)$ satisfying the condition that $\sum_{t \geq 1} t\psi(t)^{1/2} < \infty$. Furthermore, $E\left(y_t^4\right) < \infty$ elementwisely.

(e) For $k = 0, 1, 2, \cdots, k_0$, $\Sigma_x(k)$ is full-ranked and $\|\Sigma_x(k)\|_2 \asymp 1 \asymp \|\Sigma_x(k)\|_{\min}$.

(f) The maximum eigenvalue of $\Sigma_\varepsilon$ remains bounded as $N$ increase to infinity. For all $k \geq 1$, $\text{cov}(\varepsilon_{t+k}, x_t) = 0$ and $\|\Sigma_{x,\varepsilon}(k)\|_2 = o_P\left(N^{\frac{1-\delta}{2}}\right)$.

(g) $H$ is a semi-orthogonal matrix that satisfies $H'H = I_m$.

**Remark 4.1.** Similar to Lam et al. (2011), the above assumptions are regularity conditions for the constrained factor time series model. The idiosyncratic noise is assumed to be white noise, allowing for strong cross-sectional correlation. However the variance matrix of the idiosyncratic noise was diagonal in Tsai and Tsay (2010). Assumption 4.1-(a) indicates that no linear combination of the components of $x_t$ are absorbed into $\varepsilon_t$. Assumption 4.1-(b) defines $\delta$ to be the strength of the factors. As $\delta = 0$, the factor is called to be strong, which means the factor can affect all observable variables; as $\delta > 0$, the factor is called to be weak, which means the factor only affects some observable variables. Comparing with Tsai and Tsay (2010), Assumption 4.1-(b) defines the strong and weak factors in the proposed model, which together with Assumption 4.1-(c) ensure that all $r$ factors in the model are of the equal strength. In the case of varying factor strengths, the estimation problem of the constrained model will become very complex, and we leave it for future work. Readers can refer to the following two articles. Lam and Yao (2012) investigated models with two different factor strengths using a two-step estimation method. Li et al. (2017) studied models with arbitrary factor strengths from the perspectives of singular values and matrix perturbation theory. Assumption 4.1-(e) illustrates the non-negative definite matrix $M$ defined in (4) has $r$ positive eigenvalues. Assumptions 4.1-(d) and (f) relax the condition of Tsai and Tsay (2010). The $\psi$-mixing condition can imply $\text{cov}\big(\text{vec}(z_t z_t'), \text{vec}(z_{t-j} z_{t-j}')\big) < \infty$ directly, where $z_t = (\varepsilon_t', x_t')'$ in Tsai and Tsay (2010). Assumption 4.1-(f) relaxes the condition of Tsai and

Tsay (2010) and stipulates a weak correlation between $x_t$ and $\varepsilon_t$, which makes the first term of $\Sigma_f(k)\,Q' + \Sigma_{f,u}(k)$ in (4) to be a dominating part. See Lemma A.4 and Lemma A.5 in the supplementary material for details. Assumption 4.1-(g) just follows Section 3.

**Theorem 4.1.** Suppose that Assumption 4.1 holds. If $mN^{-1+\delta}T^{-1/2} = o(1)$, then

$$\|\widehat{Q} - Q\|_2 = O_P\left(\max\left(T^{-\frac{1}{2}}, \frac{m}{N^{1-\delta}}T^{-\frac{1}{2}}\right)\right).$$

Theorem 4.1 shows that we can achieve the optimal rate $O_P\left(T^{-1/2}\right)$, when $m = O\left(N^{1-\delta}\right)$. This is a different asymptotic property between constrained and unconstrained time series factor models. Specially, as $N$ and $T$ increase, and $m$ is fixed, the convergence rate is still the optimal rate regardless of the strength of the factors. In fact, it reveals a fact that increasing $N$ dimensions while keeping $m$ fixed is equivalent to increasing the sample points in the constrained space.

The ratio $m \,/\, N^{1-\delta}$ in convergence rate can be interpreted as the noise-signal ratio in Tsai and Tsay (2010). When $H$ is orthogonal, each element of $A$ on average increases $\sqrt{N^{1-\delta} \,/\, m}$, but each element of the transformed error $u_t$ remains bounded under Assumption 4.1-(f) (refer to Lemma A.1 for details). The smaller the noise-signal ratio, the faster the convergence rate. For example, let $m = N^c$ with positive constant $c$. Then the convergence rate is $N^{\delta+c-1}T^{-1/2}$. Thus, we achieve a better rate than that of the unconstrained case if $c \le 1$.

As noted in Section 2, the factor loading matrix $Q$ is only identifiable up to a linear space spanned by its columns. The $\widehat{Q}$ obtained from the eigen-vectors of $\widehat{M}$ may not be the same as the one defined in the proof of Theorem 4.1. For two orthogonal matrices $\widehat{Q}$ and $Q$, define

$$D\left(\widehat{Q}, Q\right) = \sqrt{1 - \frac{1}{m}\mathrm{Tr}\left(\widehat{Q}\widehat{Q}'QQ'\right)}.$$

Clearly, $D\left(\widehat{Q}, Q\right) \in [0,1]$. $D\left(\widehat{Q}, Q\right) = 0$ if and only if $\mathcal{M}\left(\widehat{Q}\right) = \mathcal{M}(Q)$ and $D\left(\widehat{Q}, Q\right) = 1$ if and only if $\mathcal{M}\left(\widehat{Q}\right) \perp \mathcal{M}(Q)$. Thus, we adopt $D(\widehat{Q}, Q)$ to measure the difference between the two linear spaces $\mathcal{M}\left(\widehat{Q}\right)$ and $\mathcal{M}(Q)$. Then, the following corollary shows that the convergence rate still holds in this discrepancy measure.

**Corollary 4.1.** Under the conditions of Theorem 4.1, it holds that

$$D\left(\widehat{Q}, Q\right) = O_P\left(\max\left(T^{-\frac{1}{2}}, \frac{m}{N^{1-\delta}}T^{-\frac{1}{2}}\right)\right).$$

Note that, from the proof of Corollary 4.1, we have $D\left(\widehat{Q}, Q\right) \leq \|\widehat{Q} - Q\|_2$.

**Theorem 4.2.** Suppose that Assumption 4.1 holds, and $mN^{-1+\delta}T^{-1/2} = o(1)$. Then, it follows that

$$(i)\ N^{-\frac{1}{2}}\left\|\widehat{Q}\widehat{f}_t - Ax_t\right\|_2 = O_P\left(\left\|\widehat{Q} - Q\right\|_2 N^{-\frac{\delta}{2}} + N^{-\frac{1}{2}}\right), \text{ and}$$
$$(ii)\ N^{-\frac{1}{2}}\left\|\widehat{f}_t - f_t\right\|_2 = O_P\left(\left\|\widehat{Q} - Q\right\|_2 N^{-\frac{\delta}{2}} + N^{-\frac{1}{2}}\right).$$

Theorem 4.2 illustrates that as long as $m$ increases slower than $N$, we get a faster convergence rate $O_P\left(mN^{-1+\delta/2}T^{-1/2} + N^{-1/2}\right)$ than the counterpart in factor time series model of Lam et al. (2011). When all factors are strong, we can reach the optimal rate $O_P\left(T^{-1/2} + N^{-1/2}\right)$ specified in Bai (2003). It is easy to see that the estimation of the loading spaces is consistent with fixed $N$ and $m$ in Theorem 4.1. However, the consistency of the signal estimator requires that $N$ tends to infinity.

To obtain the asymptotic normality of the estimated factors, the following two additional conditions are needed.

**Assumption 4.2.** (a) $\varepsilon_{j,t}$ are independent for different $t$ and $j$, and $E\left(\varepsilon_{j,t}\right) = 0$, $E\left(\varepsilon_{j,t}^2\right) = \sigma_j^2 < \infty$ and $E\left(|\varepsilon_{j,t}|^3\right) = o_P\left(N^{1/2}\right)$ for all $1 \leq j \leq N$ and $t \geq 1$.

(b) Let $v_{i,j}$ be the $(i,j)$-element of matrix $HQ$ and $|v_{i,j}| = O\left(N^{-1/2}\right)$. The limit

$$\lim_{N\to\infty}\sum_{k=1}^{N} v_{k,i}v_{k,j}\sigma_k^2 = \Omega_{i,j}$$

exists.

Now we present the asymptotic normality of $\hat{f}_t$.

**Theorem 4.3.** Suppose that Assumptions 4.1 and 4.2 hold, $mN^{-\frac{1-\delta}{2}}T^{-\frac{1}{2}} = o(1)$ and

$N^{\frac{1-\delta}{2}}T^{-\frac{1}{2}} = o(1)$. Then, as $N \to \infty$ and $T \to \infty$,

$$\widehat{f}_t - f_t \xrightarrow{\text{d}} N(0, \Omega),$$

where $\Omega = (\Omega_{i,j})_{r\times r}$.

Since $HQ$ is a semi-orthogonal matrix, we have $\sum_{k=1}^{N} v_{k,j}^2 = 1$. From an average perspective, each element $|v_{i,j}|$ has an order of $O\left(N^{-1/2}\right)$. Theorem 4.3 shows that when the components of $\varepsilon_t$ are independent and the growth rate of the third-order moment is much smaller than $\sqrt{N}$. The error of estimated factors follows a normal distribution approximately regardless of the distribution of $\varepsilon_t$, as $N, T$ tend to infinity.

Finally, we deal with the convergence rates of the estimated eigenvalues.

**Theorem 4.4.** If Assumption 4.1 holds, and $mN^{-1+\delta}T^{-1/2} = o(1)$, then

$$\left|\widehat{\lambda}_i - \lambda_i\right| = O_P\left(N^{2-2\delta}T^{-\frac{1}{2}} + N^{1-\delta}mT^{-\frac{1}{2}}\right), \text{ for } i = 1, 2, \cdots, r, \text{ and}$$
$$\left|\widehat{\lambda}_i\right| = O_p\left(N^{2-2\delta}T^{-1} + m^2T^{-1}\right), \text{ for } i = r+1, r+2, \cdots, m.$$

The corollary 4.2 below follows immediately from Theorem 4.4 and the fact that $\lambda_j \asymp N^{2-2\delta}$, for $j = 1, \cdots, r$ (see Lemma A.5 for details).

**Corollary 4.2.** Under the conditions of Theorem 4.4, it holds that

$$\frac{\widehat{\lambda}_{i+1}}{\widehat{\lambda}_i} \asymp 1, \text{ for } i = 1, 2, \cdots, r-1, \text{ and } \frac{\widehat{\lambda}_{r+1}}{\widehat{\lambda}_r} = O_P\left(\max\left(T^{-1}, \frac{m^2}{N^{2-2\delta}}T^{-1}\right)\right).$$

Compared with the results of Theorem 1 in Lam and Yao (2012), the convergence rate of the estimated eigenvalues for the constrained time series factor model in Theorem 4.4 is faster than that of the unconstrained factor model. Corollary 1 implies that the ratios of eigenvalue drop more sharply and quickly than the counterpart in Lam and Yao (2012) at $i = r$. This provides a useful theoretical underpinning for the estimator $\widehat{r}$ defined in (7).

# 5 Simulation Experiments

## 5.1 Estimation for Loading Matrix

In this subsection, we carry out the simulation experiments to compare the accuracy of the proposed estimation for the loading matrix between the unconstrained and constrained factor model when the number of factors is known and fixed. In the simulations, the estimation of unconstrained factor model follows the method in Lam et al. (2011). The data are generated from the model (1). The elements of constraint matrix $H$ and loading matrix $A$ are uniformly distributed on the interval $[-3, 3]$, and we divide each of them by $N^{-\delta/2}$ to make all the factors the same strength with $\delta$. The factor $x_t$ follows a VAR(1) process, i.e.,

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} = \begin{pmatrix} 0.5 & 0.1 & 0.1 \\ 0.1 & 0.5 & 0.1 \\ 0.1 & 0.1 & 0.5 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \\ x_{3,t-1} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ e_{2,t} \\ e_{3,t} \end{pmatrix},$$

where $e_t = (e_{1,t}, e_{2,t}, e_{3,t})'$ obeys a multivariate normal distribution and a multivariate $t$-distribution, respectively, as follows:

$$\text{Case 1: } e_{i,t} \sim N(0, 1);$$
$$\text{Case 2: } e_{i,t} \sim t(3).$$

The idiosyncratic noise $\varepsilon_t$ follows a time-independent multivariate normal distribution with mean 0 and variance matrix $\Sigma_\varepsilon = (\sigma_{ij})$, where $\sigma_{ij}$ is defined as

$$\sigma_{ij} = \frac{1}{2}\left\{(|i-j|+1)^{2\times 0.9} - 2\,|i-j|^{2\times 0.9} + (|i-j|-1)^{2\times 0.9}\right\}.$$

Setting $k_0 = 1$ is similar to Lam et al. (2011). Following the estimation procedure in section 3, we transform the constraint matrix and loading matrix to the orthogonal matrices. Finally, we compute the estimation error $D\left(\widehat{Q}, Q\right)$ for the combinations of $N = 100, 200, 400, 800$, $T = 100, 200, 400, 800$ and $m = 80$. The factor number $r = 3$ is known, and all of them have the same strength with two cases, i.e., $\delta = 0$ and $\delta = 0.5$, respectively. For each setting, we replicate 200 times and calculate the mean and standard errors of $D\left(\widehat{Q}, Q\right)$, which are listed in Table 1 below. The results of the constrained and unconstrained factor models are presented on the left and right sides of the symbol $\|$, respectively. As can be seen from

Table 1, whether the factors are strong or weak, $D\left(\widehat{Q},Q\right)$ of the constrained factor model outperforms its counterpart of the unconstrained factor model. In addition, when $T$ is fixed and $N$ increases, $D\left(\widehat{Q},Q\right)$ of both the constrained and unconstrained factor models remains essentially unchanged for strong factors. For weak factors, $D\left(\widehat{Q},Q\right)$ of the unconstrained factor model gradually increases, while $D\left(\widehat{Q},Q\right)$ of the constrained factor model gradually decreases. This result is consistent with the conclusion of Theorem 4.1 in Section 4 and Theorem 1 in Lam et al. (2011).

Table 1: Mean and standard errors (in brackets) of the estimation error $D\left(\widehat{Q},Q\right)$ when $m=80$ for 200 replications of simulation. The values shown are the true values multiplied by 1000.

| | $D\left(\widehat{Q},Q\right)$ | | $T=100$ | $T=200$ | $T=400$ | $T=800$ |
|---|---|---|---|---|---|---|
| Case 1 | $\delta=0$ | $N=100$ | $14_{(7)}\|\|16_{(7)}$ | $7_{(2)}\|\|8_{(3)}$ | $5_{(2)}\|\|5_{(2)}$ | $3_{(1)}\|\|4_{(1)}$ |
| | | $N=200$ | $10_{(6)}\|\|16_{(10)}$ | $6_{(2)}\|\|9_{(3)}$ | $3_{(1)}\|\|5_{(2)}$ | $2_{(1)}\|\|4_{(1)}$ |
| | | $N=400$ | $8_{(11)}\|\|18_{(22)}$ | $4_{(1)}\|\|9_{(2)}$ | $2_{(1)}\|\|5_{(1)}$ | $2_{(0)}\|\|4_{(1)}$ |
| | | $N=800$ | $5_{(5)}\|\|17_{(17)}$ | $3_{(1)}\|\|9_{(3)}$ | $2_{(0)}\|\|5_{(1)}$ | $1_{(0)}\|\|4_{(1)}$ |
| | $\delta=0.5$ | $N=100$ | $43_{(20)}\|\|49_{(23)}$ | $23_{(7)}\|\|26_{(8)}$ | $15_{(4)}\|\|17_{(5)}$ | $10_{(3)}\|\|12_{(4)}$ |
| | | $N=200$ | $35_{(14)}\|\|55_{(22)}$ | $20_{(5)}\|\|31_{(8)}$ | $12_{(3)}\|\|20_{(5)}$ | $8_{(2)}\|\|13_{(3)}$ |
| | | $N=400$ | $31_{(23)}\|\|70_{(43)}$ | $17_{(5)}\|\|37_{(12)}$ | $10_{(2)}\|\|22_{(5)}$ | $7_{(1)}\|\|16_{(4)}$ |
| | | $N=800$ | $31_{(42)}\|\|89_{(73)}$ | $15_{(5)}\|\|46_{(17)}$ | $9_{(2)}\|\|27_{(7)}$ | $6_{(1)}\|\|19_{(3)}$ |
| Case 2 | $\delta=0$ | $N=100$ | $16_{(70)}\|\|17_{(70)}$ | $5_{(3)}\|\|5_{(3)}$ | $3_{(1)}\|\|3_{(1)}$ | $2_{(0)}\|\|2_{(1)}$ |
| | | $N=200$ | $11_{(38)}\|\|13_{(22)}$ | $3_{(2)}\|\|5_{(2)}$ | $2_{(1)}\|\|3_{(1)}$ | $1_{(0)}\|\|2_{(1)}$ |
| | | $N=400$ | $4_{(2)}\|\|9_{(4)}$ | $2_{(1)}\|\|5_{(2)}$ | $1_{(0)}\|\|3_{(1)}$ | $1_{(0)}\|\|2_{(1)}$ |
| | | $N=800$ | $3_{(2)}\|\|10_{(5)}$ | $2_{(1)}\|\|5_{(2)}$ | $1_{(0)}\|\|3_{(1)}$ | $1_{(0)}\|\|2_{(0)}$ |
| | $\delta=0.5$ | $N=100$ | $34_{(52)}\|\|39_{(60)}$ | $15_{(6)}\|\|17_{(7)}$ | $9_{(3)}\|\|10_{(3)}$ | $6_{(1)}\|\|6_{(1)}$ |
| | | $N=200$ | $22_{(12)}\|\|35_{(18)}$ | $12_{(4)}\|\|20_{(6)}$ | $8_{(2)}\|\|12_{(3)}$ | $5_{(1)}\|\|8_{(2)}$ |
| | | $N=400$ | $24_{(31)}\|\|51_{(55)}$ | $10_{(4)}\|\|23_{(8)}$ | $6_{(2)}\|\|15_{(4)}$ | $4_{(1)}\|\|10_{(2)}$ |
| | | $N=800$ | $16_{(6)}\|\|50_{(19)}$ | $9_{(2)}\|\|28_{(7)}$ | $5_{(1)}\|\|17_{(4)}$ | $3_{(1)}\|\|11_{(3)}$ |

### 5.2 Asymptotic Normality for the Estimated Factors

Now we illustrate the asymptotic normality of the estimated factors through the simulation experiments. The data generation process is similar to that in Subsection 5.1. To make the constraint matrix sparse, $H$ is selected to be $H = I_m \otimes 1_{N/m}$, where $1_k$ is a column vector of $k$ ones. The components of idiosyncratic noise $\varepsilon_t$ are independent, and various settings of distribution of each noise component are considered as follows, including Student-t distribution, Chi-square distribution and Exponential distribution. For $i = 1, 2, \cdots, N$ and $t = 1, 2, \cdots, T$,

$$\text{Case 1: } \varepsilon_{i,t} \sim t(\theta_i),\ \theta_i \sim U(3,5);$$
$$\text{Case 2: } \varepsilon_{i,t} \sim \chi^2(\theta_i),\ \theta_i \sim U(3,5);$$
$$\text{Case 3: } \varepsilon_{i,t} \sim \text{Exp}(\theta_i),\ \theta_i \sim U(3,5).$$

According to Theorem 4.3, the factor error $\widehat{f}_t - f_t$ should approximately follows a normal distribution. Under the setting for $m = 50, N = 500, T = 1000$ and $r = 3$, we explore the distribution of each normalized factor error $\widehat{\Sigma}_f^{-\frac{1}{2}}\left(\widehat{f}_t - f_t\right)$, where $\widehat{\Sigma}_f$ is the sample variance matrix of $\widehat{f}_t - f_t$. We conduct the Shapiro-Wilks test, one-sample Kolmogorov-Smirnov test, Cramer-Von Mises test and Anderson Darling test for the normality testing for each factor component. With 100 simulations for $\delta = 0$ and $\delta = 0.5$ respectively, Table 2 reports the relative frequency that the null hypothesis of standard normal distribution can not be rejected at significance level 5% for three cases. Figures 1 and 2 depict the histogram and Quantile-Quantile (QQ) plot for the strong and weak factors in Case 1, respectively. The histograms and QQ plots for the other two cases are similar, and we will not present them here. The results indicate that the estimated factors are conformed to follow a normal distribution roughly, regardless of the strength or weakness of the factors.

### 5.3 Prediction Performance

We compare the performance of proposed method in Section 3 with PCE in Bai (2003) and MLE in Tsai and Tsay (2010) when the number of factors is known. The data generation process is the same as in Subsection 5.1. For each combination of $(N, T)$, the simulations are replicated 200 times. The mean and standard deviation of the root-mean-square error

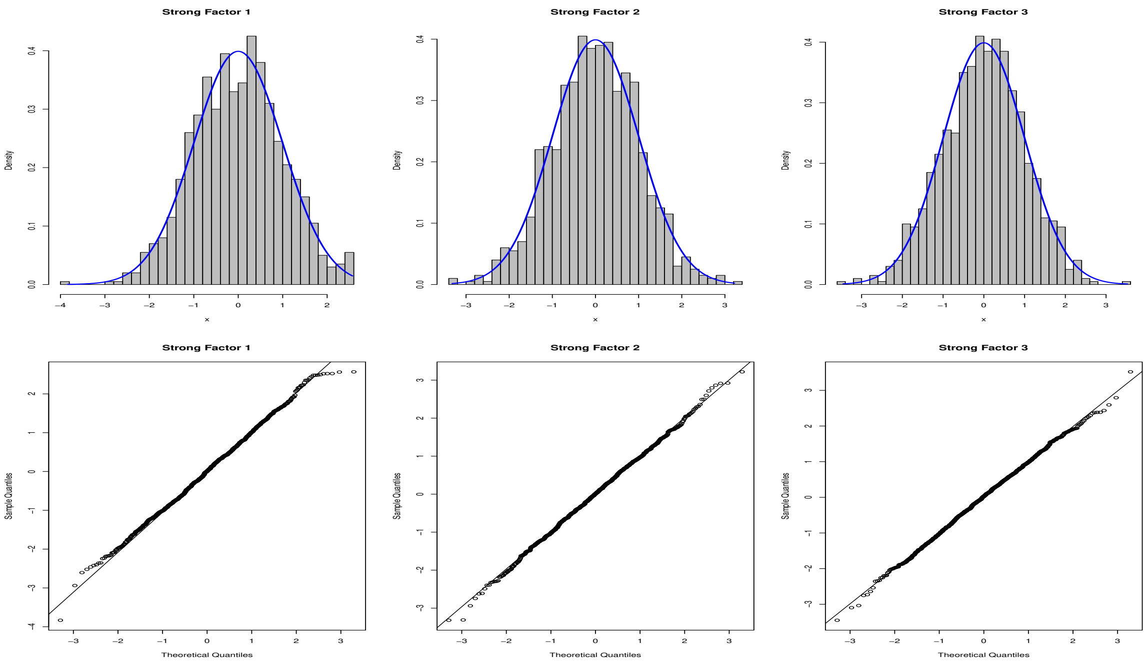


Figure 1: The histograms and QQ plots of three strong factors ($\delta = 0$) for Case 1. The upper three panels are the histograms, while the lower three panels are QQ plots.

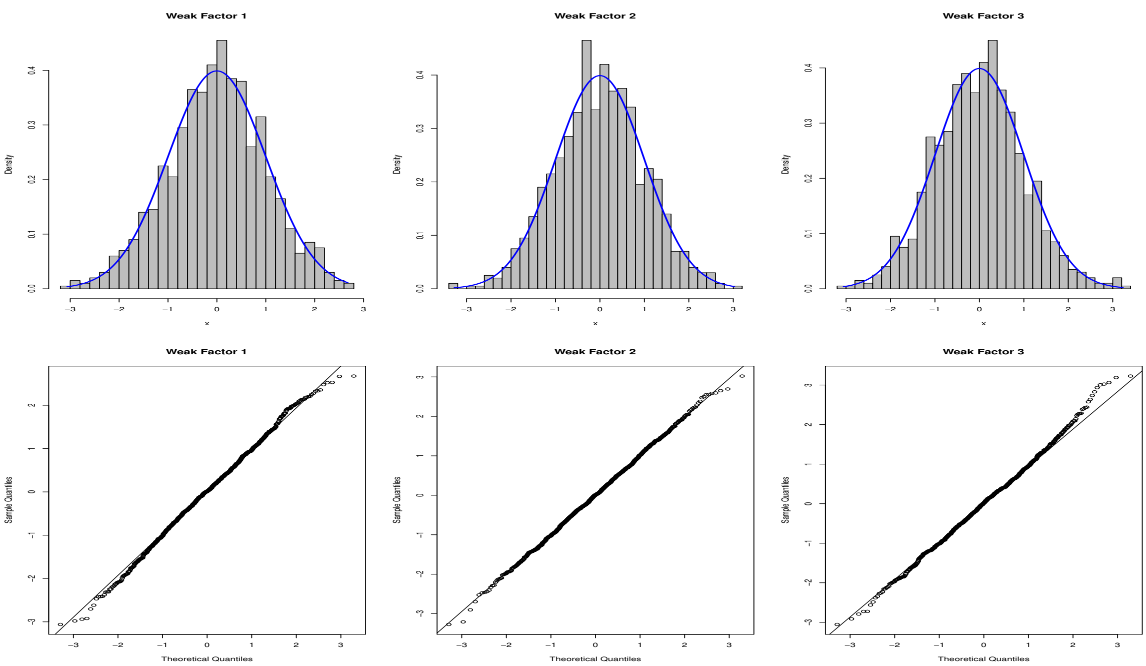


Figure 2: The histograms and QQ plots of three weak factors ($\delta = 0.5$) for Case 1. The upper three panels are the histograms, while the lower three panels are QQ plots.

(RMSE), i.e.,

$$\mathrm{RMSE} = \left( \frac{1}{NT} \sum\nolimits_{t=1}^{T} \left\| H\widehat{Q}\widehat{f}_t - HAx_t \right\|_2^2 \right)^{1/2}$$

are calculated. The results are listed in Table 3, which indicates that both PCE and MLE have better performance than our method in terms of RMSE when strong factors are present. This suggests that the covariance matrix contains more informative content of factors compared to the lagged covariance matrix. However, in the case of weak factors, MLE performs the worst in terms of RMSE. This is because the MLE in Tsai and Tsay (2010) requires the covariance matrix of the noise to be a diagonal matrix, implying independence of each idiosyncratic noise. However, in this case, the noise exhibits strong cross-sectional correlation, introducing significant interference for MLE to disentangle weak factors from observable variables. The strong cross-sectional correlation of the noise severely compromises the predictive capability of MLE in the weak factor model. In contrast, our method outperforms other methods in this case. From these observations, it can be concluded that PCE and MLE are more sensitive to the factor strength compared to our method. Remarkably, our method exhibits little change in RMSE when the factor strength is from 0 to 0.5, especially in the case of large dimensions ($N = 800$). However, the RMSE of other two methods almost doubles with changing factor strength from 0 to 0.5.

### 5.4 Estimation for the Number of Factors

When the number of factors is unknown, we are exploring the performance of the ratio-based estimator mentioned in Section 3 for estimating the number of factors in both unconstrained and constrained factor model. Let CER and ER denote the ratio-based estimator in constrained factor model and unconstrained factor model, respectively. The data generation process and parameter settings are identical to those in Subsection 5.1. For each combination of $(N, T)$, we replicate the simulation 200 times and calculate the frequency of correctly estimating the number of factors for both CER and ER in Table 4. From Table 4, in the case of large constraint dimension ($m = 80$) or strong factors ($\delta = 0$), we can observe that both ER and CER estimators have almost 100% frequency of correctly estimating the number of factors. However, in the case of both small constraint dimension ($m = 8$) and weak factors ($\delta = 0.5$), CER still has desirable performance, while ER performs relatively poorly especially for large $N = 800$ and small $T = 100$. Meanwhile, when $T$ is fixed and $N$ gradually increases, the frequency of correct estimation for the number of factors by ER estimator decreases, while CER estimator remains stable with accuracy over 96%. This implies that CER estimator is robust in both constraint dimension and factor strength. In short, when the idiosyncratic noise is white noise and there is serial correlation in the factors, the CER estimator performs well enough to determine the number of factors for the constrained factor models.

## 6 An Example on Analysis of Real Data

We now use the proposed model to analyze a real time series dataset. The dataset consists of 168 observations of monthly excess stock returns of ten U.S. companies from 1990 to 2003. These ten companies include: (1) Abbott Labs, (2) Eli Lilly, (3) Merck, (4) Pfizer, (5) Ford, (6) General Motors, (7) BP, (8) Chevron, (9) Royal Dutch, and (10) Exxon-Mobil. At the significance level of 0.05, Tsai and Tsay (2010) conducted likelihood ratio tests on six models with different grouping constraints on factor models, and concluded that the data were divided into three groups: (1-4), (5-6), and (7-10). This can be interpreted as the ten companies coming from three different industrial sectors: pharmaceutical industry, automotive industry and petroleum industry. In this case, $T = 168, N = 10, m = r = 3$, we use $H = (h_1, h_2, h_3)$ to indicate the variables for three industrial sector, that is, $h_1 = (1(4), 0(6))', h_2 = (0(4), 1(2), 0(4))', h_3 = (0(6), 1(4))'$. Table 5 shows the estimated loading matrices of the constrained and unconstrained factor models. From Table 5, it is evident that the three common factors differ across industrial sectors in the constrained factor model. Another difference between the constrained and unconstrained factor models is that the former uses 9 parameters whereas the latter has 30 parameters in the loading matrix. In this example, RMSEs for PCE, MLE and the proposed method for the constrained factor model are all 0.0388. Furthermore, we employ $\overline{\text{RMSE}}$ to represent the variation in RMSE between constrained and unconstrained factor models, i.e.,

$$\overline{\text{RMSE}} = \left| \frac{\text{RMSE}_\text{C} - \text{RMSE}_\text{U}}{\text{RMSE}_\text{U}} \right|,$$

where $\text{RMSE}_{\text{C}}$ is for the constrained factor model, while $\text{RMSE}_{\text{U}}$ is for the unconstrained one. The $\overline{\text{RMSE}}$ of PCE, MLE and the proposed method are 0.0486, 0.0458 and 0.074, respectively. This result indicates that the estimation accuracy of the three methods is nearly consistent. In contrast, our method is more suitable for this dataset analyzed by the constrained factor models.

# 7 Conclusion

This article proposes a new estimation method for the constrained factor models for high-dimensional time series by extending the estimation theory of Lam et al. (2011). We have provided the convergence rate of the proposed estimator. We assumed that the idiosyncratic noise of the constrained factor model is white noise with cross-sectional correlation and have weak correlation with the factors. Since the serial correlation of the observable sequence is provided by the common factors, the estimation of loading matrix in this article utilizes the eigen-analysis for lagged autocovariance and the number of factors is determined by the eigenvalue ratio method. The primary findings are as follows:

1. The convergence rate of the proposed estimator for factor loading matrix and eigenvalue is faster than that in Lam et al. (2011) and Lam and Yao (2012). When the constrained dimension $m$ remains unchanged, increasing the dimension $N$ does not lead to a higher estimation error for the loading matrix of the corresponding weak factors. Moreover, when $m = O\left(N^{1-\delta}\right)$, the optimal convergence rate $O\left(T^{-1/2}\right)$ can be achieved.

2. The asymptotic normality of the estimator for factors does not rely on the assumption in Chan et al. (2017) that the idiosyncratic noise follows a multivariate normal distribution. It is only required that the factor components are mutually independent and that the growth rate of their third moments is much slower than $T^{-1/2}$. As we showed in simulation of subsection 5.2, when the idiosyncratic noise follows chi-squared, t-, and exponential distributions, the factor estimation errors approximately follow a normal distribution.

3. Compared to the MLE employed in Tsai and Tsay (2010), our method demonstrates robustness to variations in factor strength. For the constrained weak factor models, our

approach outperforms the MLE in terms of estimation accuracy. Specifically, RMSE of MLE is significantly larger than that of our method. This is because when the idiosyncratic noise exhibits strong cross-sectional correlation, it interferes with the MLE's ability in identifying and estimating weak factors, whereas our method does not encounter such issues.

## Acknowledgments

The authors thank the editor-in-chief and the anonymous reviewers for their comments that significantly improved this work. The research of Qiang Xia was supported in part by the National Natural Science Foundation of China (No.12171161, 91746102) and the Natural Science Foundation of Guangdong Province of China (No.2022A1515011754).

## Supplementary Materials

The technical proofs of theoretical results are presented in the attached supplementary materials.

## References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics and Probability Letters*, 80:1806–1813.

Amengual, D. and Watson, M. (2007). Consistent estimation of the number of dynamic factors in a large n and t panel. *Journal of Business and Economic Statistics*, 25:91–96.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Chamberlain, G. and Rothschild, M. (1983). Arbitrary, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51:1281–1304.

Chan, N. H., Lu, Y., and Yau, C. Y. (2017). Factor modelling for high-dimensional time series: inference and model selection. *Journal of Time Series Analysis*, 38(2):285–307.

Choi, J. and Yuan, M. (2024). High dimensional factor analysis with weak factors. *arXiv:2402.05789v1 [econ.EM]*.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, 164:188–205.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B*, 75:603–680.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25:1319–1347.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics*, 82:540–554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic-factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100:830–840.

Gao, Z. and Tsay, R. S. (2022). Modeling high-dimensional time series: A factor model with dynamically dependent factors and diverging eigenvalues. *Journal of the American Statistical Association*, 117(539):1398–1414.

Gao, Z. and Tsay, R. S. (2023). A two-way transformed factor model for matrix-variate time series. *Econometrics and Statistics*, 27:83–101.

Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102:603–617.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40(2):694–726.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

Li, Z., Wang, Q., and Yao, J. (2017). Identifying the number of factors from singular values of a large sample auto-covariance matrix. *The Annals of Statistics*, 45(1):257–288.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77:1447–1479.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economic and Statistics*, 92:1004–1016.

Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, 95:356–379.

Stock, J. and Watson, M. (2005). *Implications of dynamic factor models for VAR analysis (No. w11467).* National Bureau of Economic Research.

Stock, J. and Watson, M. (2011). *Dynamic factor models. In Oxford handbook on economic forecasting.* Oxford University Press.

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.

Tsai, H. and Tsay, R. S. (2010). Constrained factor models. *Journal of the American Statistical Association*, 105(492):1593–1605.

Tsai, H., Tsay, R. S., Lin, E. M., and Cheng, C. W. (2016). Doubly constrained factor models with applications. *Statistica Sinica*, 26:1453–1478.

Wu, J. (2021). Estimation of high dimensional factor model with multiple threshold-type regime shifts. *Computational Statistics & Data Analysis*, 157:107153.

Xia, Q., Liang, R., and Wu, J. (2017). Transformed contribution ratio test for the number of factors in static approximate factor models. *Computational Statistics & Data Analysis*, 112:235–241.

Xia, Q., Xu, W., and Zhu, L. (2015). Consistently determining the number of factors in multivariate volatility modelling. *Statistica Sinica*, pages 1025–1044.

Zhang, B., Pan, G., Yao, Q., and Zhou, W. (2023). Factor modeling for clustering high-dimensional time series. *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2023.2183132.

Table 2: The empirical frequency of accepting the null hypothesis for the tests at 5% significance level with 100 replications.

| Normality test | | | Shapiro-Wilks | Kolmogorov-Smirnov | Cramer-Von Mises | Anderson Darling |
|---|---|---|---|---|---|---|
| Case 1 | $\delta = 0$ | $\widehat{\Sigma}_{f_1}^{-\frac{1}{2}}\left(\widehat{f}_{1,t} - f_{1,t}\right)$ | 0.83 | 1 | 0.92 | 0.87 |
| | | $\widehat{\Sigma}_{f_2}^{-\frac{1}{2}}\left(\widehat{f}_{2,t} - f_{2,t}\right)$ | 0.90 | 1 | 0.96 | 0.96 |
| | | $\widehat{\Sigma}_{f_3}^{-\frac{1}{2}}\left(\widehat{f}_{3,t} - f_{3,t}\right)$ | 0.90 | 1 | 0.94 | 0.94 |
| | $\delta = 0.5$ | $\widehat{\Sigma}_{f_1}^{-\frac{1}{2}}\left(\widehat{f}_{1,t} - f_{1,t}\right)$ | 0.88 | 1 | 0.94 | 0.83 |
| | | $\widehat{\Sigma}_{f_2}^{-\frac{1}{2}}\left(\widehat{f}_{2,t} - f_{2,t}\right)$ | 0.87 | 1 | 0.97 | 0.98 |
| | | $\widehat{\Sigma}_{f_3}^{-\frac{1}{2}}\left(\widehat{f}_{3,t} - f_{3,t}\right)$ | 0.93 | 1 | 0.93 | 0.92 |
| Case 2 | $\delta = 0$ | $\widehat{\Sigma}_{f_1}^{-\frac{1}{2}}\left(\widehat{f}_{1,t} - f_{1,t}\right)$ | 0.89 | 1 | 0.93 | 0.94 |
| | | $\widehat{\Sigma}_{f_2}^{-\frac{1}{2}}\left(\widehat{f}_{2,t} - f_{2,t}\right)$ | 0.88 | 1 | 0.91 | 0.87 |
| | | $\widehat{\Sigma}_{f_3}^{-\frac{1}{2}}\left(\widehat{f}_{3,t} - f_{3,t}\right)$ | 0.97 | 1 | 0.97 | 0.98 |
| | $\delta = 0.5$ | $\widehat{\Sigma}_{f_1}^{-\frac{1}{2}}\left(\widehat{f}_{1,t} - f_{1,t}\right)$ | 0.86 | 1 | 0.88 | 0.88 |
| | | $\widehat{\Sigma}_{f_2}^{-\frac{1}{2}}\left(\widehat{f}_{2,t} - f_{2,t}\right)$ | 0.96 | 1 | 0.94 | 0.95 |
| | | $\widehat{\Sigma}_{f_3}^{-\frac{1}{2}}\left(\widehat{f}_{3,t} - f_{3,t}\right)$ | 0.98 | 1 | 0.97 | 0.96 |
| Case 3 | $\delta = 0$ | $\widehat{\Sigma}_{f_1}^{-\frac{1}{2}}\left(\widehat{f}_{1,t} - f_{1,t}\right)$ | 0.87 | 1 | 0.84 | 0.85 |
| | | $\widehat{\Sigma}_{f_2}^{-\frac{1}{2}}\left(\widehat{f}_{2,t} - f_{2,t}\right)$ | 0.94 | 1 | 0.96 | 0.96 |
| | | $\widehat{\Sigma}_{f_3}^{-\frac{1}{2}}\left(\widehat{f}_{3,t} - f_{3,t}\right)$ | 0.96 | 1 | 0.94 | 0.94 |
| | $\delta = 0.5$ | $\widehat{\Sigma}_{f_1}^{-\frac{1}{2}}\left(\widehat{f}_{1,t} - f_{1,t}\right)$ | 0.87 | 1 | 0.91 | 0.89 |
| | | $\widehat{\Sigma}_{f_2}^{-\frac{1}{2}}\left(\widehat{f}_{2,t} - f_{2,t}\right)$ | 0.90 | 1 | 0.92 | 0.91 |
| | | $\widehat{\Sigma}_{f_3}^{-\frac{1}{2}}\left(\widehat{f}_{3,t} - f_{3,t}\right)$ | 0.93 | 1 | 0.96 | 0.96 |

Table 3: Mean and standard errors (in brackets) of RMSE for PCE, MLE and the proposed estimator.

| | | $\delta = 0$ | | | $\delta = 0.5$ | | |
|---|---|---|---|---|---|---|---|
| | RMSE | PCE | MLE | proposed estimator | PCE | MLE | proposed estimator |
| $T = 100$ | $N = 100$ | $275_{(50)}$ | $297_{(57)}$ | $436_{(124)}$ | $666_{(41)}$ | $679_{(39)}$ | $561_{(95)}$ |
| | $N = 200$ | $174_{(23)}$ | $188_{(29)}$ | $322_{(93)}$ | $457_{(44)}$ | $476_{(33)}$ | $346_{(81)}$ |
| | $N = 400$ | $116_{(12)}$ | $123_{(15)}$ | $225_{(59)}$ | $277_{(64)}$ | $347_{(24)}$ | $216_{(41)}$ |
| | $N = 800$ | $83_{(8)}$ | $83_{(8)}$ | $166_{(24)}$ | $141_{(47)}$ | $270_{(21)}$ | $142_{(27)}$ |
| $T = 200$ | $N = 100$ | $239_{(44)}$ | $248_{(48)}$ | $324_{(57)}$ | $662_{(31)}$ | $665_{(30)}$ | $433_{(110)}$ |
| | $N = 200$ | $153_{(20)}$ | $162_{(24)}$ | $224_{(38)}$ | $456_{(36)}$ | $476_{(22)}$ | $260_{(63)}$ |
| | $N = 400$ | $104_{(12)}$ | $108_{(14)}$ | $156_{(23)}$ | $276_{(59)}$ | $350_{(17)}$ | $160_{(25)}$ |
| | $N = 800$ | $73_{(7)}$ | $75_{(8)}$ | $111_{(17)}$ | $112_{(36)}$ | $269_{(18)}$ | $110_{(18)}$ |
| $T = 400$ | $N = 100$ | $223_{(42)}$ | $229_{(45)}$ | $252_{(37)}$ | $664_{(26)}$ | $666_{(25)}$ | $331_{(90)}$ |
| | $N = 200$ | $144_{(23)}$ | $148_{(26)}$ | $180_{(23)}$ | $458_{(31)}$ | $473_{(21)}$ | $191_{(31)}$ |
| | $N = 400$ | $97_{(11)}$ | $100_{(12)}$ | $125_{(13)}$ | $260_{(62)}$ | $349_{(15)}$ | $131_{(20)}$ |
| | $N = 800$ | $66_{(6)}$ | $67_{(6)}$ | $87_{(8)}$ | $105_{(32)}$ | $268_{(15)}$ | $90_{(11)}$ |
| $T = 800$ | $N = 100$ | $213_{(44)}$ | $218_{(46)}$ | $215_{(32)}$ | $665_{(22)}$ | $667_{(21)}$ | $246_{(56)}$ |
| | $N = 200$ | $135_{(21)}$ | $138_{(22)}$ | $150_{(17)}$ | $466_{(29)}$ | $479_{(20)}$ | $161_{(26)}$ |
| | $N = 400$ | $91_{(10)}$ | $93_{(10)}$ | $105_{(9)}$ | $261_{(64)}$ | $350_{(14)}$ | $108_{(12)}$ |
| | $N = 800$ | $63_{(6)}$ | $63_{(6)}$ | $72_{(6)}$ | $101_{(27)}$ | $270_{(15)}$ | $76_{(8)}$ |

The constraint dimension is held constant at $m = 80$. The values of RMSE reported are actual values multiplied by 1000.

Table 4: The frequency of correct estimation for the number of factors by both ER and CER with 200 replications

| | | $m = 8$ | | | | $m = 80$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\delta = 0$ | | $\delta = 0.5$ | | $\delta = 0$ | | $\delta = 0.5$ | |
| | | CER | ER | CER | ER | CER | ER | CER | ER |
| $T = 100$ | $N = 100$ | 1 | 1 | 0.98 | 0.89 | 1 | 1 | 0.99 | 0.99 |
| | $N = 200$ | 0.99 | 0.97 | 0.96 | 0.81 | 1 | 1 | 1 | 0.99 |
| | $N = 400$ | 1 | 1 | 0.99 | 0.79 | 0.99 | 0.99 | 1 | 1 |
| | $N = 800$ | 0.99 | 1 | 0.99 | 0.72 | 1 | 1 | 1 | 1 |
| $T = 200$ | $N = 100$ | 0.99 | 1 | 1 | 0.94 | 1 | 1 | 1 | 1 |
| | $N = 200$ | 1 | 1 | 0.99 | 0.96 | 1 | 1 | 1 | 1 |
| | $N = 400$ | 1 | 1 | 1 | 0.91 | 1 | 1 | 1 | 1 |
| | $N = 800$ | 0.99 | 1 | 0.99 | 0.89 | 1 | 1 | 1 | 1 |
| $T = 400$ | $N = 100$ | 1 | 1 | 1 | 0.96 | 1 | 1 | 1 | 1 |
| | $N = 200$ | 0.97 | 1 | 1 | 0.96 | 1 | 1 | 1 | 1 |
| | $N = 400$ | 1 | 1 | 1 | 0.99 | 1 | 1 | 1 | 1 |
| | $N = 800$ | 0.99 | 1 | 1 | 0.97 | 1 | 1 | 1 | 1 |
| $T = 800$ | $N = 100$ | 1 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 |
| | $N = 200$ | 0.99 | 1 | 1 | 0.95 | 1 | 1 | 1 | 1 |
| | $N = 400$ | 0.99 | 1 | 1 | 0.98 | 1 | 1 | 1 | 1 |
| | $N = 800$ | 0.98 | 1 | 1 | 0.93 | 1 | 1 | 1 | 1 |

Table 5: Three estimates of loading matrices in the constrained and unconstrained factor models for monthly excess stock return series.

| Loading | Row | Estimates | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $A'H'$ | 1 | -0.49 | -0.49 | -0.49 | -0.49 | 0.07 | 0.07 | -0.09 | -0.09 | -0.09 | -0.09 |
| | 2 | 0.09 | 0.09 | 0.09 | 0.09 | 0.01 | 0.01 | -0.5 | -0.5 | -0.5 | -0.5 |
| | 3 | -0.05 | -0.05 | -0.05 | -0.05 | -0.7 | -0.7 | -0.01 | -0.01 | -0.01 | -0.01 |
| $L'$ | 1 | -0.29 | -0.4 | -0.53 | -0.54 | -0.12 | 0.03 | -0.11 | -0.28 | -0.27 | 0.02 |
| | 2 | -0.07 | 0.73 | -0.11 | 0 | 0.2 | -0.07 | -0.22 | -0.54 | -0.23 | -0.05 |
| | 3 | -0.12 | 0.04 | 0.11 | 0.01 | -0.61 | -0.75 | -0.03 | -0.1 | 0.11 | -0.14 |

$A'H'$ and $L'$ denote the constrained and unconstrained model, respectively. The loading matrices are normalized for ease in comparison.