Long Memory Factor Model: On Estimation of Factor Memories*

CHEUNG Ying Lun†

Goethe University Frankfurt

January 16, 2018

Abstract

This paper considers the estimation of factor memories in the context of a high-dimensional factor model. Both factors and idiosyncratic error terms are potentially nonstationary fractional integrated processes. We propose a three-step procedure to estimate the latent factors. We then apply the fully-extended local Whittle (FELW) estimator of Abadir et al. (2007) to compute factor memories. This estimator is consistent and satisfies the same normal CLT, as if the factors are observed. Finite sample performance of the proposed procedure is evaluated in a simulation study. Finally, we apply the proposed estimator on a large dataset of macroeconomic variables.

Keywords: Approximate factor models, principal components, long memory, fractional integration
JEL: C14, C22, C38

* I would like to thank Uwe Hassler and Michael Binder for their support and guidance. I am also grateful to Yangming Bao as well as the participants of the RMSE Meeting 2017 (Rotterdam) and the 3rd Vienna Workshop on High-Dimensional Time Series in Macroeconomics and Finance (Vienna) for useful comments and suggestions. I gratefully acknowledge the financial support from the Department of Economics and Business Administration at the Goethe University. All remaining errors are my own.

† **Email**: `yinglun.cheung@econ.uni-frankfurt.de`

1. Introduction

Long memory processes, often modeled as fractionally integrated $I(d)$ processes, are important generalization to the knife-edge distinction between $I(0)$ and $I(1)$ models. It allows a time series to be persistent and even nonstationary (as opposed to an $I(0)$ process), but at the same time have temporary shocks (as opposed to an $I(1)$ process). It also nests anti-persistent (with $d < 0$) and explosive processes (with $d > 1$) naturally in one model. With this generalization, researchers can draw much richer conclusions on the long run behavior of a time series. In fact, there are vast empirical evidence of long memories in financial and economic variables.[1] The questions then arise: do the persistences of different variables have common sources? If yes, can we estimate their integration orders? To answer these questions, this paper considers a high-dimensional factor model that explicitly incorporates fraction integrations in the latent factors.

Thanks to the availability of large datasets and the increasing computation power, high-dimensional factor models have attracted much attention of economists in the last few decades. To estimate the latent factors, one usually transforms each series to stationary and then applies principal component analysis (PCA), see e.g. Stock and Watson (2002a,b), Bai and Ng (2002) and Bai (2003), among others. To accommodate nonstationarity, Bai (2004) considers factors that are $I(1)$, while Bai and Ng (2004, 2010) also allow for nonstationary idiosyncratic error terms. In particular, the latter show that the common factors can be consistently estimated (up to an invertible rotation matrix $\boldsymbol{H}$ and a level shift) by PCA upon first differencing the observed variables, which we denote as the PANIC estimator hereafter. However, fractional integration is still not considered.

[1]Hassler and Wolters (1995) and Baillie et al. (1996) provide international evidence for the persistence behavior of inflation rates; Caporale and Gil-Alaña (2007) model US unemployment rate by a variety of fractional integrated processes; Gil-Alaña and Robinson (2001) analyze UK and Japanese consumption and income series and find that a model with seasonal fractional integration can provide a good fit; while Ding et al. (1993), Lobato and Savin (1998), Andersen et al. (2001), Andersen et al. (2003), Sibbertsen (2004) and Gil-Alaña et al. (2014) find long memories in stock volatilities.

This paper sits at the confluence of the two strands of research: it explicitly incorporates the notion of fractional integration in a high-dimensional factor model. This paper develops an estimation procedure that can consistently estimate the factor memories. This task has also been considered in Luciani and Veredas (2015) and Ergemen (2016). Both papers try to compute factor memories using the PANIC factors. The former considers a large panel of realized volatilities and apply the quasi-maximum likelihood estimation of Beran (1995) on the estimated factors to compute the factor memories. However, they only consider a one-factor model and do not provide the asymptotic properties of the estimator. The latter paper makes use of the conditional-sum-of-squares (CSS) estimator and proves consistency of the memory estimate. However, that paper assumes the factor memories to be the same across all factors, and larger than that of the idiosyncratic errors. In essence, all observed data series have the same integration order.

In this paper, we consider a similar model under a more general setting. First, we allow the factor memories to differ from one another. To achieve that, we adopt the identification restrictions of Bai and Ng (2013), and show that the presence of more than one factor does not affect the estimation of factor memories as long as they are not too far from each other. We also allow the integration orders of idiosyncratic errors to be bigger than factor memories. For this purpose, we develop a three-step procedure for factor estimations:

1. Obtain the estimated factor loadings $\widehat{\boldsymbol{\lambda}}_i$ by the PANIC approach.
2. Transform the estimated loadings with the identification restrictions proposed in Bai and Ng (2013).
3. Regress the observed series X_{it} on $\widehat{\boldsymbol{\lambda}}_i$ for each t to obtain the factor estimates.

Under mild assumptions, the factor space can be consistently estimated, and the rotation matrix $\boldsymbol{H}$ converges to the identity matrix. Second, we do not make any parametric assumptions on the short memory component of the factors. We apply the fully-extended local Whittle (FELW) estimator of Abadir et al. (2007) on the estimated factors, and show that the consistency and asymptotic normality of FELW continue to hold. This estimator has the same asymptotic variance as if the true factors are observed.

This paper is organized as follows. In the next section, we demonstrate the importance of our method with a simple motivational example. We then introduce the model assumptions and describe the estimation procedures of latent factors in Section 3. Section 4 is devoted to the asymptotic properties of the estimated factor memories. The finite sample performance of the proposed method is examined in a simulation study in Section 5. We apply the proposed method to a large set of macroeconomic variables in Section 6. Finally, Section 7 concludes. Mathematical proofs and technical lemmas are relegated to the Appendix.

Notation. For an $m \times n$ matrix $\boldsymbol{A}$, we denote its Frobenius norm as $\|\boldsymbol{A}\| := \sqrt{\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}')}$, where := means 'is defined as'. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ denotes convergence in distribution. The big O and little o notations $O(\cdot)$ and $o(\cdot)$ are defined as usual, and so are their probabilistic counterparts $O_p(\cdot)$ and $o_p(\cdot)$. All vector- and matrix-valued variables are written in bold font. Whenever possible, we will use $s, t \in \{0, \dots, T\}$ for time subscript, $i \in \{1, \dots, N\}$ for cross-sectional units, and $k, l \in \{1, \dots, K\}$ to indicate a specific factor.

2. Motivation: A Simple Example

Suppose one is trying to examine the persistence of an economic or financial variable, say the market return, by estimating its memory parameter. However, since the actual series is not observable, one has to rely on an imperfect proxy (e.g., using S&P500 as a market proxy). For simplicity, assume the proxy follows a signal-plus-noise model

$$Z_t = F_t + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \left(0, \sigma^2\right), \tag{2.1}$$

where $F_t \sim I(d_F)$ is unobservable, and ε_t is an independent noise term. The aim is to estimate the integration order of F_t. By Theorem 2.5 of Abadir et al. (2007), FELW of Z_t converges in probability to the larger between d_F and 0, i.e., $\widehat{d}_Z \xrightarrow{p} \max\{d_F, 0\}$. It means that the estimator is not consistent when $d_F < 0$. In general, if $\varepsilon_t \sim I(d_e)$, then $\widehat{d}_Z \xrightarrow{p} \max\{d_F, d_e\}$. With finite samples, the presence of ε_t can affect the estimation of d_F even when $d_F > d_e$. However, if we are given a set of variables (e.g., stock or portfolio returns) that possess a factor structure, we can consistently estimate d_F by using the estimated factor $\widehat{F}_t$ instead.

To illustrate, suppose that a set of N variables are known to have the factor structure

$$X_{it} = \lambda_i F_t + e_{it}.$$

Since e_{it} can be nonstationary in practice, the PCA estimator may be inconsistent. Thus, we follow Bai and Ng (2004) and take the first difference of X_{it} before computing $\widehat{\lambda}_i$ by PCA. Then for each t we regress X_{it} on $\widehat{\lambda}_i$ to obtain $\widehat{F}_t$.[2] Finally, we apply FELW and compute $\widehat{d}_{\widehat{F}}$. The advantage of this approach over directly using the PANIC estimator will be made clear when we introduce the assumptions and asymptotic properties of our estimator.

The sample size is set as $T = N = 100$. We plot the estimated factor memory averaged over 1000 repetitions with dotted lines in Figure A.1, together with the true value drawn with the red solid lines. The gray shaded areas represent the error bounds, computed as two standard deviations above and below the average value. The left panel reports

[2]We skip the second step of our three-step procedure here, since there is no rotation problem in a single-factor model.

the results using Z_t alone, and the plot on the right makes use of the estimated factors. When $d_F < 0$, the estimated memory computed with the proxy is positively biased and stay around 0. For positive d_F, a small negative bias can be observed. This bias vanishes when factor memories are computed with the estimated factor. This result is of practical importance. It shows that whenever data is available, one should try to estimate the common factor instead of relying on a proxy. The theoretical ground will be provided in this paper.

3. Estimation of Latent Factors

In this section, we present the model assumptions and introduce the three-pass factor estimator. The data generating process is assumed to be

$$X_{it} = \boldsymbol{\lambda}_i' \boldsymbol{F}_t + e_{it}, \qquad i = 1, \ldots, N; \; t = 0, \ldots, T. \tag{3.1}$$

Eq.(3.1) describes a static factor model. Dynamic factor models with a finite number of lags can be easily transformed to this form by including the lags of factors in $\boldsymbol{F}_t$. X_{it} is the observed variable. $\boldsymbol{\lambda}_i$ and $\boldsymbol{F}_t$ are $K \times 1$ vectors of loadings and latent factors respectively. e_{it} is the idiosyncratic error. We can also stack all cross-sectional units together and write Eq.(3.1) as

$$\boldsymbol{X}_t = \boldsymbol{\Lambda} \boldsymbol{F}_t + \boldsymbol{e}_t, \qquad t = 0, \ldots, T$$

where $\boldsymbol{X}_t = (X_{1t}, \ldots, X_{Nt})'$, $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)'$ and $\boldsymbol{e}_t = (e_{1t}, \ldots, e_{Nt})'$.

To model the persistences of the common factors, we assume for each $k = 1, \ldots, K$

$$\nabla^{p_k} F_{kt} = \xi_{kt}, \qquad \xi_{kt} \sim I(d_{\xi_k}) \tag{3.2}$$

where $p_k \in \{0,1\}$ and $\nabla = (1-L)$ is the backward difference operator. Essentially, $\{F_{kt}\}$ is an $I(d_{F_k})$ process, where $d_{F_k} = p_k + d_{\xi_k}$, under the following definition taken from Abadir et al. (2007).

Definition 3.1 ($\boldsymbol{I(d)}$ process)

For $d = p + d_\xi$ where $p \in \mathbb{Z}$ and $d_\xi \in (-0.5, 0.5)$, we say that $\{F_t\}$ is an $I(d)$ process (i.e., $F_t \sim I(d)$) if

$$\nabla^p F_t = \xi_t, \qquad t = 1-p, 2-p, \ldots,$$

where ξ_t is a second-order stationary sequence with spectral density

$$\mathfrak{f}_\xi(\omega) = b_0|\omega|^{-2d_\xi} + o(|\omega|^{-2d_\xi}) \qquad \text{as } \omega \to 0,$$

where $b_0 > 0$. We say d is the memory parameter or the integration order of $\{F_t\}$.

Unlike in Ergemen (2016), p_k and d_{ξ_k} need not be the same across factors. As argued in Bai (2004), different orders of integration among factors can accommodate (fractional) cointegration. However, instead of viewing the factors as a system, we will mainly focus on the case when they can be identified and treated separately. Therefore, cointegration relations among factors are not considered in this paper, and are left for future research.

3.1. Step 1: Estimate Loadings by PANIC

Since we allow the error terms to be nonstationary, the PCA estimator can be inconsistent. Hence, we follow Bai and Ng (2004) and estimate the factor loadings by principal component analysis using differenced data. Let $x_{it} = \nabla X_{it}$ and $\boldsymbol{f}_t = \nabla \boldsymbol{F}_t$, the PANIC estimator $\widetilde{\boldsymbol{f}}_t$ is $\sqrt{T}$ times the K eigenvectors corresponding to the K largest eigenvalues of the $T \times T$ matrix $\boldsymbol{x}\boldsymbol{x}'$. Under normalization $\widetilde{\boldsymbol{f}}'\widetilde{\boldsymbol{f}}/T = \boldsymbol{I}$, the factor loadings are estimated as $\widetilde{\boldsymbol{\Lambda}} = \boldsymbol{x}'\widetilde{\boldsymbol{f}}/T$. Let M be a generic constant independent of N and T, we make the following assumptions.

Assumption 3.1 (Factor loadings)

1. *For nonrandom* $\boldsymbol{\lambda}_i$, $\|\boldsymbol{\lambda}_i\| \leq M$; *for random* $\boldsymbol{\lambda}_i$, $\mathbb{E}\|\boldsymbol{\lambda}_i\|^4 \leq M$.

2. $N^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i' \xrightarrow{p} \boldsymbol{\Sigma}_\Lambda$ *as* $N \to \infty$ *for some* $K \times K$ *positive definite matrix* $\boldsymbol{\Sigma}_\Lambda$.

Assumption 3.2 (Common factors)

Let $\boldsymbol{F}_t = (F_{1t}, \dots, F_{Kt})'$ *and* $\boldsymbol{\xi}_t = (\xi_{1t}, \dots, \xi_{KT})'$, *then for all* $k = 1, \dots, K$

1. $F_{kt} \sim I(d_{F_k})$ *with* $d_{F_k} = p_k + d_{\xi_k}$ *and* $p_k \in \{0, 1\}$.

2. $\{\xi_{kt}\}$ *is a square summable linear process*

$$\xi_{kt} = \sum_{j=0}^{\infty} a_{kj}\varepsilon_{k,t-j}, \qquad \sum_{j=0}^{\infty} a_{kj}^2 < M,\ \varepsilon_{kt} \overset{iid}{\sim} (0,1)\,.$$

3. $T^{-1}\sum_{t=1}^{T}\boldsymbol{f}_t\boldsymbol{f}_t' \xrightarrow{p} \boldsymbol{\Sigma}_f$ *as* $T \to \infty$.

4. $\mathbb{E}\|\boldsymbol{f}_t\|^4 \leq M$.

Assumption 3.3 (Error terms)

1. $\mathbb{E}[e_{it}] = 0$, *and* $\mathbb{E}|\nabla e_{it}|^8 < M$.

2. *Let* $\gamma(s,t) = \mathbb{E}[\nabla\boldsymbol{e}_s'\nabla\boldsymbol{e}_t/N] = \mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\nabla e_{is}\nabla e_{it}\right]$, $|\gamma(s,s)| \leq M$ *for all* s *and* $\sum_{s=1}^{T}|\gamma(s,t)| \leq M$ *for each* t.

3. $\mathbb{E}[\nabla e_{it}\nabla e_{jt}] = \tau_{ij,t}$ *with* $|\tau_{ij,t}| \leq |\tau_{ij}|$ *and* $\sum_{i=1}^{N}|\tau_{ij}| < M$ *for all* j.

4. $\mathbb{E}[\nabla e_{it}\nabla e_{js}] = \tau_{ij,ts}$ *and* $(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{ij,ts}| \leq M$.

5. $\mathbb{E}\left|N^{-1/2}\sum_{i=1}^{N}(\nabla e_{it}\nabla e_{is} - \mathbb{E}[\nabla e_{it}\nabla e_{is}])\right|^4 < M$ *for all* (t,s).

Assumption 3.4 (Independence)

$\boldsymbol{F}_t, e_{it}$ *and* $\boldsymbol{\lambda}_i$ *are three mutually independent groups.*

The above assumptions follow Bai and Ng (2004) closely. The first two points in Assumption 3.2 are adopted from Abadir et al. (2007). They state that F_{kt} is a linear

$I(d_{F_k})$ process. Since $f_{kt} = \nabla F_{kt} \sim I(d_{f_k})$ where $d_{f_k} = d_{F_k} - 1 < 0.5$, f_{kt} is stationary, and the other points in Assumption 3.2 can be easily satisfied.

Though not explicitly stated in Assumption 3.3, ∇e_{it} can actually have long memory as well. To see that, let $\gamma_i(s,t) = \mathbb{E}\left[\nabla e_{is} \nabla e_{it}\right]$ be the autocovariance function of ∇e_{it}, and $e_{it} \sim I(d_{e_i})$ with $1 \leq d_{e_i} < 1.5$ and ∇e_{it} satisfying the assumptions in Hassler and Kokoszka (2010). Then, $\sum_{s=1}^{T} |\gamma_i(s,t)| = O(T^{d_{e_i}-1})$ because the coefficients of the Wold representation of ∇e_{it} decay at rate $n^{d_{e_i}-2}$ at lag n. As a result,

$$\sum_{s=1}^{T} \left| \frac{1}{N} \sum_{i=1}^{N} \gamma_i(s,t) \right| \leq \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{T} |\gamma_i(s,t)| = O\left(\frac{\sum_{i=1}^{N} T^{d_{e_i}-1}}{N} \right).$$

For Assumption 3.3 to be satisfied, we need the above expression to be $O(1)$. If for example $T = N$, we can allow $O\left(\sqrt{N}\right)$ of ∇e_{it} to have long memory. The following lemma can be taken from Bai (2003) and Bai and Ng (2004).

Lemma 3.1 (Estimation of Factor Loadings)

Let $\widetilde{\boldsymbol{\lambda}}_i$ be the PANIC estimator of the factor loadings. As $N, T \to \infty$, there exists a $K \times K$ invertible matrix $\boldsymbol{H}$ such that for each i

$$\min\left\{\sqrt{T}, N\right\} \left(\widetilde{\boldsymbol{\lambda}}_i - \boldsymbol{H}^{-1}\boldsymbol{\lambda}_i\right) = O_p(1).$$

3.2. Step 2: Impose Identification Restrictions

As widely acknowledged in the literature of factor analysis, the PCA estimator is only consistent up to a rotation. Since there are K^2 free elements in the rotation matrix $\boldsymbol{H}$, we have to impose K^2 identification restrictions. To this end, we follow Bai and Ng (2013) and consider the following three sets of restrictions and transformations:

(PC1) $T^{-1}\boldsymbol{f}'\boldsymbol{f} = \boldsymbol{I}_K$ and $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ is a diagonal matrix with distinct entries. In this case, no transformation is needed and $\widehat{\boldsymbol{\Lambda}} = \widetilde{\boldsymbol{\Lambda}}$.

(PC2) $T^{-1}\boldsymbol{f}'\boldsymbol{f} = \boldsymbol{I}_K$ and $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1', \boldsymbol{\Lambda}_2')'$ where

$$\boldsymbol{\Lambda}_1 = \begin{pmatrix} \lambda_{11} & 0 & \dots & 0 \\ \lambda_{21} & \lambda_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{K1} & \lambda_{K2} & \dots & \lambda_{KK} \end{pmatrix}$$

with $\lambda_{ii} \neq 0$ for $i = 1, \dots, K$, and $\boldsymbol{\Lambda}_2$ is unrestricted. Let $\widetilde{\boldsymbol{\Lambda}}$ be the unrestricted principal component estimator, we compute the QR decomposition of $\widetilde{\boldsymbol{\Lambda}}_1' = \boldsymbol{QR}$ such that $\boldsymbol{R}$ is an upper triangular matrix with positive diagonal elements and $\boldsymbol{Q}$ is a $K \times K$ orthonormal matrix. We define $\widehat{\boldsymbol{\Lambda}} = \widetilde{\boldsymbol{\Lambda}}\boldsymbol{Q} = (\boldsymbol{R}, \widetilde{\boldsymbol{\Lambda}}_2')'$ and the transformation matrix becomes $\boldsymbol{H}^* = \boldsymbol{Q}'\boldsymbol{H}$.

(PC3) $\boldsymbol{\Lambda} = (\boldsymbol{I}_K, \boldsymbol{\Lambda}_2')'$, while $\boldsymbol{f}_t$ and $\boldsymbol{\Lambda}_2$ are not restricted. Define $\widehat{\boldsymbol{\Lambda}} = \widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Lambda}}_1^{-1}$and the transformation matrix becomes $\boldsymbol{H}^* = \widetilde{\boldsymbol{\Lambda}}_1\boldsymbol{H}$.

Note that the standard PCA implicitly restricts $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ to be a diagonal matrix. **PC1** additionally assumes that it has distinct entries for identification. **PC2** says the elements in upper triangular part of the factor loading matrix $\boldsymbol{\Lambda}$ are zero. It means that the first variable is only affected by the first factor; the second variable is only affected by the first two factors; and so on. Finally, **PC3** implies that the first K variables follow the signal plus noise model $X_{kt} = F_{kt} + e_{kt}$, $k = 1, \dots, K$.

Upon first differencing, the assumptions in Bai and Ng (2013) are satisfied. Hence, we can take the following lemma from their paper.

Lemma 3.2 (Factor identification)

*Under Assumptions 3.1-3.4, and as $N, T \to \infty$, $\boldsymbol{H} = \boldsymbol{I} + O_p(1/\min\{N, T\})$ if **PC1** is satisfied; and $\boldsymbol{H}^* = \boldsymbol{I} + O_p(T^{-\frac{1}{2}})$ if **PC2** or **PC3** is satisfied.*

Combining Lemma 3.1 and 3.2, we can now write $\widehat{\boldsymbol{\lambda}}_i$ as

$$\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i = \left(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}^{*-1}\boldsymbol{\lambda}_i\right) + \left(\boldsymbol{H}^{*-1} - \boldsymbol{I}\right)\boldsymbol{\lambda}_i = O_p(1/\min\{\sqrt{T}, N\}).$$

Throughout this paper, we will assume $T = O(N^2)$ since it is the dominant case. Hence, we have $\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i = O_p\left(T^{-1/2}\right)$.

3.3. Step 3: Estimate Factors by OLS

Since Eq.(3.1) resembles a linear regression model for each t given $\boldsymbol{\lambda}_i$, we regress X_{it} on $\widehat{\boldsymbol{\lambda}}_i$ to estimate common factors in the third step. We compute

$$\widehat{\boldsymbol{F}}_t = \left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\lambda}}_i\widehat{\boldsymbol{\lambda}}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\lambda}}_i X_{it}\right). \tag{3.3}$$

We consider the following assumptions on the idiosyncratic term.

Assumption 3.5 (Weak dependence 1)

For each t,

1. $N^{-1/2}\sum_{i=1}^{N}\boldsymbol{\lambda}_i e_{it} = O_p(1)$.
2. $N^{-1}\sum_{s=1}^{T}|\mathbb{E}\left[\nabla\boldsymbol{e}_s'\boldsymbol{e}_t\right]| = O(1)$.
3. $(NT)^{-1/2}\sum_{s=1}^{T}\left(\boldsymbol{f}_s\nabla\boldsymbol{e}_s'\boldsymbol{e}_t - \mathbb{E}\left[\boldsymbol{f}_s\nabla\boldsymbol{e}_s'\boldsymbol{e}_t\right]\right) = O_p(1)$.

The above assumptions are not hard to satisfy if $\boldsymbol{e}_t$ are weakly stationary. The first point is the same as Assumption F.3 in Bai (2003). For the second point, notice that

$$N^{-1}\sum_{s=1}^{T}|\mathbb{E}\left[\nabla\boldsymbol{e}_s'\boldsymbol{e}_t\right]| \leq 2N^{-1}\sum_{s=0}^{T}|\mathbb{E}\left[\boldsymbol{e}_s'\boldsymbol{e}_t\right]|.$$

Therefore, it is implied by Assumption E in Bai (2003). Similarly, the third part is analogous to Assumption A.1(vii) in Su and Wang (2017).

Lemma 3.3 (Estimation of latent factors)

Let $\widehat{\boldsymbol{\lambda}}_i$ be the transformed estimated loadings, and Assumptions 3.1-3.4 be satisfied. Define $\widehat{\boldsymbol{F}_t}$ as in Eq.(3.3)*, then*

$$\widehat{\boldsymbol{F}_t} = \left(\boldsymbol{I} + O_p\left(T^{-1/2}\right)\right)\boldsymbol{F}_t + \left(\boldsymbol{\Sigma}_\Lambda + o_p\left(1\right)\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\lambda}}_i e_{it}. \tag{3.4}$$

If in addition Assumption 3.5 is satisfied,

$$\widehat{\boldsymbol{F}_t} = \left(\boldsymbol{I} + O_p\left(T^{-1/2}\right)\right)\boldsymbol{F}_t + O_p\left(C_{NT}^{-1}\right)$$

where $C_{NT} = \min\{\sqrt{N}, T\}$.

Proof. See Appendix.

The above lemma combines and generalizes the results of Bai and Ng (2004) and Bai and Ng (2013). The advantage of the three-pass estimator over the PANIC estimator is two-fold. First, unlike the PANIC estimator, there is no mean shift in our estimator. This allows us to perform a range of nonlinear transformation of the estimated factor without affecting consistency. For example, the absolute value of our estimated factor is again a consistent estimator of the absolute value of the true factor. Thus, subsequent estimation of memory of the transformed factor is still valid. Second, the expression Eq.(3.4) allows us to express the estimation errors in terms of the linear combination of $\boldsymbol{F}_t$ and $\boldsymbol{e}_t$ only. Their time-series properties, and thus their effect on the subsequent estimation of factor memories, can be analyzed easily.

Remark 3.1 Since $(\boldsymbol{\Sigma}_\Lambda + o_p(1))^{-1} = O_p(1)$, the estimation error is just a weighted sum of idiosyncratic error terms. Suppose $e_{it} \sim I(d_{e_i})$ for each i and $d_{F_k} \geq \max_i d_{e_i}$ for all k, as in Ergemen (2016), then consistency of FELW of the estimated factors follows directly. However, this assumption would be too restrictive, especially when $d_{F_k} < 0$.

Remark 3.2 It is important to point out that Lemma 3.3 does not necessarily imply that $\widehat{F}_{kt} \xrightarrow{p} F_{kt}$ for all k and t, although the rotation matrix converges to the identity matrix. Suppose $\sqrt{T} < N$ and $F_{lt} \sim I(1)$ for some $l \neq k$. Then $F_{lT} = O_p(\sqrt{T})$ and the error term is only $O_p(1)$. Nonetheless, Lemma 3.3 is sufficient for our purpose.

4. Estimation of Factor Memory

Given the estimated factors, we can find the respective factor memories. Since the short memory components of the factors are not specified, we propose to use the fully-extended local Whittle (FELW) estimator of Abadir et al. (2007). For this paper to be self-contained, we first summarize the assumptions and asymptotic behaviors of FELW in Section 4.1. Additional assumptions and asymptotic results of our estimator will be given in subsequent sections.

4.1. Fully-Extended Local Whittle (FELW) Estimator

FELW is an extension of the local Whittle (LW) estimator of Robinson (1995). Phillips and Shimotsu (2004) find that LW is only consistent for $d \in (-0.5, 1]$ and asymptotically normally distributed for $d \in (-0.5, 0.75)$. To extend the admissible range of the memory parameter, FELW adopts the idea of the extended discrete Fourier transform (DFT) of Phillips (1999). Given a stochastic process $\{F_{kt}\}$, $t = 1, \ldots, T$, and let $\omega_j = 2\pi j/T$ be the j-th Fourier frequency, the extended DFT is defined as

$$w_{F_k}(j; d) = w_{F_k}(j) + k(j; d).$$

Here $w_{F_k}(j)$ is the standard DFT

$$w_{F_k}(j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} F_{kt} \mathrm{e}^{\mathrm{i}t\omega_j},$$

and $k(j;d)$ is the correction term

$$k(j;d) = \begin{cases} -\mathrm{e}^{-\mathrm{i}\omega_j} Z_0 & d \in \mathcal{I}_{-1} = [-1.5, -0.5) \\ 0 & d \in \mathcal{I}_0 = [-0.5, 0.5) \\ \mathrm{e}^{\mathrm{i}\omega_j} \sum_{r=1}^{p} Z_r & d \in \mathcal{I}_p,\ p = 1, 2 \ldots \end{cases}$$

where $Z_0 = (2\pi T)^{-1/2} \sum_{t=1}^{T} F_{kt}$ and

$$Z_r = \frac{1}{\sqrt{2\pi T}} (\nabla^{r-1} F_{kT} - \nabla^{r-1} F_{k0}), \qquad r = 1, \ldots, p.$$

FELW is defined as

$$\widehat{d}_{F_k} = \arg\min_{d \in \mathcal{I}} U_{F_k}(d) \tag{4.1}$$

where $\mathcal{I} = [\overline{d}, \underline{d}] \subset [-1.5, \infty)$ and

$$U_{F_k}(d) = \log \left(\frac{1}{m} \sum_{j=1}^{m} \omega_j^{2d} I_{F_k}(j;d) \right) - \frac{2d}{m} \sum_{j=1}^{m} \log \omega_j.$$

Here $I_{F_k}(j;d)$ is the extended periodogram given by

$$I_{F_k}(j;d) = |w_{F_k}(j;d)|^2.$$

Assumption 4.1 (Bandwidth 1)

The bandwidth is chosen such that $m \to \infty$ and $m = o(T)$.

This assumption implies that we only consider frequencies close to zero. Therefore, the estimator is robust to the presence of seasonality and other effects on the series $\{F_{kt}\}$ at nonzero spectral frequencies, as noted in Abadir et al. (2011). Abadir et al. (2007) prove the following result.

Lemma 4.1 (Consistency of FELW)

Suppose $\boldsymbol{F}_t = (F_{1t}, \ldots, F_{Kt})$ is generated by Eq.(3.2) and satisfies Assumption 3.2 for *each k. Moreover, the bandwidth m satisfies Assumption 4.1. Let $\widehat{d}_{F_k}$ denote the FELW estimator computed with $\{F_{kt}\}$, then $\widehat{d}_{F_k} \xrightarrow{p} d_{F_k}$ for each k when $T \to \infty$.*

For asymptotic normality, we need to strengthen the assumptions.

Assumption 4.2 (Spectrum 1)

Let $\alpha_k(\omega) = \sum_{j=1}^{\infty} a_{kj} \mathrm{e}^{\mathrm{i}j\omega}$, where $\{a_{kj}\}$ is defined in Assumption 3.2. Then for each k

$$\frac{\mathrm{d}}{\mathrm{d}\omega}\alpha_k(\omega) = O\left(\frac{|\alpha_k(\omega)|}{\omega}\right), \qquad \text{as } \omega \to 0^+.$$

Assumption 4.3 (Spectrum 2)

For some $\beta_k \in (0, 2]$, the spectral density of ξ_{kt} satisfies

$$\mathfrak{f}_{\xi_k}(\omega) = |\omega|^{-2d_{\xi_k}}(b_{0k} + b_{1k}|\omega|^{\beta_k} + o(|\omega|^{\beta_k})).$$

Assumption 4.4 (Bandwidth 2)

The bandwidth choice satisfies $m = o\left(T^{2\beta_k/(1+2\beta_k)}\right)$.

Lemma 4.2 (Asymptotic normality of FELW)

Under assumptions in Lemma 4.1, together with Assumptions 4.2-4.4,

$$2\sqrt{m}(\widehat{d}_{F_k} - d_{F_k}) \xrightarrow{d} \mathcal{N}(0, 1).$$

The two lemmas in this section state that d_{F_k} can be consistently estimated by FELW, if $\{F_{kt}\}$ is observable. Moreover, the estimation error is asymptotically normally distributed. In what follows, we shall prove that when $\{F_{kt}\}$ is replaced by $\{\widehat{F}_{kt}\}$, theses results continue to hold under certain conditions.

4.2. Asymptotic Theory with Stationary Errors

In this section, we first consider the case when Assumption 3.5 is satisfied. Express the k-th element of $\widehat{\boldsymbol{F}_t}$ as

$$\widehat{F}_{kt} = F_{kt} + \underbrace{\sum_{l=1}^{K} h_{kl} F_{lt}}_{=:\varphi_{kt}} + \underbrace{(\boldsymbol{\Sigma}_\Lambda + o_p(1))_k^{-1} \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\lambda}}_i e_{it}}_{=:\zeta_{kt}} \tag{4.2}$$

where $(\boldsymbol{\Sigma}_\Lambda + o_p(1))_k^{-1}$ is the k-th row of $(\boldsymbol{\Sigma}_\Lambda + o_p(1))^{-1}$. Lemma 3.3 together with Assumption 3.5 imply that $h_{kl} = O_p(1/\min\{\sqrt{T}, N\})$ and $\zeta_{kt} = O_p(1/\min\{\sqrt{N}, T\})$. Since the estimated factor is just a linear combination of the true factors and error terms, it is straightforward to show how the first-stage estimation error carries over to the second step. We first consider the asymptotic magnitudes of the DFTs of the latent factors and those of their estimation errors.

Lemma 4.3 (DFT of factors)

If $p_k = 0$, then DFTs of the k-th factor and its first difference are

$$\begin{aligned} w_{F_k}(j) &= O_p((j/T)^{-d_{F_k}}), \\ w_{\nabla F_k}(j) &= O_p((j/T)^{-(d_{F_k}-1)}) + O_p(T^{-1/2}). \end{aligned}$$

If $p_k = 1$, then

$$\begin{aligned} w_{F_k}(j) &= O_p((j/T)^{-d_{F_k}}) + O_p(j^{-1}T^{d_{F_k}}), \\ w_{\nabla F_k}(j) &= O_p((j/T)^{-(d_{F_k}-1)}). \end{aligned}$$

Proof. See Appendix.

Lemma 4.4 (DFT of estimation errors 1)

Let ζ_{kt} be the estimation error defined in Eq.(4.2). Then, under Assumption 3.5

$$w_{\zeta_k}(j) = O_p(\sqrt{T}C_{NT}^{-1}), \qquad w_{\nabla\zeta_k}(j) = O_p(jT^{-1/2}C_{NT}^{-1}).$$

Proof. See Appendix.

According to Lemma 4.4, DFT of the error term $\boldsymbol{\zeta}_t$ does not necessarily converge to zero, even though $\boldsymbol{\zeta}_t$ itself does. When $N = O(T)$, $w_{\zeta_k}(j) = O_p(1)$. It can actually dominate the DFT of the true factor, since Lemma 4.3 implies $w_{F_k}(j) = o_p(1)$ when $d_{F_k} < 0$. The following assumption is needed to ensure N is large enough.

Assumption 4.5 (Bandwidth 3)

For $k = 1, \ldots, K$, if $d_{F_k} \geq 0$,

$$m^{2d_{F_k}} T^{1-2d_{F_k}} N^{-1} \to 0. \tag{4.3}$$

If $d_{F_k} < 0$,

$$T^{1-2d_{F_k}} N^{-1} \leq M < \infty. \tag{4.4}$$

Although Eq.(4.3) and (4.4) depend on the unknown parameter d_{F_k}, sufficient conditions independent of d_{F_k} can be easily formed. Specifically, if $T = O(\sqrt{N})$, then Assumption 4.5 is satisfied uniformly over $d_{F_k} > -0.5$.

Another source of errors comes from the presence of more than one factor. As DFT is a linear operator, we have

$$w_{\varphi_k}(j) = \sum_{l=1}^{K} h_{kl} w_{F_l}(j), \qquad h_{kl} = O_p\left(T^{-1/2}\right).$$

We need the following assumption.

Assumption 4.6 (Factor memory)

Let $\overline{d} = \max_l d_{F_l}$, we assume $\overline{d} - d_{F_k} \leq 0.5$.

Note that this assumption is imposed on each factor independently. To illustrate, consider a three-factor model in which the factor memories are $(d_{F_1}, d_{F_2}, d_{F_3}) = (0.2, 0.6, 0.8)$ respectively. Then although F_{1t} violates Assumption 4.6 since $\overline{d} - d_{F_1} = 0.6 > 0.5$, theorems derived in this section continue to hold for d_{F_2} and d_{F_3}.

Theorem 4.1 (Consistency of factor memory estimation)

Suppose Assumptions 3.1-3.5 are satisfied, and $N, T, m \to \infty$ as required in Assumptions 4.1 and 4.5. Let $\widehat{\boldsymbol{F}}_t$ be defined as in Eq.(3.3) *and $\widehat{d}_{\widehat{F}_k}$ be the FELW applied on the k-th estimated factor. Then,*

$$\widehat{d}_{\widehat{F}_k} \xrightarrow{p} d_{F_k}.$$

Moreover,

$$\widehat{d}_{\widehat{F}_k} - d_{F_k} = -\frac{1}{2} Q_{m,k}(1 + o_p(1)) + O_p(m^{-1} \log m), \tag{4.5}$$

where

$$Q_{m,k} = \frac{1}{m} \sum_{j=1}^{m} (\log(j/m) + 1) \eta'_{\nabla^{p_k} \widehat{F}_k, j} + o_p\left(\left(\frac{m}{T}\right)^2\right)$$

and

$$\eta'_{\nabla^{p_k} \widehat{F}_k, j} = \frac{I_{\nabla^{p_k} \widehat{F}_k}(j)}{b_{0k} \omega_j^{-2d_{\xi_k}}}.$$

Proof. See Appendix.

Theorem 4.1 shows that one can estimate the factor memories consistently, even though the factors are not observable. Recall in Section 2 we show that memory estimation is biased when one directly apply FELW to a noisy proxy. Theorem 4.1 provides the theoretical ground to solve this problem.

Now we turn to derive the asymptotic distribution of our estimator. For this, we have to strengthen the assumptions on the sample sizes and the bandwidth choice.

Assumption 4.7 (Bandwidth 4)

We assume

$$\log(m)\sqrt{m}\left(\frac{m}{T}\right)^{2d_{F_k}} \frac{T}{N} \to 0. \tag{4.6}$$

Again $T = O\left(\sqrt{N}\right)$ is sufficient for this assumption to be satisfied, since

$$\log(m)\sqrt{m}\left(\frac{m}{T}\right)^{2d_{F_k}}\frac{T}{N} = \frac{\log(m)}{\sqrt{m}}\left(\frac{m}{T}\right)^{2d_{F_k}+1}\frac{T^2}{N} \to 0.$$

Theorem 4.2 (Asymptotic normality of factor memory estimation)
Under conditions in Theorem 4.1, together with Assumptions 4.2-4.4 and 4.7,

$$\sqrt{m}(\widehat{d}_{\widehat{F}_k} - \widehat{d}_{F_k}) = o_p(1).$$

Therefore,

$$2\sqrt{m}(\widehat{d}_{\widehat{F}_k} - d_{F_k}) \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. See Appendix.

We show in the above theorem that the difference between FELW computed with the estimated factors and that computed with the true factors is small, and converges to zero at a rate faster than $\sqrt{m}$. It implies that our estimator does not lose efficiency asymptotically. Besides asymptotic normality, the following corollary can also be useful in practice.

Corollary 4.1 (Estimation with non-overlapping subsamples)
Suppose we can divide a panel into b equal parts for some finite b, so that the a-th subsample spans the period $\mathcal{T}_a = \{(a-1)T/b+1, \ldots, aT/b\}$, $a = 1, \ldots, b$. Let $\widehat{d}^{(a)}_{\widehat{F}_k}$ be estimated using data within $\mathcal{T}_a$. Moreover, let all assumptions in Theorem 4.2 be satisfied and choose m/b as the bandwidth. Then $\widehat{d}^{(a_1)}_{\widehat{F}_k}$ is asymptotically independent of $\widehat{d}^{(a_2)}_{\widehat{F}_k}$ for $a_1 \neq a_2$.

Proof. See Appendix.

This result is useful in the following two situations. First, in many settings, T can be much larger than N, especially for financial variables. For example, there can be over

10,000 observations in a dataset of daily stock returns beginning in the 1960s, while only thousands of stocks or hundreds of portfolios are available. However, Assumption 4.7 requires N to be large compared to T. Corollary 4.1 then allows us to split the dataset into shorter intervals for estimation. Suppose the dataset can be evenly divided into b parts, then

$$\sqrt{m}(\overline{d}_{\widehat{F}_k} - d_{F_k}) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4}\right),$$

where

$$\overline{d}_{\widehat{F}_k} = \frac{1}{b}\sum_{a=1}^{b} \widehat{d}^{(a)}_{\widehat{F}_k}.$$

Note that $\overline{d}_{\widehat{F}_k}$ has the same asymptotic variance as $\widehat{d}_{F_k}$, the infeasible estimator computed with full sample of the true variable. Second, we can apply the simple tests proposed by Hassler et al. (2014) to test for changes in factor memory. Under the null hypothesis of no structural changes, the following holds: $H_0 : d^{(a_1)}_{F_k} = d^{(a_2)}_{F_k} = d_{F_k}$

$$\sqrt{\frac{2m}{b}}\left(\widehat{d}^{(a_1)}_{\widehat{F}_k} - \widehat{d}^{(a_2)}_{\widehat{F}_k}\right) \xrightarrow{d} \mathcal{N}(0,1). \tag{4.7}$$

Inference can thus be made on the difference between estimated factors using different subsamples.

4.3. Asymptotic Theory with Nonstationary Errors

Recall that in the first step of the three-pass factor estimation procedure, we compute the factor loadings with differenced data. In fact, under our assumptions, Lemma 3.1 and 3.2 hold even when the idiosyncratic terms are nonstationary. Moreover, Eq.(3.4) is still valid. We can decompose the last term there as

$$\frac{1}{N}\widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1}\sum_{i=1}^{N}\widehat{\boldsymbol{\lambda}}_i e_{it} = \widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1}\Bigg(\underbrace{\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_i e_{it}}_{=:\boldsymbol{\nu}_{1t}} + \underbrace{\frac{1}{N}\sum_{i=1}^{N}(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i)e_{it}}_{=:\boldsymbol{\nu}_{2t}}\Bigg).$$

Since both $\widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1}$ and $\widehat{\boldsymbol{\lambda}}_i$ are bounded, if $e_{it} \sim I(d_{e_i})$ and $\max_i d_{e_i} \leq \min_k d_{F_k}$, consistency follows without extra assumptions. However, this assumption is quite restrictive, especially when the factors are allowed to be stationary. To analyze the case when the idiosyncratic terms are potentially more persistent than the factors, we introduce the following additional assumptions.

Assumption 4.8 (Weak dependence 2)

For each t, $\sqrt{N}\nabla\boldsymbol{\nu}_{1t} = N^{-1/2}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\nabla e_{it} = O_p(1)$.

Assumption 4.9 (Idiosyncratic error memory)

$\overline{d}_e - d_{F_k} \leq 0.5$, where $\overline{d}_e \equiv \max_i d_{e_i}$.

Assumption 4.10 (Sample size)

$T^{3-2d_{F_k}}N^{-1} \leq M$.

Assumption 4.8 controls the effect of $\boldsymbol{\nu}_{1t}$. It is analogous to the first part of Assumption 3.5 but is less stringent. It can be satisfied even if $\boldsymbol{e}_t$ is nonstationary. Assumption 4.9 allows the memory parameters of the idiosyncratic terms to be larger than those of the factors by at most 0.5. Finally Assumption 4.10 controls the relative rate of divergence in sample sizes. At first sight it is much more stringent than Eq.(4.4) as a very large N is needed. However, if $\overline{d}_e \geq 1$, then Assumption 4.9 implies $d_{F_k} \geq 0.5$ and the condition $T^2N^{-1} \leq M$ is again sufficient.

Theorem 4.3 (Consistency under nonstationary idiosyncratic errors)

Under Assumptions 4.8-4.10 and conditions in Theorem 4.1 except Assumption 3.5, the result in Theorem 4.1 holds.

Proof. See Appendix.

We show in the above theorem that consistency of our estimator still holds under mild assumptions on e_{it}. However, a large N is needed. One reason for this is that we do not explicitly take into account the time-series property of $\boldsymbol{\nu}_{1t}$ in the proof of Theorem 4.3. We can strengthen the first point of Assumption 4.8 slightly to avoid this.

Assumption 4.11 (Upper bound of DFT of estimation errors)

There exists an $I(d_g)$ process g_t with $d_g \leq \overline{d}_e$ such that $|\sqrt{N} w_{\nu_{1k}}(j)| = O_p(|w_g(j)|)$ uniformly over $1 \leq j \leq m$, where $w_{\nu_{1k}}(j)$ and $w_g(j)$ are respectively the DFT of the k-th element of $\boldsymbol{\nu}_{1t}$ and g_t evaluated at the j-th Fourier frequency.

This assumption is an alternative way to state the weak dependence among e_{it}. It states that the DFT of $\sqrt{N}\boldsymbol{\nu}_{1t}$ is bounded by that of an $I(d_g)$ process. Since $\boldsymbol{\lambda}_i = O_p(1)$ and $(\boldsymbol{\lambda}_i, e_{jt})$ are independent for all i, j, t,

$$\begin{aligned}
\operatorname{cov}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{ki}e_{it}, \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{li}e_{i,t+s}\right) &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\operatorname{cov}(\lambda_{ki}e_{it}, \lambda_{lj}e_{j,t+s}) \\
&\sim \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\operatorname{cov}(e_{it}, e_{j,t+s}) \\
&= \operatorname{cov}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{it}, \frac{1}{\sqrt{N}}\sum_{j=1}^{N}e_{j,t+s}\right).
\end{aligned}$$

We require that the terms where $i \neq j$ to be small enough such that the above expression is bounded by $\max_i \operatorname{cov}(e_{it}, e_{i,t+s})$ as $s \to \infty$. A sufficient condition is that (e_{it}, e_{js}) are not correlated for all $i \neq j$. In this case, the above expression becomes $N^{-1}\sum_{i=1}^{N}\operatorname{cov}(e_{it}, e_{i,t+s}) \leq \max_i \operatorname{cov}(e_{it}, e_{i,t+s})$.

Theorem 4.4 (CLT under nonstationary idiosyncratic errors)

Under Assumptions 4.9 and 4.11, together with the conditions in Theorem 4.2 except Assumption 3.5, the results in Theorem 4.2 and Corollary 4.1 continue to hold.

Proof. See Appendix.

5. Simulation Studies

In this section, we examine the finite sample properties of our estimator. In the first part, we report the biases and mean squared errors (MSEs) under various settings. In the second part, we simulate a one-factor model with break in factor persistence. Assuming the break time is known, we apply Corollary 4.1 to test for the presence of break. The empirical size and power are computed.

5.1. Estimating Factor Memories

As implied by Lemma 3.3, there are two sources of errors in our factor estimate: One emerges when we estimate the factor space and is induced by the idiosyncratic error terms. The other appears when we transform factors according to the identification restrictions and is induced by the presence of more than one factor. Since the latter vanishes when there is only one factor, we consider a single- and a two-factor model to disentangle these two effects.

5.1.1. One-Factor Model

The following model is simulated for 5,000 times:

$$\begin{aligned}
X_{it} &= \lambda_i F_t + e_{it}, & i &= 1,\ldots,N, t = 0,\ldots,T \\
\nabla^{d_F} F_t &= (1-\rho_F L)\varepsilon_t^F, & \varepsilon_t^F &\overset{\text{iid}}{\sim} \mathcal{N}(0,1) \\
\nabla^{d_{e_i}} e_{it} &= \varepsilon_{it}^e, & \varepsilon_{it}^e &\overset{\text{iid}}{\sim} \mathcal{N}(0,1)
\end{aligned}$$

Factor memories are set between $d_F \in (-0.5, 1.5)$. Short memory components of the latent factors are modeled through ρ_F. We set $\rho_F \in \{0, 0.5\}$ for the absence/presence of short memory in the factor. Albeit robust to the presence of short memory, FELW is still subject to a small positive bias with finite samples, which can also be observed in the simulation study of Abadir et al. (2007). Idiosyncratic shocks e_{it} are modeled as fractional integrated noises with memory parameters $d_{e_i} \sim \mathcal{U}\left(-0.5, \overline{d}_e\right)$, where $\overline{d}_e = \max\{0.5, \min\{0.5 + d_F, 1.5\}\}$. For factors with short memory (potentially anti-persistent), we allow the idiosyncratic terms to have long memory, but still restrict them to be stationary. For long memory and nonstationary factors, we allow the error terms to be nonstationary as well, but still restrict them to be below $\min\{1.5, d_F + 0.5\}$. Fractional integration is computed using the algorithm provided by Jensen and Nielsen (2014).

Simulation results are summarized in Table A.1 and A.2 for $\rho_F = 0$ and 0.5 respectively. We report the biases and MSEs of our estimator, as well as those computed with the true factors (benchmark results hereafter) for comparison. First, comparing the results in the two tables, we observe that an AR component in factors leads to a small positive bias as expected. This bias diminishes when T increases. Other than that, the presence of short memory dynamics in factors does not bring any adverse effect to our estimator. Secondly, we compare the performance of our estimator with that of the benchmark. We find that they are close to each other, except when $T = 500$ and $d_F = -0.4$. In this case our estimator is positively biased. This agrees with our theorem because Eq.(4.4) requires $T^{1-2d_F}N^{-1} = T^{1.8}N^{-1} \leq M$ for consistency. This bias can be reduced by increasing N relative to T. For the same d_F, we do not observe such bias when $T = 100$ or $N = 500$.

To remove this bias, Corollary 4.1 suggests that one can separate a long dataset into shorter parts. The average value of the estimated factor memory using each subsample is then again consistent and asymptotically normal. Therefore, we repeat the above exercise with a focus on long panels and factors with low persistence. We set $N \in \{100, 500\}$ as before, but longer time-series lengths, $T \in \{500, 1000\}$, are chosen. Since the biases are not observed for $d > 0$, we only consider $d \in [-0.4, 0]$ here. For each combination in sample size, we split the sample into b equal parts such that $N = T/b$. For each subsample, we treat it as the full sample and carry out the three-pass approach for the latent factor. We then compute the FELW with bandwidth[3] $m = \sqrt{T/b}$. We

[3]Note that in Shimotsu (2006) as well as Corollary 4.1, it is suggested that the bandwidth to be

report the biases and MSEs of the mean estimator in Table A.3. By splitting long panels to shorter parts, we can substantially reduce the magnitude of the bias when d_F is small. Specifically, the bias with $T = 500$ and $N = 100$ drops from 0.1657 to 0.0366 after using this method. This demonstrates the effectiveness of the splitting method.

5.1.2. Two-Factor Model

In this section, we consider a two-factor model to examine the effect of the presence of more than one factor on the estimation of factor memories. We assume the same data generating process stated in the last section, except that now there are two common factors. Factor memories are set such that $d_{F_1} \in (d_{F_2} - 0.5, d_{F_2} + 0.5)$ and $d_{F_2} = 0.6$. The two factors are generated independently. To save space, we only report results for $\rho_F = 0$. The effect of a positive ρ_F is the same as in the previous simulation study. Finally, we generate $\boldsymbol{\Lambda}$ such that the following identification restrictions are satisfied:

(PC1) $\boldsymbol{\Lambda}$ is a block matrix such that

$$\boldsymbol{\Lambda} = \begin{pmatrix} \underline{\boldsymbol{\lambda}}_1 & \mathbf{0} \\ \mathbf{0} & \underline{\boldsymbol{\lambda}}_2 \end{pmatrix}$$

where $\underline{\boldsymbol{\lambda}}_1$ and $\underline{\boldsymbol{\lambda}}_2$ are both vectors of length $N/2$. The factors are normalized such that $\nabla \boldsymbol{F}' \nabla \boldsymbol{F} / T = \boldsymbol{I}_2$.

(PC2) $\boldsymbol{\Lambda}$ is unrestricted except that $\lambda_{12} = 0$, and that λ_{11} and λ_{22} are not zero. The factors are normalized such that $\nabla \boldsymbol{F}' \nabla \boldsymbol{F} / T = \boldsymbol{I}_2$.

(PC3) $\boldsymbol{\Lambda} = (\boldsymbol{I}_2, \boldsymbol{\Lambda}_2')'$, while $\boldsymbol{F}_t$ and $\boldsymbol{\Lambda}_2$ are unrestricted.

---

chosen as $m = \sqrt{T}/b$ instead, so that the asymptotic variance of $\overline{d}_{\widehat{F}}$ is the same as $\widehat{d}_F$ computed with the whole series of the true factor. However, if $T = 500$ and $N = 100$, then $m = \lfloor \sqrt{500}/5 \rfloor = 4$ is too small to yield asymptotic results. Thus, we set here $m = \sqrt{500/5} = 10$ and similarly for other sample sizes.

All unrestricted elements of $\mathbf{\Lambda}$ are generated with the standard normal distribution. We set $T = 100$ and $N \in \{100, 500\}$.

Simulation results are summarized in Table A.4. Again, we compare the result with the FELW estimator applied on true factors. Both biases and MSEs have the same magnitude as the benchmark case, except when $|d_{F_1} - d_{F_2}| = 0.4$. In this case, the bias is around 0.08. This bias is not observed when PC3 is employed.

5.2. Testing Break in Memory

We now turn to evaluate how the test for break in persistence proposed by Hassler et al. (2014) performs in the context of latent factor models. We simulate a one-factor model in which the factor is generated by

$$\nabla^{d_1} F_t = \varepsilon_t, \qquad t \in [1, T/2]$$
$$\nabla^{d_2} F_t = \varepsilon_t, \qquad t \in [T/2 + 1, T]$$

A two-sided test with null hypothesis $H_0 : d_1 = d_2$ is performed using the asymptotic distribution derived in Corollary 4.1. As noted in Hurvich and Chen (2000) and Shimotsu (2006), the finite sample variance of the local Whittle (LW) estimator is larger than its asymptotic variance. Due to the theoretical similarity between FELW and LW, it is likely that FELW also shares this property. Thus, we replace m by a number c_m to improve approximation. c_m is defined as

$$c_m = \sum_{j=1}^{m} \nu_j^2, \qquad \nu_j = \log j - \frac{1}{m} \sum_{j=1}^{m} \log j.$$

Since $c_m/m \to 1$ as $m \to \infty$, the asymptotic distribution of FELW is unchanged. We set $d_1 = 0.3$ and $d_2 \in [-0.4, 1]$. The sample sizes are set to be $T \in \{200, 1000, 10000\}$ and $N \in \{100, 500\}$. We assume the break point is at the middle of the time series and is known. Thus, the actual numbers of observations for the computation of factors and factor memories are 100, 500 and 5000 respectively.

We plot the empirical probability of rejection computed with the true (left) and estimated factor (right) in Fig.A.2. We first consider the result obtained from the true factors. The test is over-sized when T is only moderate. The empirical size is around 9.4% for $T = 200$ and 8.7% for $T = 1000$. It approaches the correct size as T becomes large. For $T = 10000$, the empirical size is 5.8%. The power plot is inverted bell-shape. For $T = 200$, the power is low even for $|d_1 - d_2| = 0.7$. This is due to a large variance of the estimator. The number of frequencies used is only $m = 9$[4] and the asymptotic variance of $\widehat{d}^{(1)}_{\widehat{F}} - \widehat{d}^{(2)}_{\widehat{F}}$ under H_0 is 1/18. Two standard deviation is then around 0.47. Thus, it is hard to reject the null hypothesis. When T increases, the power improves substantially.

Now we consider the results computed with the estimated factors. The plot on the right resembles the left one very closely, except when $d_2 \leq 0$ and $N = 100$. In this case, Eq.(4.4) is not satisfied and the estimated factor memory tends to be positively biased. Thus, the test becomes less powerful. Other than that, the performance is very similar to the test with true factor.

6. Empirical Application: Macroeconomic Data

In this section, we apply the proposed estimator to FRED-MD.[5] There are 131 series after removing series with missing data. The dataset spans the period from the beginning of 1963 till the end of 2015. Instead of applying the transformation to each series as suggested in the literature, we directly apply the proposed procedures to the raw data.

[4] For each subperiod, there are 100 observations. For the computation of FELW, two initial observations are used for computation of ∇F_1 and $\nabla^2 F_1$. Thus, $m = \lfloor \sqrt{98} \rfloor = 9$.

[5] A monthly database for macroeconomic research prepared by the Federal Reserve Bank of St. Louis, see McCracken and Ng (2016) for a detailed data description.

In the top panel of Figure A.3, we plot the histogram of the raw data's integration orders. The estimated integration orders range from zero to about 1.8. In the bottom panel, we plot the histogram using differenced data. The summary statistics of the estimated integration orders are reported in Table A.5. After differencing, most series are at least close to stationary: only 8 out of the 131 series (6.1%) have memory parameters significantly larger than 0.5 at the 5% level.

We follow Bai and Ng (2013) and apply **PC2** for identification. We arrange the data such that the first eight series are (1) All Employees: Total Nonfarm Payrolls (PAYEMS), (2) Industrial Production Index (INDPRO), (3) One-Year Treasury Constant Maturity Minus Federal Funds Rate (T1YFFM), (4) Consumer Price Index for All Urban Consumers: All items less shelter (CUSR0000SA0L2), (5) One-Year Treasury Constant Maturity Rate (GS1), (6) New Private Housing Units Authorized by Building Permits (PERMIT), (7) M2 Money Stock (M2SL), and (8) S&P500 index (SP500). This arrangement mostly follows Bai and Ng (2013), except that we replace total reserves (TOTRESNS) by M2SL as the seventh variable. The reason is that TOTRESNS increased dramatically since the collapse of Lehman Brothers in September 2008 (see Keister and McAndrews (2009)). In Figure A.4, we plot the differenced series of TOTRESNS in the bottom panel. It is clearly shown that almost all variations come from the period after September 2008, and the plot is flat and close to zero in the period before. Thus, it is not suitable to be used in the identification restriction.

Due to the large variation in scales among series, we divide each differenced series by its own standard deviation. We then apply the three-pass approach to estimate the latent factors. Since **PC2** is only a local identification condition, we still need to apply sign restrictions for global identification. Thus, we set the diagonal elements of the top 8×8 submatrix of $\widehat{\mathbf{\Lambda}}$ to be positive.

As in Bai and Ng (2013), we report the marginal explanatory power of each factor in Table A.6. We regress the differenced data series on the first k and $k-1$ PANIC factors.[6] The marginal R^2 is defined as the difference between the R^2 coefficients of the two regression models. It represents the explanatory power of each factor on a series. The interpretation of each factor is similar to that in Bai and Ng (2013): the first two factors correspond to real activities; the third and fifth represent interest rates; the forth and sixth factors are the inflation and housing factors respectively; the seventh factor corresponds to the money supply; and the eighth factor is related to the stock market.

To confirm this identification result, we follow McCracken and Ng (2016) and report the ten series[7] that load most heavily on each factor in Table A.7 and A.8 for Factors 1-4 and 5-8 respectively. The first two factors correspond to real activities. The first one focuses on labor market while the second factor is a production factor. Factor three mainly affects interest rate spreads, while Factor four is related to price level. The fifth factor affects interest rates in level. The sixth factor is the housing factor. Factor seven and eight are monetary and stock factors respectively.

The estimated factors are plotted in Figure A.5. We denote the recession periods defined by NBER with the grey shaded area. On the right panel we plot the PANIC factors. Recall that it is a consistent estimator of the differenced factors. Most noticeably, we observe that factors one, two, four and eight, corresponding to the labor market, production, price and stock market respectively, experience a large negative value around 2009. In the contrary, factor seven, which is related to money and credit, experience a large positive value. These suggest that the sign restrictions are reasonable.

Next, we consider the three-pass estimator, plotted in the left panel in Figure A.5. All factors appear to be persistent and nonstationary. The first factor captures the business cycle. It falls during recession and rises otherwise. The second factor is downward

[6] Since most series and estimated factors are nonstationary, we compute the marginal R^2 with the PANIC estimators and differenced data to avoid spurious regressions.

[7] We report the FRED series ID here. A description of each series can be found in `https://files.stlouisfed.org/files/htdocs/econ/mccracken/fred-databases/Appendix_Tables_Update.pdf`.

trending. It may seem strange at first sight that the production factor is falling over time. However, it should be noted that under **PC2** the second factor is restricted to have zero effect on the first series PAYEMS. Therefore, it should be interpreted as the production factor that does not affect total employment. In fact, the upward trend of production series has already been captured in the first factor. Similarly, the stock market factor is also falling between 1960s and early 2000s. It starts to rise only after 2010.

The estimated factor memories are reported in the first row of Table A.9. The factor memories agree with those reported in the literature. For example, the high persistence level of the labor market factor is also evidenced in Caporale and Gil-Alaña (2007). The estimated integration order of the fourth factor, the price level factor, is slightly higher than 1.5. This agrees with the results in the early work of Hassler and Wolters (1995) and Baillie et al. (1996), who find that US inflation rate has integration order of about 0.4 to 0.5.

Note that $T > N$ in our dataset. It may introduce a bias as shown in our simulation results. Therefore, we split the data into two equal parts, and set the bandwidth as $\left\lfloor \sqrt{T}/2 \right\rfloor$, as suggested in Shimotsu (2006). Most noticeably, the memory parameter of the housing factor becomes larger, while that of the monetary and stock market factors becomes smaller. Finally, we apply the test for break in integration order. We find that the memory parameters of the labor market, interest rate spread, housing and monetary factors increase significantly, while that of the stock market drops. The memory parameters of the remaining factors, namely, production, price level and interest rate factors, do not have significant changes.

7. Conclusion

In this paper, we introduce the long memory factor model that allows both factors and idiosyncratic errors to have long memory. We propose a three-step procedure to estimate the latent factors. The factor space can be consistently estimated under mild assumption, and the rotation matrix $\boldsymbol{H}$ converges to the identity matrix under the identification restriction. Given the estimated factors, we apply FELW to compute factor memories. We show that the estimator is both consistent and asymptotically normal, and has the usual asymptotic variance $1/(4m)$. One drawback of our estimator is that we require N to be large relative to T, which may not be satisfied in some datasets. We show that it can be solved by splitting the data into smaller subsamples. By choosing the bandwidth accordingly, the mean value of the memory estimators obtained in different subsamples can again achieve the same convergence results.

References

Abadir, K. M., Distaso, W., Giraitis, L., 2007. Nonstationarity-extended local Whittle estimation. Journal of Econometrics 141, 1353–1384.

Abadir, K. M., Distaso, W., Giraitis, L., 2011. An I(d) model with trend and cycles. Journal of Econometrics 163, 186–199.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.

Bai, J., 2004. Estimating cross-section common stochastic trends in nonstationary panel data. Journal of Econometrics 122, 137–183.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bai, J., Ng, S., 2004. A PANIC attack on unit roots and cointegration. Econometrica 72, 1127–1177.

Bai, J., Ng, S., 2010. Panel unit root tests with cross-section dependence: A further investigation. Econometric Theory 26, 1088–1114.

Bai, J., Ng, S., 2013. Principal components estimation and identification of static factors. Journal of Econometrics 176, 18–29.

Baillie, R. T., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59.

Baillie, R. T., Chung, C. F., Tieslau, M. A., 1996. Analysing inflation by the fractionally integrated ARFIMA-GARCH model. Journal of Applied Econometrics 11, 23–40.

Beran, J., 1995. Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. Journal of the Royal Statistical Society. Series B (Methodological) 57, 659–672.

Caporale, G. M., Gil-Alaña, L. A., 2007. Nonlinearities and fractional integration in the US unemployment rate. Oxford Bulletin of Economics and Statistics 69, 521–544.

Ding, Z., Granger, C. W., Engle, R. F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83–106.

Ergemen, Y. E., 2016. Generalized efficient inference on factor models with long-range dependence.

Gil-Alaña, L. A., Robinson, P. M., 2001. Testing of seasonal fractional integration in UK and Japanese consumption and income. Journal of Applied Econometrics 16, 95–114.

Gil-Alaña, L. A., Shittu, O. I., Yaya, O. O. S., 2014. On the persistence and volatility in European, American and Asian stocks bull and bear markets. Journal of International Money and Finance 40, 149–162.

Hassler, U., Kokoszka, P. S., 2010. Impulse responses of fractionally integrated processes with long memory. Econometric Theory 26, 1855–1861.

Hassler, U., Rodrigues, P. M. M., Rubia, A., 2014. Persistence in the banking industry: Fractional integration and breaks in memory. Journal of Empirical Finance 29, 95–112.

Hassler, U., Wolters, J., 1995. Long memory in inflation rates: International evidence. Journal of Business & Economic Statistics 13, 37–45.

Hurvich, C. M., Chen, W. W., 2000. An efficient taper for potentially overdifferenced long-memory time series. Journal of Time Series Analysis 21, 155–180.

Jensen, A. N., Nielsen, M. Ø., 2014. A fast fractional difference algorithm. Journal of Time Series Analysis 35, 428–436.

Keister, T., McAndrews, J. J., 2009. Why are banks holding so many excess reserves? FRBNY Current Issues in Economics and Finance 15, 1–10.

Lahiri, S. N., 2003. A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short- and long-range dependence. Annals of Statistics 31, 613–641.

Lobato, I. N., Savin, N., 1998. Real and spurious long-memory properties of stock-market data. Journal of Business & Economic Statistics 16, 261–268.

Luciani, M., Veredas, D., 2015. Estimating and forecasting large panels of volatilities with approximate dynamic factor models. Journal of Forecasting 34, 163–176.

McCracken, M. W., Ng, S., 2016. FRED-MD: A monthly database for macroeconomic research. Journal of Business & Economic Statistics 34, 574–589.

Phillips, P. C. B., 1999. Discrete fourier transforms of fractional processes .

Phillips, P. C. B., Shimotsu, K., 2004. Local Whittle estimation in nonstationary and unit root cases. Annals of Statistics 32, 656–692.

Robinson, P. M., 1995. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630–1661.

Shimotsu, K., 2006. Simple (but effective) tests of long memory versus structural breaks.

Sibbertsen, P., 2004. Long memory in volatilities of German stock returns. Empirical Economics 29, 477–488.

Stock, J. H., Watson, M. W., 2002a. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Stock, J. H., Watson, M. W., 2002b. Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics 20, 147–162.

Su, L., Wang, X., 2017. On time-varying factor models: Estimation and testing. Journal of Econometrics 198, 84–101.

A. Proofs

In this section, we prove the results obtained in this paper. For this purpose, we provide several auxiliary lemmas in the next subsection, which are then proved in the last subsection. In the second and third subsections, we prove the lemmas and main results respectively.

A.1. Auxiliary Lemmas

Lemma A.1

1. $\|N^{-1}\sum_{i=1}^{N}\widehat{\boldsymbol{\lambda}}_i(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i)'\| = O_p\left(1/\min\{\sqrt{T}, N\}\right)$ *under Assumptions 3.1-3.4;*
2. $\widehat{\boldsymbol{\Sigma}}_\Lambda = \boldsymbol{\Sigma}_\Lambda + o_p(1)$ *under Assumptions 3.1-3.4;*
3. $N^{-1}\sum_{i=1}^{N}\widehat{\boldsymbol{\lambda}}_i e_{it} = O_p(N^{-1/2}) + O_p(T^{-1})$ *under Assumptions 3.1-3.5.*

Lemma A.2

As $m \to \infty$, we have

$$\frac{1}{m}\sum_{j=1}^{m} j^d = \begin{cases} O(m^d), & d > -1 \\ O(m^{-1}\log(m)), & d = -1 \\ O(m^{-1}), & d < -1. \end{cases}$$

Lemma A.3

For any $d \in (-0.5, 0.5)$,

1.

$$\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\zeta_k}(j)|^2 = O_p(m^{2d}T^{1-2d}C_{NT}^{-2}),$$
$$\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\nabla\zeta_k}(j)|^2 = O_p(m^{2d+2}T^{-1-2d}C_{NT}^{-2}).$$

2. *For each* $k = 1, \ldots, K$*,*

$$\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\xi_k}(j)|^2 = \begin{cases} O_p(m^{2(d-d_{\xi_k})}T^{-2(d-d_{\xi_k})}), & d - d_{\xi_k} > -0.5 \\ O_p(m^{-1}\log(m)T), & d - d_{\xi_k} = -0.5 \\ O_p(m^{-1}T^{-2(d-d_{\xi_k})}), & d - d_{\xi_k} < -0.5. \end{cases}$$

3. *If* $d_{\xi_k} \geq 0$*,*
$$\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\nabla^{-1}\xi_k}(j)|^2 = O_p(m^{-1}T^{2-2d+2d_{\xi_k}}).$$

If $d_{\xi_k} < 0$*,*

$$\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\nabla^{-1}\xi_k}(j)|^2 = \begin{cases} O_p(m^{-2+2d-2d_{\xi_k}}T^{2-2d+2d_{\xi_k}}), & d - d_{\xi_k} > 0.5 \\ O_p(m^{-1}\log(m)T^{2-2d+2d_{\xi_k}}), & d - d_{\xi_k} = 0.5 \\ O_p(m^{-1}T^{2-2d+2d_{\xi_k}}), & d - d_{\xi_k} < 0.5 \end{cases}$$

4.
$$\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\nabla\xi_k}(j)|^2 = O_p((m/T)^{2+2d-2d_{\xi_k}}) + O_p(m^{2d}T^{-1-2d}).$$

Lemma A.4

Let $w_{\nu_{2k}}(j)$ *be the DFT of the* k*-th element of* $\boldsymbol{\nu}_{2t}$*, evaluated at the* j*-th Fourier frequency.*

Then, under the conditions in Theorem 4.3,

$$w_{\nu_{2k}}(j) = \frac{1}{N\sqrt{T}} \sum_{i=1}^{N} (O_p(j^{-d_{e_i}} T^{d_{e_i}}) + O_p(T^{d_{e_i}})) \tag{A.1}$$
$$\leq O_p(j^{-\overline{d}_e} T^{\overline{d}_e - 1/2}) + O_p(T^{\overline{d}_e - 1/2}).$$

A.2. Proof of Lemmas

Proof of Lemma 3.3. Writing $X_{it} = \widehat{\boldsymbol{\lambda}}_i' \boldsymbol{F}_t + (\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i)' \boldsymbol{F}_t + e_{it}$,

$$\widehat{\boldsymbol{F}}_t = \boldsymbol{F}_t + \left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\lambda}}_i \widehat{\boldsymbol{\lambda}}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\lambda}}_i \left[(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i)' \boldsymbol{F}_t + e_{it} \right] \right)$$
$$= \left(\boldsymbol{I} + N^{-1} \widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1} \sum_{i=1}^{N} \widehat{\boldsymbol{\lambda}}_i (\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i)' \right) \boldsymbol{F}_t + N^{-1} \widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1} \sum_{i=1}^{N} \widehat{\boldsymbol{\lambda}}_i e_{it}. \tag{A.2}$$

The desired result is given by Lemma A.1. ■

Proof of Lemma 4.3. Consider the case when $p_k = 0$. We write the scaled periodogram as

$$\eta_{kj} := \frac{I_{\xi_k}(j)}{\mathfrak{f}_{\xi_k}(j)}, \qquad j = 1, \ldots, m.$$

As shown in Abadir et al. (2007), under some additional regularity conditions in Lahiri (2003), as $T \to \infty$

$$\mathbb{E}\left[\eta_{kj}\right] = 1 + o(1) + O(j^{-1} \log j), \qquad \operatorname{var}(\eta_{kj}) \leq M$$

for $1 \leq j \leq m$. Therefore,

$$\eta_j = \frac{I_{\xi_k}(j)}{\mathfrak{f}_{\xi_k}(j)} = \frac{|w_{\xi_k}(j)|^2}{\mathfrak{f}_{\xi_k}(j)} = O_p(1).$$

By Assumption 3.2,

$$|w_{\xi_k}(j)| = O_p\left(\sqrt{\mathfrak{f}_{\xi_k}(j)} \right) = O_p\left(\left(\frac{j}{T} \right)^{-d_{\xi_k}} \right).$$

Now for $p_k = 1$, by Lemma 4.4 of Abadir et al. (2007), we have the equality

$$w_{F_k}(j) = (1 - \mathrm{e}^{\mathrm{i}\omega_j})^{-1} w_{\xi_k}(j) - \frac{\mathrm{e}^{\mathrm{i}\omega_j}}{1 - \mathrm{e}^{\mathrm{i}\omega_j}} \frac{1}{\sqrt{2\pi T}} (F_{kT} - F_{k0}).$$

By Laurent series,

$$(1-\mathrm{e}^{\mathrm{i}\omega_j})^{-1} = \frac{1}{2} + \mathrm{i}(\omega_j^{-1} + o(\omega_j)) = O(\omega_j^{-1}).$$

Moreover, $F_{kT} - F_{k0} = O_p\left(T^{d_{\xi_k}+1/2}\right)$ (see e.g. Baillie (1996)). Thus,

$$\begin{aligned} w_{F_k}(j) &= O_p\left(\omega_j^{-d_{\xi_k}-1}\right) + O_p\left(\omega_j^{-1}T^{d_{\xi_k}}\right) \\ &= O_p\left(j^{-d_{F_k}}T^{d_{F_k}}\right) + O_p\left(j^{-1}T^{d_{F_k}}\right). \end{aligned}$$

■

Proof of Lemma 4.4. The discrete Fourier transform of ζ_{kt} is

$$w_{\zeta_k}(j) = (2\pi T)^{-\frac{1}{2}} \sum_{t=1}^{T} \zeta_{kt}\mathrm{e}^{\mathrm{i}t\omega_j}.$$

By the Cauchy-Schwarz inequality,

$$|w_{\zeta_k}(j)| \leq \left(\sum_{t=1}^{T} |\zeta_{kt}|^2\right)^{\frac{1}{2}} \left(\frac{1}{2\pi T}\sum_{t=1}^{T} |\mathrm{e}^{\mathrm{i}t\omega_j}|^2\right)^{\frac{1}{2}}.$$

The first term on the right hand side is $O_p(\sqrt{T}C_{NT}^{-1})$ by Lemma 3.3, while the second term is $O(1)$ by the fact that $|\mathrm{e}^{\mathrm{i}t\omega_j}| = 1$. Now consider the DFT of $\nabla\zeta_{kt}$. By Lemma 4.4 in Abadir et al. (2007),

$$w_{\nabla\zeta_k}(j) = (1-\mathrm{e}^{\mathrm{i}\omega_j})w_{\zeta_k}(j) + (2\pi T)^{-1/2}\mathrm{e}^{\mathrm{i}\omega_j}(\zeta_T - \zeta_0).$$

Since $1-\mathrm{e}^{\mathrm{i}\omega_j} = O(\omega_j)$ as $\omega_j \to 0$ by Taylor series, $w_{\nabla\zeta_k}(j) = O_p(jT^{-1/2}C_{NT}^{-1}) + O_p(T^{-1/2}C_{NT}^{-1})$ and the proof is completed. ■

A.3. Proof of Main Results

Proof of Theorem 4.1. The sufficient conditions for consistency are stated in Theorem 2.3 of Abadir et al. (2007). Consider the normalized periodogram $\eta'_{\nabla^{p_k}\widehat{F}_k,j} = I_{\nabla^{p_k}\widehat{F}_k}(j)/(b_{0k}\omega_j^{-2d_{\xi_k}})$, we shall prove the weak law of large number property

$$\frac{1}{m}\sum_{j=1}^{m}\eta'_{\nabla^{p_k}\widehat{F}_k,j}\xrightarrow{p}1. \tag{A.3}$$

Another condition is for $0 \le \gamma < 1$,

$$\frac{1}{m}\sum_{j=1}^{\lfloor\alpha m\rfloor}(j/m)^{-\gamma}\eta'_{\nabla^{p_k}\widehat{F}_k,j}\xrightarrow{p}0 \qquad \text{as } T\to\infty,\ \alpha\to 0. \tag{A.4}$$

Writing

$$\begin{aligned}\widehat{F}_{kt} &= F_{kt} + \sum_{l=1}^{K} h_{kl}F_{lt} + \zeta_{kt}\\ &= F_{kt} + \varphi_{kt} + \zeta_{kt},\end{aligned}$$

and note that

$$\begin{aligned}|w_{\nabla^{p_k}\widehat{F}}(j)|^2 &= |w_{\nabla^{p_k}F_k}(j) + w_{\nabla^{p_k}\varphi_k}(j) + w_{\nabla^{p_k}\zeta_k}(j)|^2\\ &= |w_{\xi_k}(j)|^2 + \Delta_j\end{aligned}$$

where

$$\begin{aligned}\Delta_j \le\ & 2|w_{\xi_k}(j)||w_{\nabla^{p_k}\varphi_k}(j)| + 2|w_{\xi_k}(j)||w_{\nabla^{p_k}\zeta_k}(j)| + 2|w_{\nabla^{p_k}\varphi_k}(j)||w_{\nabla^{p_k}\zeta_k}(j)|\\ &+ |w_{\nabla^{p_k}\varphi_k}(j)|^2 + |w_{\nabla^{p_k}\zeta_k}(j)|^2.\end{aligned} \tag{A.5}$$

Define respectively

$$
\begin{aligned}
S_{\xi_k}(d_{\xi_k}) &:= \frac{1}{m}\sum_{j=1}^{m}\frac{|w_{\xi_k}(j)|^2}{b_{0k}\omega_j^{-2d_{\xi_k}}},\\
S_{\nabla^{p_k}\varphi_k}(d_{\xi_k}) &:= \frac{1}{m}\sum_{j=1}^{m}\frac{|w_{\nabla^{p_k}\varphi_k}(j)|^2}{b_{0k}\omega_j^{-2d_{\xi_k}}},\\
S_{\nabla^{p_k}\zeta_k}(d_{\xi_k}) &:= \frac{1}{m}\sum_{j=1}^{m}\frac{|w_{\nabla^{p_k}\zeta_k}(j)|^2}{b_{0k}\omega_j^{-2d_{\xi_k}}}.
\end{aligned}
$$

Now writing

$$
\frac{1}{m}\sum_{j=1}^{m}\eta'_j = S_{\xi_k}(d_{\xi_k}) + R(d_{\xi_k}), \qquad \text{where } R(d_{\xi_k}) \le \frac{1}{m}\sum_{j=1}^{m}\frac{\Delta_j}{b_0\omega_j^{-2d_{\xi_k}}}.
$$

$S_{\xi_k}(d_{\xi_k}) \xrightarrow{p} 1$ by Robinson (1995) and

$$
\begin{aligned}
R(d_{\xi_k}) \le\ & 2S_{\nabla^{p_k}F_k}(d_{\xi_k})^{\frac{1}{2}}S_{\nabla^{p_k}\varphi_k}(d_{\xi_k})^{\frac{1}{2}} + 2S_{\nabla^{p_k}F_k}(d_{\xi_k})^{\frac{1}{2}}S_{\nabla^{p_k}\zeta_k}(d_{\xi_k})^{\frac{1}{2}}\\
&+ 2S_{\nabla^{p_k}F_k}(d_{\varphi_k})^{\frac{1}{2}}S_{\nabla^{p_k}\zeta_k}(d_{\xi_k})^{\frac{1}{2}} + S_{\nabla^{p_k}\varphi_k} + S_{\nabla^{p_k}\zeta_k}.
\end{aligned}
$$

Thus, it suffices to show $S_{\nabla^{p_k}\varphi_k} \xrightarrow{p} 0$ and $S_{\nabla^{p_k}\zeta_k} \xrightarrow{p} 0$ to prove Eq.(A.3). By part 1 of Lemma A.3,

$$
\begin{aligned}
S_{\zeta_k} &= O_p(m^{2d_{\xi_k}}T^{1-2d_{\xi_k}}C_{NT}^{-2}) = o_p(1)\\
S_{\nabla\zeta_k} &= O_p(m^{2d_{\xi_k}+2}T^{-1-2d_{\xi_k}}C_{NT}^{-2}) = o_p(1)
\end{aligned}
\tag{A.6}
$$

due to Assumption 4.5. Besides, for $p_k = 0$,

$$
\begin{aligned}
S_{\varphi_k}(d_{\xi_k}) &\le K \max_{1\le l\le K}\frac{1}{m}\sum_{j=1}^{m} b_{0k}\omega_j^{2d_{\xi_k}}|h_{kl}|^2|w_{F_l}(j)|^2\\
&= O_p(T^{-1})\max_{1\le l\le K}\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d_{\xi_k}}|w_{F_l}(j)|^2.
\end{aligned}
$$

Suppose $d_{F_l} = \max_k d_{F_k}$. If $p_l = 0$, i.e., when all factors are stationary,

$$S_{\varphi_k}(d_{\xi_k}) = \begin{cases} O_p(m^{2(d_{\xi_k}-d_{\xi_l})}T^{-2(d_{\xi_k}-d_{\xi_l})-1}), & d_{\xi_k} - d_{\xi_l} > -0.5 \\ O_p(m^{-1}\log(m)), & d_{\xi_k} - d_{\xi_l} = -0.5 \\ O_p(m^{-1}T^{-2(d_{\xi_k}-d_{\xi_l})-1}), & d_{\xi_k} - d_{\xi_l} < -0.5. \end{cases}$$

The last case is ruled out since it is assumed that $d_{F_k} - d_{F_l} \geq -0.5$ for all k, l. Moreover, if $p_k = 0$ and $p_l = 1$, then $d_{\xi_l} < 0$. In this case,

$$S_{\varphi_k}(d_{\xi_k}) = \begin{cases} O_p(m^{2(d_{\xi_k}-d_{F_l})}T^{-2(d_{\xi_k}-d_{F_l})-1}), & d_{\xi_k} - d_{F_l} > -0.5 \\ O_p(m^{-1}\log(m)), & d_{\xi_k} - d_{F_l} = -0.5 \\ O_p(m^{-1}T^{-2(d_{\xi_k}-d_{F_l})-1}), & d_{\xi_k} - d_{F_l} < -0.5. \end{cases}$$

The last case is again ruled out. Therefore, $S_{\varphi_k}(d_{\xi_k}) = o_p(1)$ in all cases. Since the case $p_k = p_l = 1$ is the same as that of $p_k = p_l = 0$, the proof for Eq.(A.3) is completed.

To show (A.4), write

$$T_{\nabla^{p_k}\zeta_k}(\alpha) := \frac{1}{m}\sum_{j=1}^{\lfloor \alpha m \rfloor}\left(\frac{j}{m}\right)^{-\gamma}\omega_j^{2d_{\xi_k}}|w_{\nabla^{p_k}\zeta_k}(j)|^2.$$

Without loss of generality, let $\widetilde{m} = \lfloor \alpha m \rfloor \to \infty$. Since $|w_{\zeta_k}(j)|^2 = O_p(TC_{NT}^{-2})$,

$$\begin{aligned} T_{\zeta_k}(\alpha) &= \left(\frac{1}{\widetilde{m}}\sum_{j=1}^{\widetilde{m}} j^{-\gamma+2d_{\xi_k}}\right)\alpha m^{\gamma}T^{-2d_{\xi_k}}O_p(TC_{NT}^{-2}) \\ &= \begin{cases} O_p(\alpha^{2d_{\xi_k}-\gamma+1}m^{2d_{\xi_k}}T^{1-2d_{\xi_k}}C_{NT}^{-2}), & 2d_{\xi_k} - \gamma > -1 \\ O_p(m^{\gamma-1}\log(\widetilde{m})T^{1-2d_{\xi_k}}C_{NT}^{-2}), & 2d_{\xi_k} - \gamma = -1 \\ O_p(m^{\gamma-1}T^{1-2d_{\xi_k}}C_{NT}^{-2}), & 2d_{\xi_k} - \gamma < -1 \end{cases} \\ &= \begin{cases} o_p(m^{2d_{\xi_k}}T^{1-2d_{\xi_k}}C_{NT}^{-2}), & 2d_{\xi_k} - \gamma > -1 \\ o_p(T^{1-2d_{\xi_k}}C_{NT}^{-2}), & 2d_{\xi_k} - \gamma \leq -1 \end{cases} \end{aligned}$$

Now for $p_k = 1$, since $|w_{\nabla\zeta_k}(j)|^2 = O_p(j^2 T^{-1} C_{NT}^{-2})$,

$$
\begin{aligned}
T_{\nabla\zeta_k}(\alpha) &= \left(\frac{1}{\widetilde{m}} \sum_{j=1}^{\widetilde{m}} j^{2-\gamma+2d_{\xi_k}}\right) \alpha m^{\gamma} T^{-2d_{\xi_k}} O_p(T^{-1} C_{NT}^{-2}) \\
&= O_p(\alpha^{3-\gamma+2d_{\xi_k}} m^{2+2d_{\xi_k}} T^{-1-2d_{\xi_k}} C_{NT}^{-2}) \\
&= o_p(m^{2d_{F_k}} T^{1-2d_{F_k}} C_{NT}^{-2})
\end{aligned}
$$

Finally, we consider

$$
T_{\nabla^{p_k}\varphi_k}(\alpha) := \frac{1}{m} \sum_{j=1}^{\lfloor \alpha m \rfloor} \left(\frac{j}{m}\right)^{-\gamma} \omega_j^{2d_{\xi_k}} |w_{\nabla^{p_k}\varphi_k}(j)|^2.
$$

Again let $d_{F_l} = \max_k d_{F_k}$,

$$
T_{\nabla^{p_k}\varphi_k}(\alpha) \le O_p(T^{-1}) \frac{1}{m} \sum_{j=1}^{\lfloor \alpha m \rfloor} \left(\frac{j}{m}\right)^{-\gamma} \omega_j^{2d_{\xi_k}} |w_{\nabla^{p_k} F_l}(j)|^2.
$$

For $p_k = p_l = 0$, note that $\omega_j^{2d_{\xi_k}} |w_{\nabla^{p_k} F_l}(j)|^2 = O_p(j^{2d_{\xi_k} - 2d_{\xi_l}} T^{-(2d_{\xi_k} - 2d_{\xi_l})})$,

$$
\begin{aligned}
T_{\varphi_k}(\alpha) &\le O_p(T^{-1-(2d_{\xi_k} - 2d_{\xi_l})}) \frac{1}{\widetilde{m}} \sum_{j=1}^{\widetilde{m}} j^{-\gamma+2d_{\xi_k}-2d_{\xi_l}} \alpha m^{\gamma} \\
&= \begin{cases} O_p(\alpha^{1-\gamma} m^{2d_{\xi_k}-2d_{\xi_l}} T^{-1-(2d_{\xi_k}-2d_{\xi_l})}), & 2d_{\xi_k} - 2d_{\xi_l} > \gamma - 1 \\ O_p(m^{\gamma-1} \log(\widetilde{m}) T^{-1-(2d_{\xi_k}-2d_{\xi_l})}), & 2d_{\xi_k} - 2d_{\xi_l} = \gamma - 1 \\ O_p(m^{\gamma-1} T^{-1-(2d_{\xi_k}-2d_{\xi_l})}), & 2d_{\xi_k} - 2d_{\xi_l} < \gamma - 1 \end{cases} \\
&= \begin{cases} o_p(m^{2d_{F_k}-2d_{F_l}} T^{-1-(2d_{F_k}-2d_{F_l})}), & 2d_{F_k} - 2d_{F_l} > \gamma - 1 \\ o_p(T^{-1-(2d_{F_k}-2d_{F_l})}), & 2d_{F_k} - 2d_{F_l} \le \gamma - 1 \end{cases}
\end{aligned}
$$

For $p_k = 0$ and $p_l = 1$, $d_{\xi_l} < 0$ and

$$\begin{aligned}
T_{\varphi_k}(\alpha) &\le O_p(T^{-1-(2d_{\xi_k}-2d_{F_l})})\frac{1}{\widetilde{m}}\sum_{j=1}^{\widetilde{m}} j^{-\gamma+2d_{\xi_k}-2d_{F_l}}\alpha m^{\gamma} \\
&= \begin{cases} O_p(\alpha^{1-\gamma} m^{2d_{\xi_k}-2d_{F_l}} T^{-1-(2d_{\xi_k}-2d_{F_l})}), & 2d_{\xi_k}-2d_{F_l} > \gamma-1 \\ O_p(m^{\gamma-1})\log(\widetilde{m})T^{-1-(2d_{\xi_k}-2d_{F_l})}), & 2d_{\xi_k}-2d_{F_l} = \gamma-1 \\ O_p(m^{\gamma-1})T^{-1-(2d_{\xi_k}-2d_{F_l})}), & 2d_{\xi_k}-2d_{F_l} < \gamma-1 \end{cases} \\
&= \begin{cases} o_p(m^{2d_{F_k}-2d_{F_l}} T^{-1-(2d_{F_k}-2d_{F_l})}), & 2d_{F_k}-2d_{F_l} > \gamma-1 \\ o_p(T^{-1-(2d_{F_k}-2d_{F_l})}), & 2d_{F_k}-2d_{F_l} \le \gamma-1 \end{cases}
\end{aligned}$$

The case $p_k = p_l = 1$ is the same as when $p_k = p_l = 0$. Thus, $T_{\nabla^{p_k}\varphi_k}(\alpha) = o_p(1)$. The same arguments in the proof of Eq.(A.3) apply and thus Eq.(A.4) is proved. ■

Proof of Theorem 4.2. In view of the proof of Theorem 2.5 in Abadir et al. (2007), we shall consider

$$m^{-1}\sum_{j=1}^{m}(\log(j/m)+1)\,\eta'_{\nabla^{p_k}\widehat{F}_k,j} = m^{-1}\sum_{j=1}^{m}(\log(j/m)+1)\,\eta'_{\xi_k,j} + q_m, \tag{A.7}$$

where $\eta'_{\xi_k,j} = |w_{\xi_k}(j)|^2/(b_{0k}\omega_j^{-2d_{\xi_k}})$ and similarly for $\eta'_{\nabla^{p_k}\widehat{F}_k,j}$. We shall prove that $q_m = o_p(m^{-1/2})$ to ensure the error terms to be small enough such that it is still negligible after being multiplied by $m^{1/2}$, and so it does not affect the asymptotic distribution. Under our assumptions,

$$m^{-1}\sum_{j=1}^{m}(\log(j/m)+1)\,\eta'_{\xi_k,j} = O_p(m^{-1/2}).$$

Thus, it suffices to show that

$$m^{-1}\sum_{j=1}^{m}(\log(j/m)+1)\,\omega^{2d_{\xi_k}}|w_{\nabla^{p_k}\zeta_k,j}|^2 = o_p(m^{-1/2}) \tag{A.8}$$

$$m^{-1}\sum_{j=1}^{m}(\log(j/m)+1)\,\omega^{2d_{\xi_k}}|w_{\nabla^{p_k}\varphi_k,j}|^2 = o_p(m^{-1/2}) \tag{A.9}$$

By Assumption 4.7 and Eq.(A.6),

$$\left| m^{-1} \sum_{j=1}^{m} \left(\log(j/m)+1\right) \omega^{2d_{\xi_k}} |w_{\nabla^{p_k}\zeta_k,j}|^2 \right| \leq O(\log(m)) S_{\nabla^{p_k}\zeta_k,j}(d_{\xi_k}) = o_p(m^{-1/2}).$$

The same can be proved for Eq.(A.9). ■

Proof of Corollary 4.1. As shown in Theorem 4.2, $\widehat{d}^{(a_1)}_{\widehat{F}_k} = \widehat{d}^{(a_1)}_{F_k} + o_p((m/b)^{-1/2})$ and $\widehat{d}^{(a_2)}_{\widehat{F}_k} = \widehat{d}^{(a_2)}_{F_k} + o_p((m/b)^{-1/2})$ respectively. Thus,

$$\sqrt{\frac{m}{b}}(\widehat{d}^{(a_1)}_{\widehat{F}_k} - \widehat{d}^{(a_2)}_{\widehat{F}_k}) = \sqrt{\frac{m}{b}}(\widehat{d}^{(a_1)}_{F_k} - \widehat{d}^{(a_2)}_{F_k}) + o_p(1).$$

Moreover, Shimotsu (2006) showed that $\widehat{d}^{(a_1)}_{F_k}$ and $\widehat{d}^{(a_2)}_{F_k}$ are asymptotically independent when $p_k = 0$. The result can be extended to $p_k = 1$ because the asymptotic expression of estimation error $\widehat{d}_{\widehat{F}_k} - d_{F_k}$ depends only on $\eta'_{\xi_k,j}$, which is independent of p_k (cf. Eq.(4.5)). The result follows from the asymptotic normality of the estimated memories. ■

Proof of Theorem 4.3. In view of Lemma A.4, it can be shown analogously to the proofs of Theorem 4.1 and 4.2 that $\boldsymbol{\nu}_{2t}$ does not affect the asymptotic properties of our estimator if $\max_i d_{e_i} - d_{F_k} \leq 0.5$. Hence, it suffices to show the error term $\boldsymbol{\nu}_{1t}$ satisfies either Eq.(2.40) or Eq.(2.41) and Eq.(2.42) in Abadir et al. (2007). For $p_k = 0$, we prove the former.

$$\sum_{t=1}^{T} \|\nabla \boldsymbol{\nu}_{1t}\| = \sum_{t=1}^{T} \left\| \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\lambda}_i \nabla e_{it} \right\| = O_p(TN^{-1/2}) = O_p(T^{-1/2+d_{\xi_k}})$$

since by Assumption 4.11, $T^{3/2-d_{\xi_k}}N^{-1/2} \leq M$. For $p_k = 1$, we prove Eq.(2.41) and Eq.(2.42) in Abadir et al. (2007). For $d_{\xi_k} \geq 0$,

$$\sum_{t=1}^{T} \|\nabla \boldsymbol{\nu}_{1t}\|^2 = \sum_{t=1}^{T} \left\| \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\lambda}_i \nabla e_{it} \right\|^2 = O_p(TN^{-1}) = O_p(T^{2d_{\xi_k}})$$

because $T^{1-2d_{\xi_k}}N^{-1} = T^{3-2d_{F_k}}N^{-1} \leq M$. For $d_{\xi_k} < 0$,

$$\sum_{t=1}^{T} \|\nabla \boldsymbol{\nu}_{1kt}\| = \sum_{t=1}^{T} \left\| \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\lambda}_i \nabla e_{it} \right\| = O_p(TN^{-1/2}) = O_p(T^{1/2+d_{\xi_k}}).$$

Thus, $\boldsymbol{\nu}_{1t}$ satisfies the conditions of Theorem 2.5(ii) in Abadir et al. (2007). ■

Proof of Theorem 4.4. Treating $\sqrt{N}\boldsymbol{\nu}_{1t}$ and $\sqrt{T}\boldsymbol{\nu}_{2t}$ as $I(\bar{d}_e)$ processes, Theorem 4.4 can be proved following the same lines as Theorem 4.1 and 4.2. ■

A.4. Proof of Auxiliary Lemmas

Proof of Lemma A.1. 1. By Cauchy-Schwarz inequality, let $\boldsymbol{\nu}_i = \widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i$,

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\lambda}}_i(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i)'\right\| &\leq \left\|\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\nu}_i\boldsymbol{\nu}_i'\right\| + \left\|\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\boldsymbol{\nu}'\right\| \\ &= (I) + (II).\end{aligned}$$

Consider (I),

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\nu}_i\boldsymbol{\nu}_i'\right\| &\leq \frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\nu}_i\|^2 \\ &= O_p(1/\min\{T, N^2\})\end{aligned}$$

due to Lemma 3.1 and 3.2. Similarly,

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\boldsymbol{\nu}_i'\right\| &\leq \left(\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\lambda}_i\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\nu}_i\|^2\right)^{1/2} \\ &= O_p\left(1/\min\{\sqrt{T}, N\}\right).\end{aligned}$$

Thus, $(I) + (II) = O_p\left(1/\min\{\sqrt{T}, N\}\right)$.

2. Letting $\boldsymbol{\nu}_i = \widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i$,

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_\Lambda &= \frac{1}{N}\sum_{i=1}^N \widehat{\boldsymbol{\lambda}}_i \widehat{\boldsymbol{\lambda}}_i' \\
&= \frac{1}{N}\sum_{i=1}^N (\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_i)(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_i)' \\
&= \frac{1}{N}\sum_{i=1}^N \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i + \frac{1}{N}\sum_{i=1}^N \boldsymbol{\nu}_i\boldsymbol{\lambda}_i' + \frac{1}{N}\sum_{i=1}^N \boldsymbol{\lambda}_i\boldsymbol{\nu}_i' + \frac{1}{N}\sum_{i=1}^N \boldsymbol{\nu}_i\boldsymbol{\nu}_i' \\
&\xrightarrow{p} \boldsymbol{\Sigma}_\Lambda
\end{aligned}
$$

due to Assumption 3.1.

3. Note that

$$
\begin{aligned}
\widehat{\boldsymbol{\lambda}}_i &= \boldsymbol{\lambda}_i + \left(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}^{*-1}\boldsymbol{\lambda}_i\right) + \left(\boldsymbol{H}^{*-1} - \boldsymbol{I}\right)\boldsymbol{\lambda}_i \\
&= \boldsymbol{\lambda}_i + \boldsymbol{G}'\left(\widetilde{\boldsymbol{\lambda}}_i - \boldsymbol{H}^{-1}\boldsymbol{\lambda}_i\right) + \left(\boldsymbol{H}^{*-1} - \boldsymbol{I}\right)\boldsymbol{\lambda}_i
\end{aligned}
$$

where $\boldsymbol{G}$ equals to $\boldsymbol{I}$ under **PC1**, $\boldsymbol{Q}'$ under **PC2** and $(\widetilde{\boldsymbol{\Lambda}}_1^{-1})'$ under **PC3**. Therefore, we have

$$
\frac{1}{N}\sum_{i=1}^N \widehat{\boldsymbol{\lambda}}_i e_{it} = \boldsymbol{H}^{*-1}\frac{1}{N}\sum_{i=1}^N \boldsymbol{\lambda}_i e_{it} + \boldsymbol{G}'\frac{1}{N}\sum_{i=1}^N (\widetilde{\boldsymbol{\lambda}}_i - \boldsymbol{H}^{-1}\boldsymbol{\lambda}_i)e_{it}.
$$

The first term is $O(N^{-1/2})$ by Assumption 3.5. Following Bai (2003),

$$
\widetilde{\boldsymbol{\lambda}}_i - \boldsymbol{H}^{-1}\boldsymbol{\lambda}_i = \boldsymbol{H}'\frac{1}{T}\sum_{s=1}^T \boldsymbol{f}_s \nabla e_{is} + \boldsymbol{\epsilon}_i
$$

where $\boldsymbol{\epsilon}_i = O_p(1/\min\{N,T\})$.

$$
\begin{aligned}
\frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T \boldsymbol{f}_s \nabla e_{is} e_{it} &= \frac{1}{NT}\sum_{s=1}^T \boldsymbol{f}_s \nabla \boldsymbol{e}_s'\boldsymbol{e}_t \\
&= \frac{1}{NT}\sum_{s=1}^T \boldsymbol{f}_s \mathbb{E}\left[\nabla \boldsymbol{e}_s'\boldsymbol{e}_t\right] + \frac{1}{NT}\sum_{s=1}^T \boldsymbol{f}_s(\nabla \boldsymbol{e}_s'\boldsymbol{e}_t - \mathbb{E}\left[\nabla \boldsymbol{e}_s'\boldsymbol{e}_t\right]).
\end{aligned}
$$

Since $\max_s \|\boldsymbol{f}_s\| = O_p(1)$, and due to Assumption 3.5,

$$\left\| \frac{1}{NT} \sum_{s=1}^{T} \boldsymbol{f}_s \mathbb{E}\left[\nabla \boldsymbol{e}_s' \boldsymbol{e}_t\right] \right\| \le \frac{1}{NT} \max_s \|\boldsymbol{f}_s\| \left\| \sum_{s=1}^{T} \mathbb{E}\left[\nabla \boldsymbol{e}_s' \boldsymbol{e}_t\right] \right\| \\ = O_p(T^{-1}).$$

Now consider the second term,

$$\begin{aligned} &\frac{1}{NT} \sum_{s=1}^{T} \boldsymbol{f}_s (\nabla \boldsymbol{e}_s' \boldsymbol{e}_t - \mathbb{E}\left[\nabla \boldsymbol{e}_s' \boldsymbol{e}_t\right]) \\ =&\frac{1}{NT} \sum_{s=1}^{T} (\boldsymbol{f}_s \nabla \boldsymbol{e}_s' \boldsymbol{e}_t - \mathbb{E}\left[\boldsymbol{f}_s \nabla \boldsymbol{e}_s' \boldsymbol{e}_t\right]) + \frac{1}{NT} \sum_{s=1}^{T} (\boldsymbol{f}_s \mathbb{E}\left[\nabla \boldsymbol{e}_s' \boldsymbol{e}_t\right] - \mathbb{E}\left[\boldsymbol{f}_s \nabla \boldsymbol{e}_s' \boldsymbol{e}_t\right]) \\ =&O_p((NT)^{-1/2}) + O_p(T^{-1}) \end{aligned}$$

again by Assumption 3.5. Finally,

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\epsilon}_i e_{it} \right\| \le \left(\frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{\epsilon}_i\|^2 \right)^{1/2} \left(\frac{1}{N} \sum_{i=1}^{N} e_{it}^2 \right)^{1/2} = O_p\left(1/\min\{N,T\}\right).$$

Collecting terms,

$$\frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\lambda}}_i e_{it} = O_p\left(N^{-1/2}\right) + O_p\left(T^{-1}\right). \qquad \blacksquare$$

Proof of Lemma A.2. 1. If $d > -1$, by dividing m^d

$$\frac{1}{m} \sum_{j=1}^{m} \left(\frac{j}{m} \right)^d \to \int_0^1 x^d dx = \left[\frac{1}{d+1} x^{d+1} \right]_0^1 = \frac{1}{d+1} = O(1).$$

2. If $d = -1$,

$$\frac{1}{m} \sum_{j=1}^{m} j^{-1} < m^{-1}(\log(m) + 1) = O(m^{-1} \log(m)).$$

3. If $d < -1$,

$$\frac{1}{m}\sum_{j=1}^{m} j^d \leq m^{-1}M = O(m^{-1}). \qquad \blacksquare$$

Proof of Lemma A.3. 1. Since $|w_{\zeta_k}(j)| = O_p(\sqrt{T}C_{NT}^{-1})$,

$$\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\zeta_k}(j)|^2 = O_p(T^{1-2d}C_{NT}^{-2})\frac{1}{m}\sum_{j=1}^{m} j^{2d}.$$

The summation part is $O(m^{2d})$ by Lemma A.2 since $2d > -1$. Thus the result follows. We can prove the second part in the same way.

2. First note that by Lemma 4.3, for each k we have

$$\omega_j^{2d_{\xi_k}}|w_{\xi_k}(j)|^2 = O_p(1)$$

Thus, for each $1 \leq j \leq m$ and for all d, $\omega_j^{2d}|w_{\xi_k}(j)|^2 = O_p(j^{2d-2d_{\xi_k}}T^{2d_{\xi_k}-2d})$ and

$$\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\xi_k}(j)|^2 = O_p(1)T^{2d_{\xi_k}-2d}\frac{1}{m}\sum_{j=1}^{m} j^{2d-2d_{\xi_k}}.$$

The result follows from Lemma A.2.

3. For $d_{\xi_k} \geq 0$, $w_{\nabla^{-1}\xi_k}(j) = O_p(j^{-1}T^{d_{\xi_k}+1})$.

$$\begin{aligned}\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\nabla^{-1}\xi_k}(j)|^2 &= O_p(1)\frac{1}{m}\sum_{j=1}^{m} j^{2d-1}T^{2-2d+2d_{\xi_k}} \\ &= O_p(m^{-1}T^{2-2d+2d_{\xi_k}})\end{aligned}$$

since $2d - 2 < -1$. For $d_{\xi_k} \geq 0$, $w_{\nabla^{-1}\xi_k}(j) = O_p((j/T)^{-d_{\xi_k}-1})$.

$$\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\nabla^{-1}\xi_k}(j)|^2 = O_p(1)T^{2-2d+2d_{\xi_k}}\frac{1}{m}\sum_{j=1}^{m} j^{-2+2d-2d_{\xi_k}}$$

and the result follows again by Lemma A.2.

4. Since $w_{\nabla\xi_k}(j) = O_p((j/T)^{1-d_{\xi_k}}) + O_p(T^{-1/2})$,

$$\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}|w_{\nabla\xi_k}(j)|^2 = O_p(1)\left[\frac{1}{m}\sum_{j=1}^{m} j^{2+2d-2d_{\xi_k}}T^{-2-2d+2d_{\xi_k}} + \frac{1}{m}\sum_{j=1}^{m} j^{2d}T^{-2d-1}\right]$$
$$= O_p((m/T)^{2+2d-2d_{\xi_k}}) + O_p(m^{2d}T^{-1-2d})$$

since $2 + 2d - 2d_{\xi_k} > -1$. ■

Proof of Lemma A.4. Since DFT is a linear operator, and $\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i$ does not vary over time,

$$w_{\nu_{2k}}(j) = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\lambda}_{ik} - \lambda_{ik})w_{e_i}(j).$$

Now $\widehat{\lambda}_{ik} - \lambda_{ik} = O_p(T^{-1/2})$ by Lemma 3.1 and 3.2, while $w_{e_i}(j) = O_p(j^{-d_{e_i}}T^{d_{e_i}}) + O_p(T^{d_{e_i}})$ can be proved following the same lines of the proof of Lemma 4.3. Thus, Eq.(A.1) follows. ■

Table A.1: Biases and MSEs of FELW estimators for $\rho_F = 0$

	$N = 100$				$N = 500$			
d_F	bias($\widehat{d}_F$)	bias($\widehat{d}_{\widehat{F}}$)	mse($\widehat{d}_F$)	mse($\widehat{d}_{\widehat{F}}$)	bias($\widehat{d}_F$)	bias($\widehat{d}_{\widehat{F}}$)	mse($\widehat{d}_F$)	mse($\widehat{d}_{\widehat{F}}$)
(i) $T = 100$								
-0.4	-0.0080	0.0337	0.0758	0.0771	-0.0054	0.0023	0.0782	0.0768
-0.2	-0.0275	-0.0110	0.0756	0.0738	-0.0287	-0.0269	0.0796	0.0789
0.0	-0.0304	-0.0278	0.0749	0.0752	-0.0336	-0.0336	0.0777	0.0769
0.2	-0.0290	-0.0259	0.0838	0.0831	-0.0331	-0.0342	0.0822	0.0827
0.4	-0.0258	-0.0248	0.0837	0.0835	-0.0161	-0.0170	0.0833	0.0822
0.6	-0.0227	-0.0198	0.0767	0.0756	-0.0201	-0.0203	0.0759	0.0761
0.8	-0.0253	-0.0237	0.0720	0.0722	-0.0304	-0.0304	0.0714	0.0707
1.0	-0.0299	-0.0233	0.0730	0.0725	-0.0401	-0.0363	0.0751	0.0754
1.2	-0.0358	-0.0326	0.0743	0.0741	-0.0257	-0.0218	0.0743	0.0741
1.4	-0.0277	-0.0273	0.0672	0.0668	-0.0368	-0.0352	0.0715	0.0709
(ii) $T = 500$								
-0.4	0.0039	0.1615	0.0210	0.0491	0.0037	0.0508	0.0209	0.0231
-0.2	-0.0157	0.0351	0.0202	0.0210	-0.0146	-0.0035	0.0205	0.0207
0.0	-0.0148	-0.0045	0.0209	0.0207	-0.0171	-0.0147	0.0203	0.0203
0.2	-0.0134	-0.0027	0.0207	0.0208	-0.0152	-0.0139	0.0205	0.0206
0.4	0.0043	0.0128	0.0251	0.0252	-0.0032	-0.0016	0.0257	0.0256
0.6	0.0028	0.0087	0.0201	0.0202	0.0054	0.0060	0.0200	0.0201
0.8	-0.0127	-0.0081	0.0200	0.0201	-0.0082	-0.0070	0.0200	0.0198
1.0	-0.0159	-0.0092	0.0198	0.0198	-0.0177	-0.0145	0.0211	0.0210
1.2	-0.0134	-0.0107	0.0212	0.0209	-0.0132	-0.0113	0.0212	0.0211
1.4	-0.0084	-0.0072	0.0190	0.0189	-0.0043	-0.0032	0.0203	0.0201

Note: The table reports the biases and MSEs of FELW estimators of factor memories. The data generating process is described in Section 5.1.1 with $\rho_F = 0$. $\widehat{d}_F$ and $\widehat{d}_{\widehat{F}}$ are FELW estimators of memory parameters of the true factor series $\{F_t\}$ and the factor estimate $\{\widehat{F}_t\}$ defined in Section 3 respectively.

Table A.2: Biases and MSEs of the FELW estimator for $\rho_F = 0.5$

	$N = 100$				$N = 500$			
d_F	bias($\widehat{d}_F$)	bias($\widehat{d}_{\widehat{F}}$)	mse($\widehat{d}_F$)	mse($\widehat{d}_{\widehat{F}}$)	bias($\widehat{d}_F$)	bias($\widehat{d}_{\widehat{F}}$)	mse($\widehat{d}_F$)	mse($\widehat{d}_{\widehat{F}}$)
(iii) $T = 100$								
-0.4	0.0900	0.1011	0.0833	0.0849	0.0906	0.0915	0.0860	0.0858
-0.2	0.0797	0.0831	0.0828	0.0815	0.0798	0.0792	0.0861	0.0858
0.0	0.0831	0.0828	0.0855	0.0858	0.0790	0.0787	0.0849	0.0850
0.2	0.0887	0.0882	0.0950	0.0950	0.0837	0.0824	0.0916	0.0920
0.4	0.0855	0.0838	0.0891	0.0893	0.0966	0.0955	0.0909	0.0900
0.6	0.0826	0.0820	0.0831	0.0827	0.0861	0.0844	0.0830	0.0826
0.8	0.0830	0.0817	0.0818	0.0813	0.0780	0.0776	0.0800	0.0796
1.0	0.0852	0.0882	0.0836	0.0845	0.0739	0.0766	0.0843	0.0849
1.2	0.0806	0.0818	0.0797	0.0795	0.0898	0.0911	0.0829	0.0827
1.4	0.0783	0.0794	0.0715	0.0715	0.0704	0.0718	0.0754	0.0754
(iv) $T = 500$								
-0.4	0.0204	0.0773	0.0210	0.0275	0.0204	0.0335	0.0211	0.0216
-0.2	0.0069	0.0216	0.0200	0.0201	0.0083	0.0111	0.0203	0.0205
0.0	0.0088	0.0113	0.0208	0.0208	0.0067	0.0073	0.0202	0.0201
0.2	0.0109	0.0138	0.0210	0.0212	0.0090	0.0091	0.0206	0.0207
0.4	0.0270	0.0287	0.0245	0.0246	0.0189	0.0192	0.0247	0.0247
0.6	0.0213	0.0226	0.0199	0.0200	0.0232	0.0233	0.0199	0.0199
0.8	0.0096	0.0107	0.0200	0.0200	0.0138	0.0142	0.0203	0.0203
1.0	0.0079	0.0105	0.0197	0.0198	0.0061	0.0080	0.0209	0.0210
1.2	0.0112	0.0128	0.0214	0.0215	0.0119	0.0129	0.0217	0.0217
1.4	0.0173	0.0175	0.0194	0.0193	0.0210	0.0215	0.0207	0.0206

Note: The table reports the biases and MSEs of FELW estimators of factor memories. The data generating process is described in Section 5.1.1 with $\rho_F = 0.5$. $\widehat{d}_F$ and $\widehat{d}_{\widehat{F}}$ are FELW estimators of memory parameters of the true factor series $\{F_t\}$ and the factor estimate $\{\widehat{F}_t\}$ defined in Section 3 respectively.

Table A.3: Biases and MSEs of the FELW estimator for long panels

	$N = 100$				$N = 500$			
d_F	bias($\overline{d}_F$)	bias($\overline{d}_{\widehat{F}}$)	mse($\overline{d}_F$)	mse($\overline{d}_{\widehat{F}}$)	bias($\overline{d}_F$)	bias($\overline{d}_{\widehat{F}}$)	mse($\overline{d}_F$)	mse($\overline{d}_{\widehat{F}}$)
(v) $T = 500$								
-0.4	-0.0030	0.0366	0.0157	0.0144	-0.0012	0.0479	0.0207	0.0229
-0.3	-0.0146	0.0101	0.0161	0.0132	-0.0059	0.0170	0.0207	0.0211
-0.2	-0.0275	-0.0113	0.0164	0.0134	-0.0103	0.0008	0.0195	0.0193
-0.1	-0.0310	-0.0219	0.0166	0.0135	-0.0129	-0.0082	0.0205	0.0203
0.0	-0.0336	-0.0269	0.0165	0.0135	-0.0175	-0.0158	0.0208	0.0208
(vi) $T = 1000$								
-0.4	-0.0039	0.0355	0.0078	0.0078	0.0037	0.0526	0.0104	0.0131
-0.3	-0.0157	0.0099	0.0078	0.0065	-0.0047	0.0189	0.0100	0.0101
-0.2	-0.0241	-0.0082	0.0086	0.0066	-0.0112	-0.0000	0.0103	0.0102
-0.1	-0.0308	-0.0214	0.0085	0.0066	-0.0161	-0.0112	0.0103	0.0103
0.0	-0.0322	-0.0273	0.0091	0.0074	-0.0144	-0.0122	0.0102	0.0102

Note: The table reports the biases and MSEs of FELW estimators of factor memories. The data generating process is described in Section 5.1.1 with $\rho_F = 0$. The data is split into b non-overlapping windows such that $T/b = N$. For each $a = 1, \ldots, b$, we estimate the latent factor using data within $t \in \mathcal{T}_a = \{(a-1)T/b + 1, \ldots, aT/b\}$ and obtain an estimate of factor memory. The final estimate of factor memory $\overline{d}_{\widehat{F}}$ is computed as the mean value of estimators across subsamples. Similarly, $\overline{d}_F$ is average FELW estimator of memory parameters of the true factor series $\{F_t\}$ in different subsamples.

Table A.4: Biases and MSEs of the FELW estimator under two-factor model

d_{F_1}	bias($\widehat{d}_{F_1}$)	bias($\widehat{d}_{F_2}$)	bias($\widehat{d}_{\widehat{F}_1}$)	bias($\widehat{d}_{\widehat{F}_2}$)	mse($\widehat{d}_{F_1}$)	mse($\widehat{d}_{F_2}$)	mse($\widehat{d}_{\widehat{F}_1}$)	mse($\widehat{d}_{\widehat{F}_2}$)
(xi) **PC1**: $N = 100, T = 100$								
0.2	-0.0262	-0.0152	0.0731	-0.0277	0.0842	0.0753	0.0977	0.0779
0.4	-0.0199	-0.0114	0.0072	-0.0223	0.0810	0.0757	0.0823	0.0769
0.6	-0.0187	-0.0099	-0.0181	-0.0065	0.0753	0.0751	0.0758	0.0760
0.8	-0.0286	-0.0118	-0.0369	0.0126	0.0731	0.0750	0.0734	0.0744
1.0	-0.0271	-0.0024	-0.0398	0.0588	0.0769	0.0745	0.0773	0.0811
(xii) **PC1**: $N = 500, T = 100$								
0.2	-0.0277	-0.0178	0.0880	-0.0289	0.0797	0.0747	0.0975	0.0747
0.4	-0.0247	-0.0213	0.0007	-0.0320	0.0849	0.0776	0.0857	0.0788
0.6	-0.0172	-0.0190	-0.0172	-0.0200	0.0765	0.0759	0.0753	0.0746
0.8	-0.0248	-0.0080	-0.0350	0.0172	0.0710	0.0769	0.0719	0.0765
1.0	-0.0338	-0.0052	-0.0453	0.0632	0.0762	0.0746	0.0777	0.0824
(xiii) **PC2**: $N = 100, T = 100$								
0.2	-0.0278	-0.0147	0.0848	-0.0417	0.0806	0.0752	0.1054	0.0813
0.4	-0.0219	-0.0184	0.0164	-0.0357	0.0842	0.0738	0.0842	0.0759
0.6	-0.0176	-0.0151	-0.0162	-0.0126	0.0724	0.0746	0.0730	0.0746
0.8	-0.0329	-0.0124	-0.0507	0.0229	0.0720	0.0763	0.0742	0.0759
1.0	-0.0344	-0.0007	-0.0572	0.0761	0.0742	0.0715	0.0791	0.0872
(xiv) **PC2**: $N = 500, T = 100$								
0.2	-0.0293	-0.0184	0.0777	-0.0469	0.0776	0.0767	0.1035	0.0840
0.4	-0.0185	-0.0175	0.0208	-0.0381	0.0839	0.0768	0.0862	0.0800
0.6	-0.0250	-0.0223	-0.0251	-0.0210	0.0755	0.0779	0.0728	0.0770
0.8	-0.0243	-0.0126	-0.0404	0.0155	0.0732	0.0759	0.0737	0.0800
1.0	-0.0260	-0.0045	-0.0516	0.0769	0.0735	0.0729	0.0803	0.0875
(xc) **PC3**: $N = 100, T = 100$								
0.2	-0.0303	-0.0152	0.0042	-0.0202	0.0803	0.0769	0.0815	0.0759
0.4	-0.0271	-0.0246	-0.0138	-0.0283	0.0836	0.0732	0.0844	0.0738
0.6	-0.0191	-0.0165	-0.0174	-0.0144	0.0732	0.0742	0.0715	0.0732
0.8	-0.0235	-0.0202	-0.0278	-0.0124	0.0690	0.0732	0.0688	0.0745
1.0	-0.0345	-0.0125	-0.0406	0.0137	0.0723	0.0745	0.0725	0.0746
(xiv) **PC3**: $N = 500, T = 100$								
0.2	-0.0368	-0.0165	-0.0079	-0.0203	0.0805	0.0736	0.0830	0.0740
0.4	-0.0250	-0.0183	-0.0172	-0.0207	0.0828	0.0744	0.0824	0.0744
0.6	-0.0200	-0.0153	-0.0186	-0.0150	0.0730	0.0746	0.0731	0.0739
0.8	-0.0235	-0.0141	-0.0270	-0.0048	0.0723	0.0742	0.0730	0.0716
1.0	-0.0343	-0.0190	-0.0384	0.0074	0.0742	0.0719	0.0737	0.0711

Note: The table reports the biases and MSEs of FELW estimators of factor memories. The data generating process is described in Section 5.1.1 with two factors. Memory of the second factor is fixed as $d_{F_2} = 0.6$, while that of the first factor varies as listed in the table. Factor loadings are generated according to restrictions described in Section 5.1.2. Factors are estimated under the respective transformation described in Section 3. $\widehat{d}_{\widehat{F}_1}$ and $\widehat{d}_{\widehat{F}_2}$ are respectively FELW estimators of the memory parameters of the first and second estimated factors, while $\widehat{d}_{F_1}$ and $\widehat{d}_{\widehat{F}_1}$ are respectively the FELW estimators of the memory parameters of the true and estimated factors.

Table A.5: Summary statistics of the estimated integration orders of data series in level and difference

	Mean	Min	5%	25%	Median	75%	95%	Max
$\widehat{d}_X$	1.0683	−0.0550	0.2664	0.8965	1.0764	1.3405	1.6894	1.7873
$\widehat{d}_{\nabla X}$	0.0729	−1.0000	−0.7751	−0.1028	0.0787	0.3421	0.6897	0.8165

Note: This table reports the summary statistics of estimated integration orders of the raw and differenced data series. The estimated integration orders are computed by FELW with $m = \lfloor T^{0.5} \rfloor$.

Table A.6: Marginal R^2 of each factor on the restricted variables (using M2SL)

	Factor							
Series	1	2	3	4	5	6	7	8
(1)PAYEMS	0.8384	0.0001	0.0000	0.0006	0.0001	0.0000	0.0114	0.0008
(2)INDPRO	0.3953	0.4837	0.0000	0.0000	0.0000	0.0000	0.0004	0.0000
(3)T1YFFM	0.0176	0.0004	0.9013	0.0000	0.0000	0.0000	0.0001	0.0000
(4)CUSR0000SA0L2	0.0077	0.0019	0.0007	0.7457	0.0000	0.0000	0.0004	0.0000
(5)GS1	0.0610	0.0046	0.0253	0.0035	0.8121	0.0000	0.0000	0.0000
(6)PERMIT	0.0089	0.0003	0.0166	0.0003	0.0129	0.5383	0.0002	0.0000
(7)M2SL	0.0063	0.0318	0.0005	0.0615	0.0002	0.0083	0.4219	0.0011
(8)SP500	0.0218	0.0089	0.0010	0.0220	0.0022	0.0032	0.0068	0.4585

Note: This table reports the marginal explanatory power of each factor on the first eight series. The (i, j)-th element of the above table is the marginal R^2, denoted as $mR_i(j)$. It is computed as the difference between the R^2 coefficient of regressing the i-th series on the first j factors, and the R^2 coefficient of regressing the i-th series on the first $j-1$ factors. The factors are computed by the three-pass procedure with identification restrictions **PC2** with eight factors. The first eight factors are listed in the first column.

Table A.7: Highest marginal R^2 - Factor 1-4

$mR^2(1)$		$mR^2(2)$	
PAYEMS	0.8384	IPCONGD	0.5860
USGOOD	0.6928	IPFINAL	0.5836
USTPU	0.6807	IPFPNSS	0.5385
SRVPRD	0.6392	INDPRO	0.4837
USWTRADE	0.6224	IPMANSICS	0.4579
MANEMP	0.5209	CUMFNS	0.4549
USTRADE	0.4838	IPDCONGD	0.4502
DMANEMP	0.4746	IPMAT	0.2877
USCONS	0.4691	IPDMAT	0.2446
IPMANSICS	0.4311	IPNCONGD	0.2206
$mR^2(3)$		$mR^2(4)$	
T1YFFM	0.9013	CUSR0000SA0L5	0.8760
TB6SMFFM	0.8436	CPIAUCSL	0.8741
T5YFFM	0.8356	CPIULFSL	0.8725
T10YFFM	0.7699	CUSR0000SAC	0.8593
TB3SMFFM	0.7016	DNDGRG3M086SBEA	0.8331
AAAFFM	0.6716	PCEPI	0.8189
BAAFFM	0.5774	CPITRNSL	0.8060
COMPAPFFx	0.5265	CUSR0000SA0L2	0.7457
FEDFUNDS	0.3533	WPSFD49207	0.7304
GS10	0.2137	WPSFD49502	0.7292

Note: This table lists the 10 series that load most heavily on the first eight factors. For each factor, we report the ten largest marginal R^2 and the respective series.

Table A.8: Highest marginal R^2 - Factor 5-8

$mR^2(5)$		$mR^2(6)$	
GS1	0.8121	HOUST	0.5412
TB6MS	0.8114	PERMIT	0.5383
TB3MS	0.7517	PERMITMW	0.3697
CP3Mx	0.7310	PERMITS	0.3051
GS5	0.6458	HOUSTS	0.2759
GS10	0.5396	HOUSTMW	0.2280
FEDFUNDS	0.5362	CES0600000007	0.1732
AAA	0.4923	ISRATIOx	0.1485
BAA	0.3817	PERMITNE	0.1365
BAAFFM	0.2521	HOUSTNE	0.1342
$mR^2(7)$		$mR^2(8)$	
M2SL	0.4219	S.P.500	0.4585
CPIMEDSL	0.4144	S.P..indust	0.4532
MZMSL	0.3845	S.P.div.yield	0.3220
CUSR0000SAS	0.3522	VXOCLSx	0.3096
DSERRG3M086SBEA	0.3462	EXCAUSx	0.2329
CES0600000008	0.2635	NAPM	0.1838
NONREVSL	0.1639	NAPMNOI	0.1688
CES3000000008	0.1340	S.P.PE.ratio	0.1399
M1SL	0.1312	NAPMPI	0.1322
M2REAL	0.1250	NONBORRES	0.1246

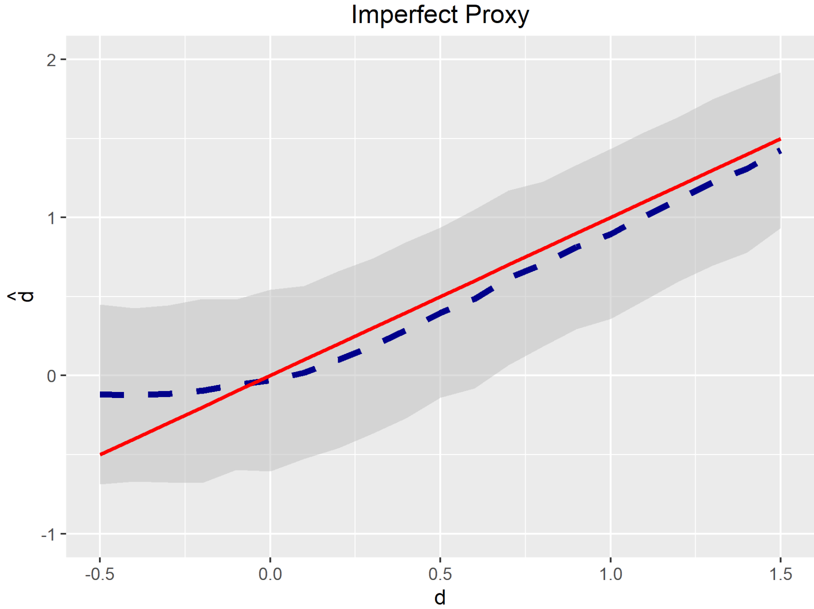
Note: This table lists the 10 series that load most heavily on the first eight factors. For each factor, we report the ten largest marginal R^2 and the respective series.

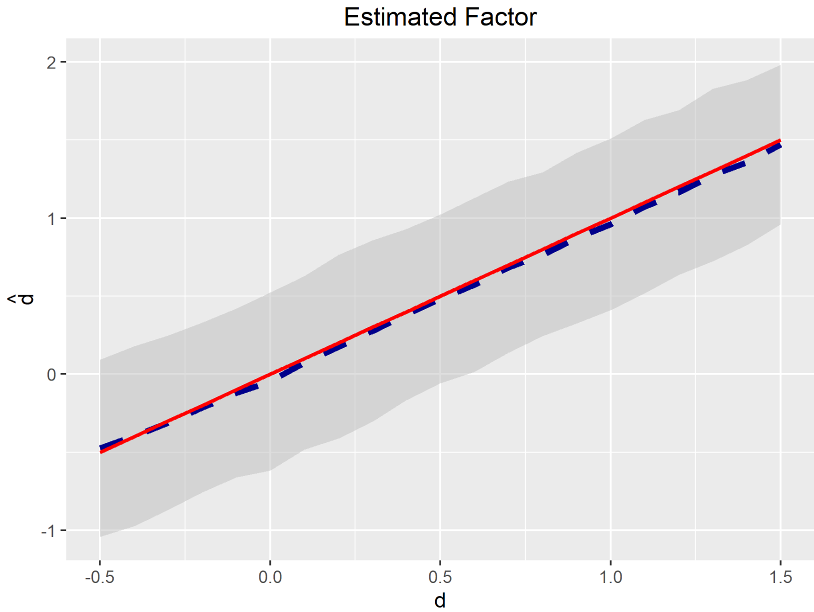
Table A.9: Estimated factor memories

	Factor							
Estimator	1	2	3	4	5	6	7	8
Full sample								
$\widehat{d}_{\widehat{F}}$	1.3011	0.8967	0.6572	1.5737	1.0628	0.8893	1.5000	1.3463
Split sample								
$\overline{d}_{\widehat{F}}$	1.2105	0.9685	0.8215	1.7580	1.2391	1.2769	1.1923	0.7631
$\widehat{d}^{(1)}_{\widehat{F}}$	0.9480	0.9957	0.5000	1.6535	1.1294	1.0538	0.9774	1.0263
$\widehat{d}^{(2)}_{\widehat{F}}$	1.4729	0.9413	1.1430	1.8626	1.3489	1.5000	1.4072	0.5000
Change in memory	✓		✓			✓	✓	✓

Note: This table reports the estimated integration order of each factor. $\widehat{d}_{\widehat{F}}$ is the factor memory computed with full sample, while $\widehat{d}^{(a)}_{\widehat{F}}$ is the factor memory computed with the a-th subsample, and $\overline{d}_{\widehat{F}}$ is the average value among the factor memories computed with subsamples. The test of change in memory is taken at the 5% level.

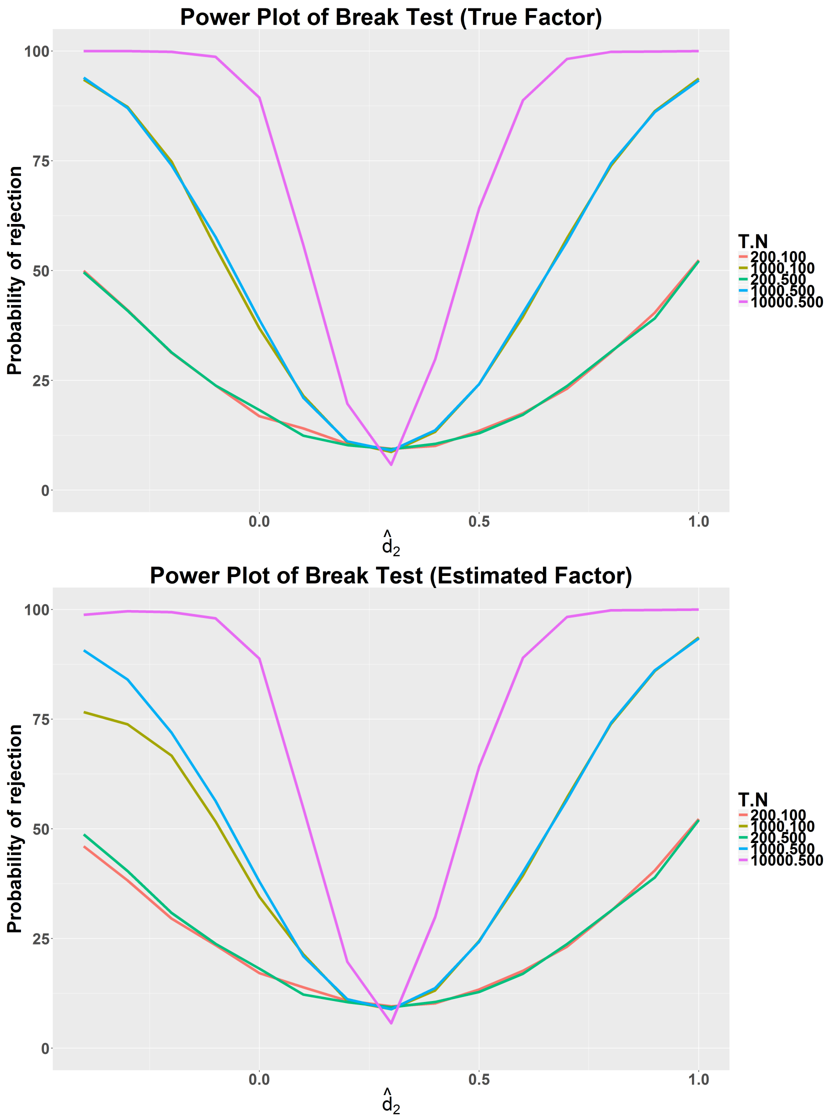
Figure A.1: FELW estimation of the proxy (left) and the estimated factor (right)





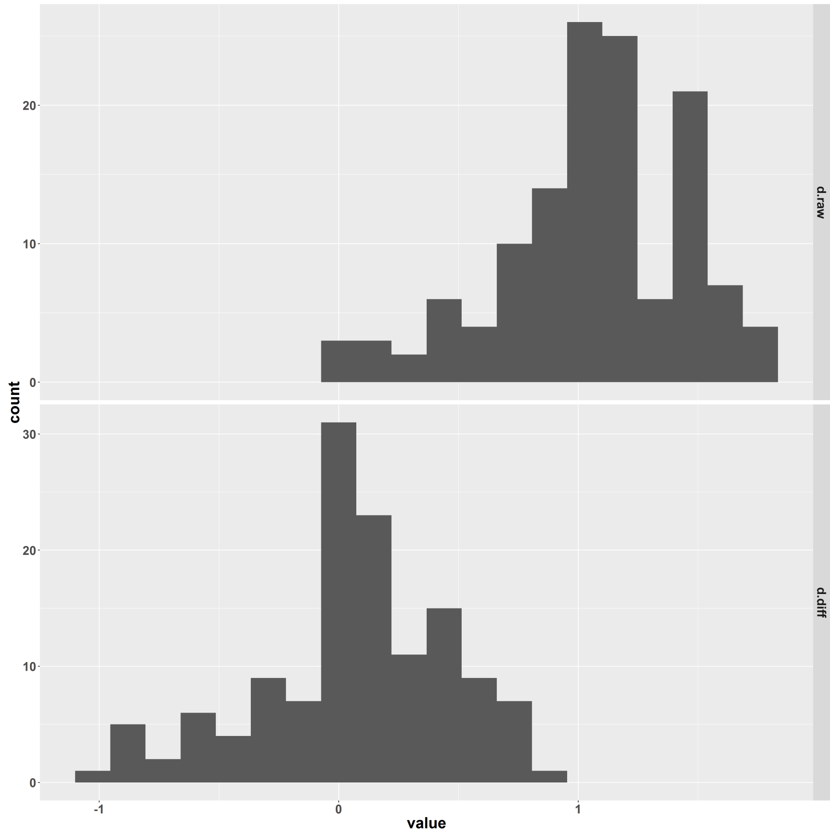
Note: The above figures plot the estimated memory of the imperfect proxy Z_t (left) and the estimated factor $\widehat{F}_t$ (right) respectively. The dotted lines represent the average over 1000 replications. The gray shades areas is error bound of the estimator, computed as two standard deviations above and below the mean of the estimators. The red lines are the true values of factor memories.

Figure A.2: Empirical probability of rejection of break test with true factor (left) and estimated factor (right)



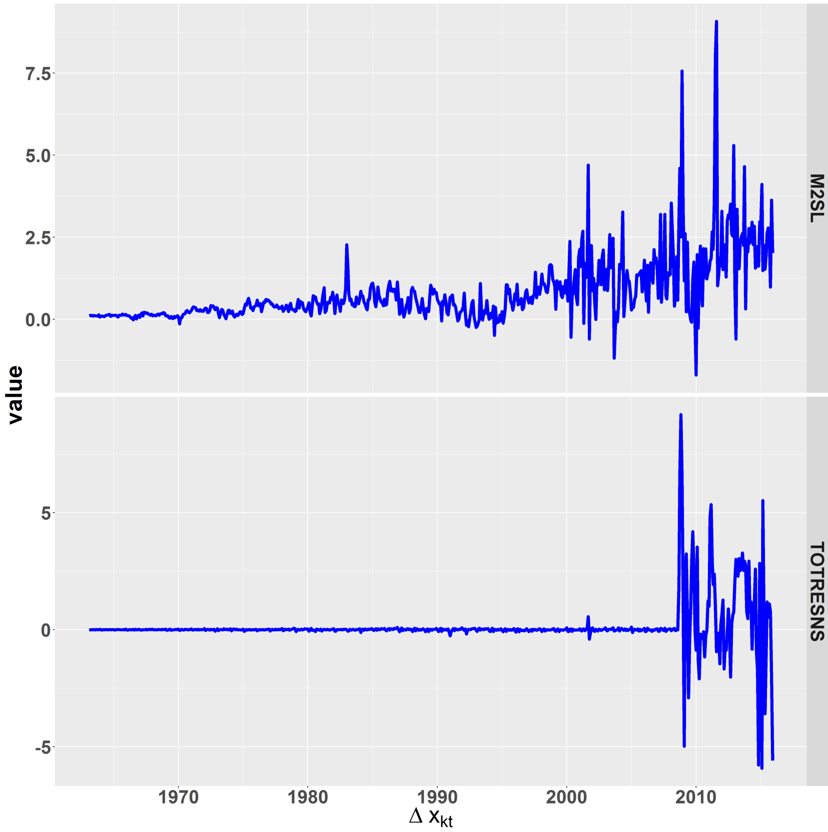
Note: The above figures plot the probabilities of rejection in a test against the null hypothesis of no break in factor memory. Graph in the left (right) panel is created with the true (estimated) factors.

Figure A.3: Estimated integration orders of raw and differenced data series



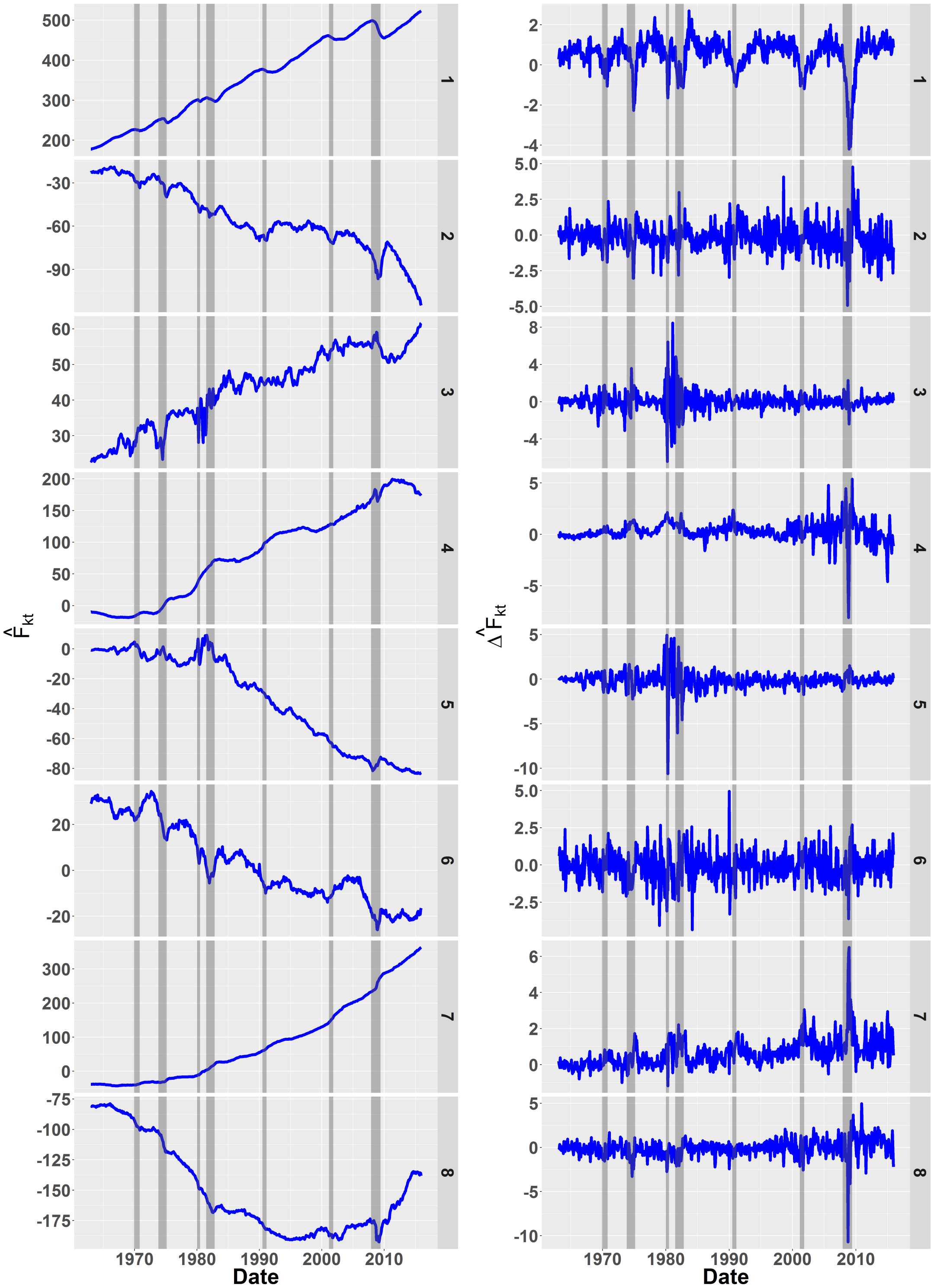
Note: The above figures plot the histograms of estimated integration order of the data series in FRED-MD. The plot on top (bottom) is created with raw (differenced) data. The estimated integration orders are computed by FELW with $m = \lfloor T^{0.5} \rfloor$.

Figure A.4: Differenced M2SL and TOTRESNS series



Note: The top panel of the above figure plots the differenced M2SL series. The bottom panel plots the differenced TOTRESNS series.

Figure A.5: Estimated factors with three-pass approach (left) and estimated differenced factors with PANIC approach(right)



Note: This figure plots the estimated factors. On the left panel we plot the three-pass estimators, while the PANIC estimators (i.e. PCA estimator applied on differenced data) are plotted on the right. The shaded area denotes the recession periods defined by NBER.