DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

Center for Research in Econometric Analysis of Time Series

Generalized Efficient Inference on Factor Models with Long-Range Dependence

Yunus Emre Ergemen

CREATES Research Paper 2016-5

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Generalized Efficient Inference on Factor Models with Long-Range Dependence*

Yunus Emre Ergemen†

January 29, 2016

Abstract

A dynamic factor model is considered that contains stochastic time trends allowing for stationary and nonstationary long-range dependence. The model nests standard $I(0)$ and $I(1)$ behaviour smoothly in common factors and residuals, removing the necessity of a priori unit-root and stationarity testing. Short-memory dynamics are allowed in the common factor structure and possibly heteroskedastic error term. In the estimation, a generalized version of the principal components (PC) approach is proposed to achieve efficiency. Asymptotics for efficient common factor and factor loading as well as long-range dependence parameter estimates are justified at standard parametric convergence rates. The use of the method for the selection of number of factors and testing for latent components is discussed. Finite-sample properties of the estimates are explored via Monte-Carlo experiments, and an empirical application to U.S. economy diffusion indices is included.

Keywords: Factor models; long-range dependence; principal components; efficiency; hypothesis testing.
JEL Codes: C12, C13, C33

---

*I am grateful to Carlos Velasco, Javier Hualde and Peter M. Robinson for their constructive comments. I acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

†CREATES and Department of Economics and Business Economics, Aarhus University. Fuglesangs Allé 4, Building 2628, 8210 Aarhus V, DENMARK. E-mail address: ergemen@econ.au.dk.

1 Introduction

Consider the factor model for an observable array $\{x_{it}, i \geq 1, t \geq 1\}$:

$$x_{it} = \psi_i(L)' g_t + e_{it} \tag{1}$$

where g_t is a $k \times 1$ vector of dynamic unobservable common factors, and $\psi_i(L) = \psi_{0i} + \psi_{1i}L + \ldots + \psi_{mi}L^m$ is a $k \times 1$ polynomial of factor loadings, with k initially unknown. In (1), the idiosyncratic errors are modeled as

$$e_{it} = \Delta_t^{-\vartheta_i} \rho_i(L) \epsilon_{it}$$

where $\rho_i(L) = 1 - \rho_{1,i}L - \ldots - \rho_{p_i,i}L^{p_i}$ and ϑ_i introduce short-memory dynamics and fractional long-range dependence, respectively, that may differ across cross-section units. The vector of common factors can also display fractional long-range dependence,

$$g_t = \Delta_t^{-\delta_0} z_t,$$

where δ_0 is common for all factors and when $\delta_0 > \vartheta_{max}$, x_{it} is essentially an $I(\delta_0)$ series.

With $\Delta = 1 - L$, L denoting the lag operator such that $L^j y_t = y_{t-j}$, Δ^ξ has the expansion

$$\Delta^\xi = \sum_{j=0}^{\infty} \pi_j(\xi) L^j, \quad \pi_j(\xi) = \frac{\Gamma(j - \xi)}{\Gamma(-\xi)\Gamma(j + 1)}$$

for $\xi > 0$, and Δ_t^ξ truncates this expansion to

$$\Delta_t^\xi = \sum_{j=0}^{t} \pi_j(\xi) L^j.$$

The fractional differencing operator Δ^ξ introduces possible stationary (when $0 < \xi < 1/2$) and nonstationary (when $\xi \geq 1/2$) long memory and the truncation of this filter is motivated by the allowance for $\xi \geq 1/2$, when $\Delta^{-\xi}$ does not converge.

The structures on the error and the common factor vector allow for short-memory dynamics and general long-range dependence that removes the necessity of a priori unit-root or stationarity testing. This is particularly important because it is impossible to identify the source of persistence via unit-root testing when data is solely generated by unobservables as in (1). With the allowance of general persistence characteristics, the model in (1) also generalizes Bai and Ng (2004)'s framework in which they model observable series assuming $I(1)$ common factors while the idiosyncratic errors can be either $I(0)$ or $I(1)$.

Several commonly used alternative specifications are readily nested under (1). For instance, the model corresponds to a fractionally integrated panel data model with fixed effects as proposed by Robinson and Velasco (2015) when $f_t = (1, \ldots, 1)'$ and $\psi_i(L) = \psi_i$. Similarly, f_t might correspond to deterministic trends in which case, (1) corresponds to the model employed by Gao and Robinson (2014) when $\psi_i(L) = \psi_i$.

In the literature, generalized principal components estimation has been considered to obtain efficient factor structure estimates, possibly among many others, by Breitung and Tenhofen (2011), Choi (2011) and Choi (2012). In these recent works, the attention was restricted to tackling serial dependence and heteroskedasticity in factor models. Breitung and Tenhofen (2011) and Choi (2012) have stationary $I(0)$ setups while Choi (2011) allows for $I(1)$ common factors. Traditionally, only $I(0)$ and $I(1)$ cases are considered, see also Bai and Ng (2004) and Bai and Ng (2013), but the possibility of fractional long-range dependence is overlooked although empirical studies show evidence for it. For example, Luciani and Veredas (2012) and Ergemen and Velasco (2015) find that financial-market realized volatility serving as a common factor in a panel of industry realized volatilities is fractionally integrated.

The primary interest in using (1) is in the efficient estimation of the common factor structure when there is not only serial dependence and heteroskedasticity but also possible fractional long-range dependence in the model. This paper contributes to the literature in many ways. First, generalized PC estimation for fractionally integrated factor models is introduced for the first time. Second, efficient estimation of the common factor structure components is shown, which will lead to better forecasts, more precise policy analysis and higher local power in hypothesis testing. Third, the model can be used parsimoniously in empirical analyses, in contrast to the available ones in the literature that require indicator-by-indicator a priori transformations.

Among others, Bai and Ng (2004) first-difference their model containing $I(1)$ common factors alongside $I(0)$ or $I(1)$ idiosyncratic errors to obtain consistent PC estimates of the common-factor structure and thus of the error term under large N and T. We propose a three-step procedure to obtain efficient factor structure estimates. We first get initial estimates of the common factor structure adopting Bai and Ng (2004)'s method and then estimate the memory of the common factors parametrically from them, which we use at the last step to obtain the efficient factor structure estimates. We establish the asymptotics for these estimates. We also include a simulation study based on Monte Carlo experiments and an empirical application to US diffusion indices to highlight the parsimony of the method.

Throughout the paper, the notation "$(N,T)_j$" denotes joint asymptotics under which both cross-section size N and time-series length T are increasing; " $\to_d$ " denotes convergence in distribution, and $\|A\| = (trace(AA'))^{1/2}$. All mathematical proofs and intermediate technical results are collected in an appendix at the end of the paper.

2 First Differences Estimation of Common Factors

We first-difference (1) to obtain

$$\Delta x_{it} = \psi_i(L)'\Delta g_t + \Delta e_{it}. \tag{2}$$

The differenced model in (2) has a static factor representation. First let us write

$$\Delta X_t = \Psi \Delta G_t + \Delta e_t$$

where $X_t = (x_{1t}, \ldots, x_{Nt})'$, $\Psi = (\Psi_0, \ldots, \Psi_m)$ that is obtained from $\Psi(L) = \Psi_0 + \Psi_1 L + \ldots + \Psi_m L^m$ with $\Psi_j = (\psi_{j1}, \ldots, \psi_{jN})'$ $(j = 0, \ldots, m)$, $G_t = (g_t', \ldots, g_{t-m}')'$, and $e_t = (e_{1t}, \ldots, e_{Nt})'$. Let $r \leq (m+1)k$ be the rank of Ψ. Then there exists an $N \times r$ matrix Λ such that $\Psi \Delta G_t = \Lambda \Delta F_t$, where $\Delta F_t = R \Delta G_t$ and R is a nonsingular $r \times (m+1)k$ matrix. The model in (2) can then be written in the static representation using full matrix notation as

$$\Delta X = \Delta F \Lambda' + \Delta e \tag{3}$$

where $X = (X_2, \ldots, X_T)'$ and $e = (e_2, \ldots, e_T)'$ are $T \times N$ matrices. The columns of the $T \times r$ matrix $F = (F_2, \ldots, F_T)'$ collect the observations of the r static factors. For each i and t, equation (3) can be written as

$$\Delta x_{it} = \lambda_i' \Delta F_t + \Delta e_{it}. \tag{4}$$

Let M denote a generic positive constant, and set $\vartheta_{max} = \max_i \vartheta_{i0}$. Then, we introduce the following conditions to study (4).

Assumption 1. $\delta_0 \in \mathcal{D} = [\underline{\delta}, \bar{\delta}] \subseteq [0, 1]$ and $\vartheta_{max} \leq \delta_0 < \underline{\delta} + 1/2$.

Assumption 2. λ_i are either deterministic satisfying $\|\lambda_i\| \leq M$ or they are random and satisfy $E\|\lambda_i\|^4 \leq M$, and $N^{-1}\sum_{i=1}^{N} \lambda_i \lambda_i' \to_p \Sigma_\Lambda > 0$.

Assumption 3. $z_t \sim iid(0, \Sigma_z)$, $\Sigma_z > 0$, and $E\|z_t\|^4 < M$.

Assumption 4. (i) $\epsilon_{it} \sim iid(0, \sigma_i^2)$, $E|\epsilon_{it}|^8 \leq M$, $\sum_{j=0}^{\infty} j|\rho_{ij}| < M$, $\omega_i^2 = \rho_i(1)\sigma_i^2 > 0$;

(ii) $E(\epsilon_{it}\epsilon_{jt}) = \tau_{ij}$ with $\sum_{i=1}^{N} |\tau_{ij}| \leq M$ for all j;

(ii) For all (t, s), $E\left|N^{-1/2}\sum_{i=1}^{N} [\epsilon_{is}\epsilon_{it} - E(\epsilon_{is}\epsilon_{it})]\right|^4 \leq M$.

Assumption 5. ϵ_{it}, z_t and λ_i are mutually independent groups.

Assumption 6. $E\|F_0\| \leq M$, and for all i, $E|e_{i0}| \leq M$.

Assumption 1 restricts the allowed range of factor memory parameter values so that the first-differenced model is stationary and allows for both cointegrated (when $\delta_0 > \vartheta_{max}$) and spurious cases (when $\delta_0 = \vartheta_{max}$). So for example, Bai and Ng (2004)'s nonstationary setup is nested when $\delta_0 = \vartheta_i = 1$, and the stationary setups of Stock and Watson (2002) and Bai and Ng (2002) are nested when $\delta_0 = \vartheta_i = 0$ for all i. The condition $\delta_0 - \underline{\delta} < 1/2$ together with $\delta_0 \geq \vartheta_{max}$ implies that the maximum error memory not be too apart from the lower bound of the allowed range of values for δ_0 and is necessary to establish consistency of the memory parameter at the standard $\sqrt{T}$ convergence rate. It could be dispensed with but then the convergence rate would depend on unknown memory parameters, rendering the use of the method in practice impossible.

The remaining conditions, i.e. Assumptions 2-6, are standard in the PC estimation of common factor structures. Assumption 2 ensures the identification of the factor structure. Assumption 3 imposes a standard moment condition on the iid innovations of common factors. Assumption 4(i) allows for weak serial correlation, and 4(ii) and 4(iii) allow for weak cross-section correlation so (1) has an approximate factor model structure. Assumption 5 imposes independence among model components, which might be stronger than needed but it makes the proofs much more tractable. Finally, Assumption 6 bounds the initial conditions, which is commonly used in the analysis of nonstationary $I(1)$ variables whose study generally requires a priori differencing.

Let

$$f_t := \Delta F_t,$$

then from (3), the principal component estimate of f_t, say $\hat{f}_t$, is obtained as $\sqrt{T-1}$ times the r eigenvectors corresponding to the r largest eigenvalues of $(\Delta X)(\Delta X)'/N(T-1)$ because the minimization problem

$$V(r) = \min_{\Lambda,F} \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta x_{it} - \lambda_i' f_t)^2$$

is identical to maximizing $tr(f'(\Delta X)(\Delta X)'f)$ concentrating out Λ and using the identifying restriction

$$f^{'}f/(T-1) = I_r.$$

Then the loadings matrix Λ is estimated by

$$\hat{\Lambda} = (\Delta X)'\hat{f}/(T-1).$$

These estimates span the factor space and the corresponding factor loadings space, respectively, only up to a rotation. That is, $\hat{f}$ estimates fH_f and $\hat{\Lambda}$ estimates $\Lambda H_f^{'-1}$ where the transpose of the $r \times r$ rotation matrix, H_f, is defined as

$$H_f' = \hat{V}^{-1}(\hat{f}^{'}f/(T-1))(\Lambda'\Lambda/N) \tag{5}$$

with $\hat{V} = \hat{\Lambda}'\hat{\Lambda}/N$.

We have the following result that establishes the consistency of first-differenced common factor estimates.

Theorem 1. *Under Assumptions 1-6 and if $\Lambda'\Lambda/N$ is diagonal with distinct entries,*

$$\min\left\{\sqrt{N}, T\right\}(\hat{f}_t - f_t) = O_p(1), \text{ for a fixed } t.$$

The convergence rate for common factor estimates is $\min\left\{\sqrt{N}, T\right\}$ because the factor loadings are also unknown and need to be estimated. If the factor loadings were known, the convergence rate would be just $\sqrt{N}$. The result contrasts with those established by Bai (2003) and Bai and Ng (2004) in that the extra condition we have on $\Lambda'\Lambda/N$ asymptotically identifies f_t, not just a rotation of the factor space. This is because the rotation matrix H_f boils down to I_r as $(N,T)_j \to \infty$, as we show in the appendix.

3 Consistent Factor Memory Estimation

Estimation of a memory parameter has traditionally been carried out directly on observable series. Although the common factor structure is unobservable in (4), the estimates $\hat{f}_t$ are observable and in this section, we will use them to estimate the integration orders of common factors. Given that the factor estimates are consistent at a parametric rate, we show that the memory parameter can be estimated as though the factors were observable and subsequently justify its consistency uniformly in the allowed set of memory values given in Assumption 1.

Based on the previously obtained common factor estimates, we estimate δ_0 based on a conditional-sum-of-squares criterion using which, the estimate is only implicitly defined as

$$\hat{\delta} = \arg\min_{\delta} \frac{1}{T}\left(\Delta_t^{\delta-1}\hat{f}_t\right)'\left(\Delta_t^{\delta-1}\hat{f}_t\right). \tag{6}$$

We then establish the $\sqrt{T}-$consistency of the memory parameters in the next theorem.

Theorem 2. *Under the conditions of Theorem 1, as $(N,T)_j \to \infty$,*

$$\hat{\delta} - \delta_0 = O_p\left(T^{-1/2}\right).$$

As we show in the appendix, this result is established replacing the common factor estimates by unobserved true common factors, given the result in Theorem 1, as though they were observable and additionally treating the estimation errors that are negligible imposing $\delta_0 - \underline{\delta} < 1/2$ and as $N \to \infty$. Note, however, that there is no restriction on the relative growth rates of N and T.

4 Generalized Principal Components Estimation of Common Factors

In this section, we show how efficient estimation of the common factor structure can be carried out based on the previously obtained consistent estimates. Let us introduce the notation $\omega_{it}(\tau) = \Delta_t^{\tau-1}\omega_{it}$.

We consider the optimization problem

$$\begin{aligned} &min\ tr\left[\left(X(\hat{\delta}) - f(\hat{\delta})\Lambda'\right)'\left(X(\hat{\delta}) - f(\hat{\delta})\Lambda'\right)\right] \\ &\quad s.t.\ f(\hat{\delta})'f(\hat{\delta})/T = I_r \\ &\qquad \Lambda'\Lambda \textit{ is diagonal with distinct entries}, \end{aligned} \tag{7}$$

which allows for the direct (asymptotic) identification of common factors and thus their corresponding loadings as discussed in Section 2.

Define for a fixed t,

$$\Gamma_t(\delta_0) = \lim_{N\to\infty}\sum_{i,j=1}^{N} E\left(\lambda_i'\lambda_j e_{it}(\delta_0)e_{jt}(\delta_0)\right),$$

and for a fixed i,

$$\Phi_i(\delta_0) = \lim_{T\to\infty}\sum_{s,t=1}^{T} E\left(f_s'(\delta_0)f_t(\delta_0)e_{is}(\delta_0)e_{it}(\delta_0)\right).$$

We have the following result that establishes the asymptotic behaviour for the efficient common factor structure estimates.

Theorem 3. *Under the conditions of Theorem 2, the optimization problem in (7) and if $N = o(T^\zeta)$ with $1/2 < \zeta < 1$,*

- *for a fixed t,*

$$\sqrt{N}(\hat{f}_t(\hat{\delta}) - f_t(\delta_0)) \to_d \mathcal{N}\left(0, \Sigma_\Lambda^{-1}\Gamma_t(\delta_0)\Sigma_\Lambda^{-1}\right);$$

- *for a fixed i,*

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) \to_d \mathcal{N}\left(0, \Phi_i(\delta_0)\right).$$

Consistent estimates for covariance matrices can be envisaged along the lines of Bai and Ng (2006). The relative growth rate of N to T requires that T grow faster than N but also that N grow faster than $\sqrt{T}$, which is because as $NT^{-1} \to 0$ the common factor estimate $\hat{f}_t(\hat{\delta})$ is asymptotically normally distributed but asymptotic normality of $\hat{\lambda}_i$ further requires that $\sqrt{T}N^{-1} \to 0$ as $(N,T)_j \to \infty$.

5 Estimation of the Number of Factors

The analysis in previous sections has assumed the number of factors to be known although this generally is not the case in practice. In this section, we consider the estimation of the number of common factors based on the efficient common factor estimates obtained earlier.

In order to estimate the number of common factors, r, we consider

$$x_{it}(\hat{\delta}) = \lambda_i' f_t(\hat{\delta}) + e_{it}(\hat{\delta}), \tag{8}$$

to write down the NT-normalized sum of squared residuals

$$V(k, \hat{f}^k(\hat{\delta})) = \min_{\Lambda} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it}(\hat{\delta}) - \lambda_i^{k\prime} \hat{f}_t^k(\hat{\delta}) \right)^2 \tag{9}$$

for k factors, where $k < \min\{N,T\}$. We then consider penalty functions, $g(N,T)$, such that criteria of the form

$$PC(k) = V(k, \hat{f}^k(\hat{\delta})) + kg(N,T),$$

or equivalently the information criteria

$$IC(k) = ln(V(k, \hat{f}^k(\hat{\delta}))) + kg(N,T),$$

will consistently estimate r.

Considering the fact that $\hat{\delta}$ is a $\sqrt{T}$-consistent estimate of δ_0, using which common factors are asymptotically exactly whitened, it can be shown adopting

Bai and Ng (2002) that for $r \leq k_{max}$ and $\hat{k} = argmin_{0 \leq k \leq k_{max}} IC(k)$,

$$\lim_{N,T \to \infty} Prob[\hat{k} = r] = 1$$

if

$$g(N,T) \to 0 \quad \text{and} \quad C_{NT}^2 g(N,T) \to \infty$$

as $(N,T)_j \to \infty$ where $C_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$. We can then make use of the following widely used specifications for $g(N,T)$,

$$g_1(N,T) = \left(\frac{N+T}{NT}\right) ln\left(\frac{NT}{N+T}\right),$$
$$g_2(N,T) = \left(\frac{N+T}{NT}\right) lnC_{NT}^2,$$

to determine the number of factors.

Other approaches can also be adopted. Hallin and Liska (2007) propose information criteria based on a frequency-domain approach where the penalty function contains a lag-window estimate of the spectral density matrix. A recent study by Ahn and Horenstein (2013) show that the use of information criteria that requires that the maximum number of common factors that allowed in the analysis be known leads to performance distortions in finite samples. So they propose using ratios of adjacent eigenvalues of the covariance matrix of the data. Although their methodology is developed for static factor models, under the static representation we described earlier, the eigenvalue ratio test can also be used to determine the number of factors. We do not develop such theory in this paper and our main aim is to assess how (possibly neglected) long memory properties affect the number of common factors, which we study based on Monte Carlo simulations in Section 7.

6 Testing for Latent Components

In applied economic analysis, there is an interest in obtaining statistical inference on the estimates of latent structures. Although the model in (1) has a simplistic nature, many useful specification tests can be performed on it. For example, it is possible to test if λ_i or some components of it are zero using a $q \times r$ restriction matrix Ξ, with $q \leq r$, such that under the null $\Xi\lambda_i = \bar{\lambda}_i$. Then under the null hypothesis,

$$T\left(\Xi\hat{\lambda}_i - \bar{\lambda}_i\right)'\left(\Xi\hat{\Phi}_i(\hat{\delta})\Xi'\right)^{-1}\left(\Xi\hat{\lambda}_i - \bar{\lambda}_i\right) \to_d \chi^2_q,$$

which is essentially an efficient Wald test. This approach can be directly used for testing for the existence of individual fixed effects or cross-section dependence both in stationary and nonstationary panel data models since failing to reject the null of $\lambda_i \equiv 0$ would imply redundancy of latent heterogeneity terms in the model. On the other hand, when the null is rejected, constancy of the common factors could be tested using the same approach, which essentially corresponds to a test for individual fixed effects in panel data models. Finally, rejecting both of these nested hypotheses implies there is only dynamic cross-section dependence but no individual fixed effects in a panel data model.

To simply illustrate these ideas, let us assume there is only one common factor. Then, for instance, individual fixed effects or cross-section dependence can be tested using the null hypothesis,

$$H_0^{CSD} : \lambda_i = 0,$$

which implies that the latent common factor structure does not exist. The test statistic can then be constructed as

$$t = \frac{\hat{\lambda}_i}{s.e.(\hat{\lambda}_i)} \to_d \mathcal{N}(0,1)$$

under the conditions of Theorem 3.

When H_0^{CSD} cannot be rejected, (1) does not have fixed effects or cross-section dependence. However, when H_0^{CSD} is rejected, there are two possibilities: fixed effects only or both fixed effects and cross-section dependence. If H_0^{CSD} is rejected, fixed effects only hypothesis can be tested based on testing whether f_t is time invariant, for example fitting a time polynomial as

$$f_t = \sum_{k=0}^{m} a_k t^k + u_f, \qquad u_f \sim iid(0, \upsilon)$$

where a_k are real valued and possibly unknown, which can be estimated along the lines of Robinson (2012). Then, the constancy of the common factor can be tested by

$$H_0^{FE} : a_k = 0, \quad k = 1, \dots, m,$$

using an F test in which the test statistic is compared to the $F_{m,\infty}$ critical value.

If H_0^{CSD} is rejected and H_0^{FE} cannot be rejected, the model features only fixed effects. If both null hypotheses are rejected, then only cross-section dependence is incorporated into the model.

7 Simulation Study

In this section, we carry out Monte Carlo simulations to assess the finite-sample behaviour of the factor structure estimates and the impact of efficient factor estimation on the selection of the number of factors. The shocks are drawn as $\epsilon_{it} \sim N(0, \sigma_i^2(1 - \rho_i^2))$ to generate the idiosyncratic series

$$e_{it} = \frac{\Delta^{-\vartheta_{i0}} \epsilon_{it}}{1 - \rho_i L}$$

where $\rho_i \sim U[0.5, 0.9]$. Factor loadings, λ_i, are drawn from $U(-0.5, 1)$ not to restrict the sign, and serially correlated common factors are generated by

$$g_t = \frac{\Delta^{-\delta_0} z_t}{1 - \psi L}$$

where $z_t \sim iidN(0, (1 - \psi^2))$. We focus on different cross-section and time-series sizes, N and T; as well as different values of δ_0 and ϑ_{i0}. Simulations are based on 1,000 replications.

7.1 Factor Structure Estimation

To assess the behaviour of the estimates under different specifications, we initially assume that there is only one factor. We then take $\sigma_i = \sqrt{2}$ in the homoskedastic case and assume $\sigma_i \sim iidN(\sqrt{2}, 0.25)$ in the heteroskedastic case, while also using $\psi = 0, 0.5$ to study pure fractional dynamics and allow for serial correlation, respectively, in the factor. We consider the cross-section sizes and time-series lengths, respectively, of $N = 32, 64$ and $T = 64, 128$. For the factor memory parameter, we consider the values of $\delta_0 = 0.4, 0.8$, and the idiosyncratic error memory parameter takes the values of $\vartheta_i = 0.2, 0.3$. For the estimation of the memory parameter, we consider both (6) and the widely used exact local Whittle approach by Shimotsu (2010) to set contrasts.

We first consider the case in which short memory and autocorrelation are allowed in the idiosyncratic errors but there is no heteroskedasticity. Table 1 collects these results. According to the results in Table 1, increasing T for a fixed N reduces the bias magnitude in the factor estimates for all three cases while the same is generally true for factor loadings when N increases for a fixed T, as suggested by Theorems 1 and 2. Having normalized the variance of the common factor to unity, most efficient factor loadings estimates are obtained employing the $\hat{\delta}$-whitened factor, where $\hat{\delta}$ is obtained from (6). In practice, sometimes a semiparametric approach, e.g. the two-step exact local Whittle method by Shimotsu (2010), is used but as expected, $\hat{\delta}_{elw}$ leads to larger variance distortions in the factor loadings estimate rendering it less efficient.

Once heteroskedasticity is also allowed in the idiosyncratic errors, whose results are shown in Table 2, magnitudes of the bias in factor estimates is reduced. Bias magnitudes for factor loadings are slightly inflated as opposed to the homoskedastic case and the performance of the estimates is slightly exacerbated although $\hat{\lambda}_i(\hat{f}_t(\hat{\delta}))$ remains to be the most efficient estimate.

Table 3 shows the case in which idiosyncratic errors are homoskedastic and the common factor is not in its static representation with it being allowed to carry serial correlation. Bias in factor estimates is roughly twice as much for all N, T combinations although for factor loadings estimates, bias is not affected much, in particular for larger N and T. As for the performance, the allowance for serial correlation appears to lead to efficiency gains in the factor loadings estimates.

When heteroskedasticity is introduced in addition to factor serial correlation, in Table 4, bias in factor estimates is reduced. Bias in factor loadings estimates does not respond much to the allowance for idiosyncratic heteroskedasticity although compared to the results in Table 3, performance of the estimates is slightly worsened.

We finally show that the negligence of even mild (stationary) long memory in data leads to a large bias in the factor and loadings estimates, possibly rendering their consistent estimation impossible. For this study, we fix $\delta_0 = 0.4$ and $\vartheta_i = 0.2$ and run the simulations in the easiest case in that there is no serial correlation in the common factor and idiosyncratic errors are homoskedastic. Table 5 shows the bias contrast of factor-structure estimates obtained under the negligence of long memory in the common factor versus efficient estimates of them. According to these results, normalization of factor variance to unity does not work when

$T = 64$, and the factor estimate is highly biased when long memory is overlooked although efficient estimation circumvents these problems. Therefore, it would be wise not to overlook the possibility for mild persistence, even in cases in which a unit root may be rejected by a standard test.

Table 1: Bias and SE Profiles of Factor Structure Estimates ($\delta_0 = 0.8, \vartheta_{i0} = 0.3, \sigma_i = \sqrt{2}, \psi = 0$)

		Common Factor			Factor Loadings		
		$\hat{f}_t$	$\hat{f}_t(\hat{\delta})$	$\hat{f}_t(\hat{\delta}_{elw})$	$\hat{\lambda}_i(\hat{f}_t)$	$\hat{\lambda}_i(\hat{f}_t(\hat{\delta}))$	$\hat{\lambda}_i(\hat{f}_t(\hat{\delta}_{elw}))$
n=32							
	T=64	-0.0338	-0.0341	-0.0312	0.0011	-0.0003	0.0025
		(1.0000)	(1.0000)	(1.0000)	(0.3800)	(0.3757)	(0.4063)
	T=128	0.0078	0.0075	0.0068	0.0013	0.0008	-0.0035
		(1.0000)	(1.0000)	(1.0000)	(0.3631)	(0.3565)	(0.3669)
n=64							
	T=64	0.0814	0.0794	0.0856	-0.0001	0.0022	-0.0013
		(1.0000)	(1.0000)	(1.0000)	(0.3527)	(0.3470)	(0.3807)
	T=128	-0.0182	-0.0212	-0.0227	-0.0021	-0.0019	-0.0024
		(1.0000)	(1.0000)	(1.0000)	(0.3391)	(0.3319)	(0.3381)

Note. In the estimation, factor variance is normalized to 1.0000. Memory parameter estimate $\hat{\delta}_{elw}$ is obtained based on the two-step exact local Whittle procedure by Shimotsu (2010). Standard errors of the estimates are reported in parentheses.

Table 2: Bias and SE Profiles of Factor Structure Estimates ($\delta_0 = 0.8, \vartheta_{i0} = 0.3, \sigma_i \sim iidN(\sqrt{2}, 0.25), \psi = 0$)

		Common Factor			Factor Loadings		
		$\hat{f}_t$	$\hat{f}_t(\hat{\delta})$	$\hat{f}_t(\hat{\delta}_{elw})$	$\hat{\lambda}_i(\hat{f}_t)$	$\hat{\lambda}_i(\hat{f}_t(\hat{\delta}))$	$\hat{\lambda}_i(\hat{f}_t(\hat{\delta}_{elw}))$
n=32							
	T=64	-0.0007	0.0036	0.0100	0.0014	-0.0026	0.0034
		(1.0000)	(1.0000)	(0.9999)	(0.3982)	(0.3947)	(0.4540)
	T=128	-0.0086	-0.0050	-0.0117	-0.0039	0.0012	0.0012
		(1.0000)	(1.0000)	(1.0000)	(0.3754)	(0.3692)	(0.3918)
n=64							
	T=64	-0.0288	-0.0281	-0.0235	0.0029	-0.0007	0.0007
		(1.0000)	(1.0000)	(1.0000)	(0.3648)	(0.3623)	(0.4247)
	T=128	-0.0317	-0.0297	-0.0311	0.0018	-0.0008	0.0000
		(1.0000)	(1.0000)	(1.0000)	(0.3426)	(0.3363)	(0.3480)

Note. In the estimation, factor variance is normalized to 1.0000. Memory parameter estimate $\hat{\delta}_{elw}$ is obtained based on the two-step exact local Whittle procedure by Shimotsu (2010). Standard errors of the estimates are reported in parentheses.

Table 3: Bias and SE Profiles of Factor Structure Estimates ($\delta_0 = 0.8, \vartheta_{i0} = 0.3, \sigma_i = \sqrt{2}, \psi = 0.5$)

		Common Factor			Factor Loadings		
		$\hat{f}_t$	$\hat{f}_t(\hat{\delta})$	$\hat{f}_t(\hat{\delta}_{elw})$	$\hat{\lambda}_i(\hat{f}_t)$	$\hat{\lambda}_i(\hat{f}_t(\hat{\delta}))$	$\hat{\lambda}_i(\hat{f}_t(\hat{\delta}_{elw}))$
n=32							
	T=64	-0.0608	-0.0596	-0.0610	0.0028	0.0019	0.0011
		(1.0000)	(1.0000)	(1.0000)	(0.3583)	(0.3553)	(0.3816)
	T=128	0.0160	0.0157	0.0179	0.0028	-0.0028	-0.0011
		(1.0000)	(1.0000)	(1.0000)	(0.3387)	(0.3352)	(0.3556)
n=64							
	T=64	0.1377	0.1404	0.1424	-0.0013	-0.0008	0.0004
		(1.0000)	(1.0000)	(1.0000)	(0.3269)	(0.3225)	(0.3536)
	T=128	-0.0317	-0.0332	-0.0335	-0.0002	-0.0009	-0.0012
		(1.0000)	(1.0000)	(1.0000)	(0.3116)	(0.3070)	(0.3228)

Note. In the estimation, factor variance is normalized to 1.0000. Memory parameter estimate $\hat{\delta}_{elw}$ is obtained based on the two-step exact local Whittle procedure by Shimotsu (2010). Standard errors of the estimates are reported in parentheses.

Table 4: Bias and SE Profiles of Factor Structure Estimates ($\delta_0 = 0.8, \vartheta_{i0} = 0.3, \sigma_i \sim iidN(\sqrt{2}, 0.25), \psi = 0.5$)

		Common Factor			Factor Loadings		
		$\hat{f}_t$	$\hat{f}_t(\hat{\delta})$	$\hat{f}_t(\hat{\delta}_{elw})$	$\hat{\lambda}_i(\hat{f}_t)$	$\hat{\lambda}_i(\hat{f}_t(\hat{\delta}))$	$\hat{\lambda}_i(\hat{f}_t(\hat{\delta}_{elw}))$
n=32							
	T=64	-0.0033	-0.0013	0.0107	0.0025	0.0032	-0.0012
		(1.0000)	(1.0000)	(0.9999)	(0.3783)	(0.3757)	(0.7735)
	T=128	-0.0147	-0.0162	-0.0152	0.0002	-0.0005	0.0028
		(1.0000)	(1.0000)	(1.0000)	(0.3523)	(0.3483)	(0.3938)
n=64							
	T=64	-0.0490	-0.0510	-0.0413	-0.0020	-0.0001	0.0008
		(1.0000)	(1.0000)	(0.9999)	(0.3422)	(0.3405)	(0.4236)
	T=128	-0.0549	-0.0552	-0.0511	0.0019	0.0030	0.0001
		(1.0000)	(1.0000)	(1.0000)	(0.3166)	(0.3119)	(0.3403)

Note. In the estimation, factor variance is normalized to 1.0000. Memory parameter estimate $\hat{\delta}_{elw}$ is obtained based on the two-step exact local Whittle procedure by Shimotsu (2010). Standard errors of the estimates are reported in parentheses.

Table 5: Bias and SE Profiles under Neglected Factor Long Memory ($\delta_0 = 0.4, \vartheta_{i0} = 0.2, \sigma_i = \sqrt{2}, \psi = 0$)

		Common Factor		Factor Loadings	
		$\hat{F}_t$	$\hat{f}_t(\hat{\delta})$	$\hat{\lambda}_i(\hat{F}_t)$	$\hat{\lambda}_i(\hat{f}_t(\hat{\delta}))$
n=32					
	T=64	0.4365	-0.0060	0.0080	0.0004
		(0.8962)	(1.0000)	(1.0715)	(0.4128)
	T=128	0.1212	-0.0022	0.0028	-0.0028
		(0.9924)	(1.0000)	(0.9348)	(0.3874)
n=64					
	T=64	0.4092	0.0176	0.0029	0.0012
		(0.9192)	(1.0000)	(1.0291)	(0.3929)
	T=128	0.0804	-0.0024	-0.0018	-0.0024
		(0.9965)	(1.0000)	(0.8762)	(0.3656)

Note. In the estimation, the factor variance is normalized to 1.0000 at the usual T rate. Standard errors of the estimates are reported in parentheses.

7.2 Estimating the Number of Factors

In this section, we explore how the efficient estimation of the factor structure affects the number of common factors chosen. Without loss of generality, we fix the number of factors to $r = 3$ and consider the combinations of $\delta_0 = 0, 0.4, 0.8, 1$ and $\vartheta_{max} = 0, 0.2, 0.3, 1$. We take $(N, T) = (40, 100)$ and $(130, 648)$ with the latter representing the sample size in the empirical application presented in Section 8.

We begin by considering the cases in which persistence is neglected and first differencing is carried out following Bai and Ng (2004)'s suggestion. Table 6 presents these simulation outcomes.

According to these results, negligence of persistence in common factors does not affect the number of factors selected. However, ignoring even mild long memory in the error term leads to the overestimation of the number of factors. This finding contrasts with the results in Table 7 in that the number of factors is robustly estimated so long as $\delta_0 \geq \vartheta_{max}$.

Bai and Ng (2004) suggest taking first differences to estimate the common factor structure when there is suspicion of $I(1)$ behaviour in data and claim that overdifferencing would also be fine. This is true for consistent estimation of the factor structure. However, as the last two columns of Table 6 show, the number of factors is correctly selected only when errors are $I(1)$ and it is overestimated in all other cases. This is not the case in Table 7 which shows that the number of factors are correctly estimated in all cases but when $(\delta_0, \vartheta_{max}) = (0, 1)$ that renders common factor estimates inconsistent.

Tables 6 and 7 also show the probability of selecting the correct number of common factors in each cases. The results in Table 7 suggest that as long as $\delta_0 \geq \vartheta_{max}$, this probability is at least 0.936 with it being very close to one in most of the cases, which contrasts with the results in Table 6 in that this probability is close to one only when $\delta_0 = \vartheta_{max} = 1$ under first differencing.

Table 6: Number of Common Factors under Negligence of Long Memory and First Differencing ($N = 40, T = 100$ and $k_{max} = 10$)

			Neglected Memory		First Differencing	
r	δ_0	ϑ_{max}	$\hat{k}$	$P(\hat{k} = r)$	$\hat{k}$	$P(\hat{k} = r)$
3	0	0	2.8	0.708	5.1	0.188
3	0.4	0	2.9	0.824	4.9	0.282
3	0.4	0.2	3.7	0.933	4.5	0.512
3	0	0.2	3.6	0.944	4.7	0.393
3	0.8	0.3	4.8	0.306	4	0.758
3	1	1	9.9	0.000	2.8	0.996
3	1	0	3.1	0.911	4.6	0.434
3	0	1	9.9	0.000	3.1	0.883

Table 7: Estimation of the Number of Common Factors ($k_{max} = 10$)

			$N = 40, T = 100$		$N = 130, T = 648$	
r	δ_0	ϑ_{max}	$\hat{k}$	$P(\hat{k} = r)$	$\hat{k}$	$P(\hat{k} = r)$
3	0	0	2.8	0.997	3	0.982
3	0.4	0	3.2	0.973	2.8	0.986
3	0.4	0.2	2.9	0.993	2.9	0.944
3	0	0.2	3.3	0.973	2.9	0.952
3	0.8	0.3	3.4	0.936	2.8	0.987
3	1	1	2.9	0.990	3	0.998
3	0	1	9.9	0.000	9.9	0.000

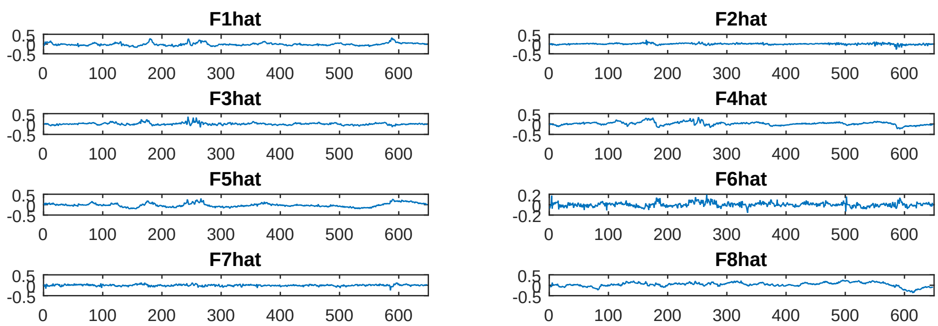
8 Diffusion Indices for US Economy

We analyze the US macroeconomic activity based on 130 monthly indicators for the time span of January 1960 - December 2014 to create diffusion indices, which may be later used for forecasting purposes. We downloaded the publicly available data from Federal Reserve Economic Database (FRED) and removed the missing values.

We note that earlier studies employing this dataset or that of Stock and Watson (2005) all require a variable-by-variable treatment in that a priori differencing to some of the data is necessary to make them stationary so that common factors can be extracted. Using our methodology, on the other hand, all variables can be treated at once because their persistence characteristics are altogether captured at the memory estimation step.

Following our estimation steps, we find that there are 9 common factors selected based on the criterion $g_1(N, T)$ and 8 common factors based on the criterion $g_2(N, T)$ given in Section 5. The latter finding parallels the results obtained by Jurado et al. (2015) and McCracken and Ng (2015). We plot the estimated original common factors, after integrating them back to their original integration orders, in Figure 1.

Figure 1: Estimated Diffusion Indices for the US Economy, 1960:1-2014:12



Based on these common factor estimates, the main interest is in assessing what fraction of the total variation each one of them explains. To compare our findings to those obtained by McCracken and Ng (2015), we regress the i-th series in the dataset on the set of r estimated common factors. For $k = 2, \ldots, r$ we obtain $R_i^2(k)$ for each i. We then calculate $mR^2(k) = N^{-1} \sum_{i=1}^{N} mR_i^2(k)$ where $mR_i^2(k) = R_i^2(k) - R_i^2(k-1)$, $k = 2, \ldots, r$ with $mR_i^2(1) = R_i^2(1)$. Our regressions also indicate on which indicators the factors heavily load. These findings are collectively presented in Table 8 along with integration orders of the diffusion indices that are estimated based on the parametric CSS criterion used in the basic setup by Ergemen and Velasco (2015).

Table 8: Average Importance of Each Factor and Explained Indicators

	mR^2	Related Key Indicators	CSS Memory Estimates
$\hat{F}_1$	0.1575	industrial production, employment	0.7169
$\hat{F}_2$	0.0655	term interest rate spreads, inventories	0.2446
$\hat{F}_3$	0.0535	inflation	0.5292
$\hat{F}_4$	0.0598	housing and interest rate	0.7732
$\hat{F}_5$	0.0551	housing and interest rate	0.6782
$\hat{F}_6$	0.0304	real activity, employment	0.3008
$\hat{F}_7$	0.0239	stock market	0.2796
$\hat{F}_8$	0.0316	exchange rates	0.7133
Total	0.4773		(0.0301)

According to these results, the first common factor explains 0.1575 of the variation in the data and loads heavily on industrial production and employment variables so it can be seen as an real activity factor, just like the sixth factor that contributes only by 0.0304 in explaining the total variation. The first factor exhibits high persistence with the estimated memory parameter around 0.72 while for the sixth factor, this estimate is 0.3 indicating that these two factors jointly capture the nonstationary and stationary dynamics, respectively, in the real activity. The second factor explains 0.0655 of the total variation in the data and is related to forward-looking variables as it loads heavily on term interest rate spreads and inventories. Its memory estimate is around 0.24 indicating stationary behavior. The third factor captures price variables so it can be seen as an inflation factor, which explains 0.0535 of the variation, and shows mildly nonstationarity with a memory estimate of 0.53. The fourth and fifth common factors are related to housing and interest rates explaining 0.0598 and 0.0551 of the total variation, respectively, exhibiting high persistence with their respective estimated integration orders 0.77 and 0.68. The seventh factor is related to stock market variables and covers 0.0239 whereas the eighth factor explains 0.0316 of the variation and is related to exchange rates. While the seventh factor exhibits stationarity of order 0.28, the eighth factor is nonstationary with an estimated memory value of 0.71.

The findings on where the estimated diffusion indices load heavily are in line with the literature and are similar to the results obtained by Jurado et al. (2015) and McCracken and Ng (2015). However, using our approach, we can also estimate the integration orders of these indices, which is important in understanding the dynamics in a more complete and accurate way, especially if a forecasting or factor-augmented regression study is to be undertaken based on them. In contrast with the literature, our findings show that the average importance of factors is not decreasing in the number of factors used because each factor itself may have different persistence characteristics and this leads to its capturing such dynamics in the data as well. Last but not least, there is no need to treat each indicator separately using our approach because different series do not require different a priori transformations. So our approach is easier to use than those readily available in the literature additionally providing a more accurate treatment.

9 Final Remarks

In this paper, we have studied a dynamic factor model that incorporates possibly fractional long-range dependence also allowing for short-memory dynamics. We have shown the consistency of first-differenced common factor estimates, and later used these initial estimates to obtain $\sqrt{T}$-consistent memory estimate of the original factors. This then allowed us to exactly whiten the factors and obtain efficient estimates of them at the final step, whose asymptotic behavior we have justified.

We have also shown the use of efficient factor structure estimates in selecting the number of factors and justified through Monte Carlo simulations that neglected memory in common factors leads to inaccurate selection of the number of factors. Furthermore, we have discussed how efficient factor structure estimates could be used in designing efficient Wald tests to test for latent components in stationary and nonstationary panel data models.

Finally, we have presented an empirical application to US diffusion indices and found that those indices actually exhibit fractional long-range dependence, which has not been shown in the literature before. We further discussed that our approach is more parsimonious than the available ones in the literature providing a more accurate treatment.

Appendix

A Proof of Theorem 1

We begin by observing that first differencing leads the series x_{it} to become asymptotically stationary under Assumption 1 because $\Delta x_{it} \sim I(\delta_0 - 1)$, with $\delta_0 \leq 1$ and $\delta_0 \geq \vartheta_{max}$ so that

$$\frac{1}{N}\frac{1}{T-1}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta x_{it})^2 \rightarrow_p \sigma_x^2 > 0$$

under Assumptions 2-4.

Therefore showing the result that

$$\min\left\{\sqrt{N}, T\right\}(\hat{f}_t - f_t H_f) = O_p(1), \ \textit{for a fixed } t,$$

can be shown proceeding the same way as in the proof of Lemma 2 by Bai and Ng (2004) under the model assumptions.

Next, we want to establish, adopting Bai and Ng (2013), that

$$H_f = I_r + O_p(\varsigma_{NT}^{-2}) \tag{10}$$

with $\varsigma_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$ if $\Lambda'\Lambda$ is diagonal with distinct entries. With

$$H_f' = \hat{V}^{-1}(\hat{f}'f/(T-1))(\Lambda'\Lambda/N),$$

we first check that

$$\begin{aligned}\frac{\hat{f}'f}{T-1} &= \frac{(\hat{f} - fH_f)'f}{T-1} + \frac{H_f'f'f}{T-1} \\ &= \frac{H_f'f'f}{T-1} + O_p(\varsigma_{NT}^{-2}) \end{aligned} \tag{11}$$

since $(\hat{f} - fH_f)'f/(T-1) = O_p(\varsigma_{NT}^{-2})$ by Lemma B.2 of Bai (2003). Right-multiplying both sides of (11) by H_f gives

$$\frac{\hat{f}'fH_f}{T-1} = \frac{H_f'f'fH_f}{T-1} + O_p(\varsigma_{NT}^{-2}). \tag{12}$$

Then,

$$\frac{\hat{f}' f H_f}{T-1} = \frac{\hat{f}'(fH_f - \hat{f} + \hat{f})}{T-1} = O_p(\varsigma_{NT}^{-2}) + I_r \tag{13}$$

because $\hat{f}'(fH_f - \hat{f})/(T-1) = O_p(\varsigma_{NT}^{-2})$ as above and $\hat{f}'\hat{f}/(T-1) = I_r$ under the identifying restriction.

Equating 12 and 13,

$$\begin{aligned} I_r &= H_f' \overbrace{\frac{f'f}{T-1}}^{=I_r} H_f + O_p(\varsigma_{NT}^{-2}) \\ &= H_f' H_f + O_p(\varsigma_{NT}^{-2}), \end{aligned}$$

so asymptotically H_f is an orthogonal matrix with eigenvalues equal to 1 or -1.

Next, we show that H_f is diagonal. From (11) and using $f'f/(T-1) = I_r$, we have that

$$\begin{aligned} H_f' &= \hat{V}^{-1}(\hat{f}'f/(T-1))(\Lambda'\Lambda/N), \\ &= \hat{V}^{-1}H_f'(\Lambda'\Lambda/N) + O_p(\varsigma_{NT}^{-2}). \end{aligned} \tag{14}$$

Multiply (14) on both sides by $\hat{V}$ and transpose to get

$$(\Lambda'\Lambda/N)H_f = H_f\hat{V} + O_p(\varsigma_{NT}^{-2}), \tag{15}$$

which shows that asymptotically H_f is a matrix containing the eigenvectors of $(\Lambda'\Lambda/N)$ that is diagonal with distinct eigenvalues by assumption. So then, each eigenvalue is associated with a unique unitary eigenvector, and this establishes that H_f is asymptotically diagonal. Without loss of generality, we can assume the eigenvalues of H_f are 1's and in that case, (10) is shown. Furthermore, from (15),

$$(\Lambda'\Lambda/N) = \hat{V} + O_p(\varsigma_{NT}^{-2}).$$

□

B Proof of Theorem 2

The CSS criterion in (6) can be written as

$$L_{N,T}(\delta)=\frac{1}{T}\left(\Delta_t^{\delta-1}\left(\hat{f}_t-f_tH_f\right)+\Delta_t^{\delta-1}(f_tH_f)\right)'\left(\Delta_t^{\delta-1}\left(\hat{f}_t-f_tH_f\right)+\Delta_t^{\delta-1}(f_tH_f)\right). \tag{16}$$

We argue that the squared estimation-error term in (16),

$$\frac{1}{T}\left(\Delta_t^{\delta-1}\left(\hat{f}_t-f_tH_f\right)\right)'\left(\Delta_t^{\delta-1}\left(\hat{f}_t-f_tH_f\right)\right)$$

is negligible as $(N,T)_j\to\infty$ because under Assumptions 5 and 6, Assumption 4 is necessary and sufficient for the conditions imposed in Lemma A.1 of Bai and Ng (2004). So the results in Lemma A.2 of Bai and Ng (2004) hold under Assumption 1 that imposes $\vartheta_{max}-\underline{\delta}<1/2$ and $\delta_0-\underline{\delta}<1/2$, see similar arguments used in the proofs of Theorems 4 and 5 of Ergemen and Velasco (2015). Therefore,

$$\max_{1\le k\le T}\frac{1}{T}\left\|\sum_{t=1}^{k}\left(\hat{f}_t-f_tH_f\right)\right\|=O_p\left((NT)^{-1/2}+T^{\delta_0-\underline{\delta}-1/2}\right).$$

Then the $\sqrt{T}-$consistency of the memory estimate can be established from

$$\frac{1}{T}\left(\Delta_t^{\delta-1}(\Delta_t^{1-\delta_0}z_t)\right)'\left(\Delta_t^{\delta-1}(\Delta_t^{1-\delta_0}z_t)\right)=\frac{1}{T}\left(\Delta_t^{\delta-\delta_0}z_t\right)'\left(\Delta_t^{\delta-\delta_0}z_t\right)$$

following exactly the same steps followed by Hualde and Robinson (2011).

Finally the cross-term in (16) is bounded and small by Cauchy-Schwarz inequality, and the proof is then complete. $\square$

C Proof of Theorem 3

First, let us write

$$\begin{aligned}\hat{f}_t(\hat{\delta})-f_t(\delta_0)&=\hat{f}_t(\hat{\delta})-\hat{f}_t(\delta_0)+\hat{f}_t(\delta_0)-f_t(\delta_0)\\&=\hat{f}_t(\hat{\delta})-\hat{f}_t(\delta_0)+(\hat{f}_t-f_t)(\delta_0)\\&=\hat{f}_t(\hat{\delta})-\hat{f}_t(\delta_0)+O_p\left(\frac{1}{\sqrt{N}}+\frac{1}{T}\right)\end{aligned}$$

by the result established in Theorem 1. Furthermore, using the Mean Value Theorem,

$$\hat{f}_t(\hat{\delta}) - \hat{f}_t(\delta_0) = \dot{\hat{f}}_t(\delta^{\dagger})(\hat{\delta} - \delta_0)$$
$$= o_p\left(T^{-1/2}\right),$$

arguing as Robinson and Hidalgo (1997) with previously established $\hat{\delta} - \delta_0 = O_p(T^{-1/2})$ in Theorem 2 that is stronger than needed but simplifies the proof. Then,

$$\hat{f}_t(\hat{\delta}) - f_t(\delta_0) = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{T}\right) + o_p(T^{-1/2})$$

and

$$\sqrt{N}\left(\hat{f}_t(\hat{\delta}) - f_t(\delta_0)\right) = O_p(1) + O_p\left(\frac{\sqrt{N}}{T}\right) + o_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right).$$

Thus the asymptotic normality of the common factor estimates requires that $NT^{-1} \to 0$ as $(N,T)_j \to \infty$ while for consistency of the estimate there is no rate requirement.

Asymptotic normality can then be established writing

$$\sqrt{N}\left(\hat{f}_t(\delta_0) - f_t(\delta_0)\right) = \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it}(\delta_0) + O_p\left(\frac{\sqrt{N}}{\varsigma_{NT}^2}\right)$$

if $\sqrt{N}/T \to 0$ with $\varsigma_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$, leading to

$$\sqrt{N}\left(\hat{f}_t(\delta_0) - f_t(\delta_0)\right) \to_d \mathcal{N}\left(0, \Sigma_\Lambda^{-1}\Gamma_t(\delta_0)\Sigma_\Lambda^{-1}\right)$$

for a fixed t.

To show the results for $\hat{\lambda}_i$, let us write

$$\sqrt{T}\left(\hat{\lambda}_i - \lambda_i\right) = \left(\frac{f(\hat{\delta})'f(\hat{\delta})}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(\hat{\delta}) e_{it}(\hat{\delta}) + O_p\left(\frac{\sqrt{T}}{\varsigma_{NT}^2}\right)$$

which establishes the asymptotic normal distribution using $\hat{\delta} - \delta_0 = O_p(T^{-1/2})$ and the identifying restriction $f(\hat{\delta})'f(\hat{\delta})/T = I_r$ if $\sqrt{T}/N \to 0$ since $\varsigma_{NT} =$

$\min\left\{\sqrt{N}, \sqrt{T}\right\}$. In particular,

$$\sqrt{T}\left(\hat{\lambda}_i - \lambda_i\right) \to_d \mathcal{N}\left(0, \Phi_i\right)$$

for a fixed i. $\square$

References

Ahn, S. C. and A. R. Horenstein (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

——— (2004): "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72(4), 1127–1177.

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-Augmented Regressions," *Econometrica*, 74, 1133–1150.

——— (2013): "Principal Components Estimation and Identification of Static Factors," *Journal of Econometrics*, 176, 18–29.

Breitung, J. and J. Tenhofen (2011): "GLS Estimation of Dynamic Factor Models," *Journal of the American Statistical Society*, 106(495), 1150–1166.

Choi, I. (2011): "Efficient Estimation of Nonstationary Factor Models," *Working Paper*.

——— (2012): "Efficient Estimation of Factor Models," *Econometric Theory*, 28(2), 274–308.

Ergemen, Y. E. and C. Velasco (2015): "Estimation of Fractionally Integrated Panels with Fixed-Effects and Cross-Section Dependence," *CREATES Research Paper 2015-35*.

Gao, J. and P. Robinson (2014): "Inference on Nonstationary Time Series with Moving Mean," *Econometric Theory*, http://dx.doi.org/10.1017/S0266466614000875, 1–27.

Hallin, M. and R. Liska (2007): "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102, 603–617.

Hualde, J. and P. M. Robinson (2011): "Gaussian Pseudo-Maximum Likelihood Estimation of Fractional Time Series Models," *The Annals of Statistics*, 39(6), 3152–3181.

Jurado, K., S. Ludvigson, and S. Ng (2015): "Measuring Macroeconomic Uncertainty," *American Economic Review*, 105, 1177–1216.

Luciani, M. and D. Veredas (2012): "A Model for Vast Panels of Volatilities," *Banco de España Working Paper Series.*

McCracken, M. and S. Ng (2015): "FRED-MD: A Monthly Database for Macroeconomic Research," *Federal Reserve Bank of St. Louis, Working Paper 2015-012B.*

Robinson, P. M. (2012): "Inference on Power Law Spatial Trends," *Bernoulli*, 18(2), 644677.

Robinson, P. M. and J. Hidalgo (1997): "Time Series Regression with Long-Range Dependence," *The Annals of Statistics*, 25(1), 77–104.

Robinson, P. M. and C. Velasco (2015): "Efficient Inference on Fractionally Integrated Panel Data Models with Fixed Effects," *Journal of Econometrics*, 185, 435–452.

Shimotsu, K. (2010): "Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend," *Econometric Theory*, 26(2), 501–540.

Stock, J. H. and M. W. Watson (2002): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

Research Papers 2016

CREATES
Center for Research in Econometric Analysis of Time Series

2015-48: Harri Pönkä: The Role of Credit in Predicting US Recessions

2015-49: Palle Sørensen: Credit policies before and during the financial crisis

2015-50: Shin Kanaya: Uniform Convergence Rates of Kernel-Based Nonparametric Estimators for Continuous Time Diffusion Processes: A Damping Function Approach

2015-51: Tommaso Proietti: Exponential Smoothing, Long Memory and Volatility Prediction

2015-52: Mark Podolskij, Christian Schmidt and Mathias Vetter: On U- and V-statistics for discontinuous Itô semimartingale

2015-53: Mark Podolskij and Nopporn Thamrongrat: A weak limit theorem for numerical approximation of Brownian semi-stationary processes

2015-54: Peter Christoffersen, Mathieu Fournier, Kris Jacobs and Mehdi Karoui: Option-Based Estimation of the Price of Co-Skewness and Co-Kurtosis Risk

2015-55 Kadir G. Babaglou, Peter Christoffersen, Steven L. Heston and Kris Jacobs: Option Valuation with Volatility Components, Fat Tails, and Nonlinear Pricing Kernels

2015-56: Andreas Basse-O'Connor, Raphaël Lachièze-Rey and Mark Podolskij: Limit theorems for stationary increments Lévy driven moving averages

2015-57: Andreas Basse-O'Connor and Mark Podolskij: On critical cases in limit theory for stationary increments Lévy driven moving averages

2015-58: Yunus Emre Ergemen, Niels Haldrup and Carlos Vladimir Rodríguez-Caballero: Common long-range dependence in a panel of hourly Nord Pool electricity prices and loads

2015-59: Niels Haldrup and J. Eduardo Vera-Valdés: Long Memory, Fractional Integration, and Cross-Sectional Aggregation

2015-60: Mark Podolskij, Bezirgen Veliyev and Nakahiro Yoshida: Edgeworth expansion for the pre-averaging estimator

2016-01: Matei Demetrescum, Christoph Hanck and Robinson Kruse: Fixed-b Inference in the Presence of Time-Varying Volatility

2016-02: Yunus Emre Ergemen: System Estimation of Panel Data Models under Long-Range Dependence

2016-03: Bent Jesper Christensen and Rasmus T. Varneskov: Dynamic Global Currency Hedging

2016-04: Markku Lanne and Jani Luoto: Data-Driven Inference on Sign Restrictions in Bayesian Structural Vector Autoregression

2016-05: Yunus Emre Ergemen: Generalized Efficient Inference on Factor Models with Long-Range Dependence