DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Parametric Estimation of Long Memory in Factor Models

Yunus Emre Ergemen

CREATES Research Paper 2022-10

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

Email: oekonomi@au.dk
Tel: +45 9352 1936

Parametric Estimation of Long Memory in Factor Models

Yunus Emre Ergemen*

June 9, 2022

Abstract

A dynamic factor model is proposed in that factor dynamics are driven by stochastic time trends describing arbitrary persistence levels. The proposed model is essentially a long memory factor model, which nests standard $I(0)$ and $I(1)$ behavior smoothly in common factors. In the estimation, principal components analysis (PCA) and conditional sum of squares (CSS) estimations are employed. For the dynamic model parameters, centered normal asymptotics are established at the usual parametric rates, and their small-sample properties are explored via Monte-Carlo experiments. The method is then applied to a panel of U.S. industry realized volatilities.

Keywords: Factor models; long memory; conditional sum of squares; principal components analysis; realized volatility.
JEL Codes: C12, C13, C33

*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

Given the increasing availability of large data sets, factor models are extensively used as a dimension reduction tool in several applications. For example, Kapetanios (2004) and Cristadoro et al. (2005) build economic indicators while Bernanke et al. (2005) and Favero et al. (2005) derive policy results employing factor models. There has also been a keen focus on using factors for forecasting, see, for recent reviews, Uematsu and Yamagata (2022), Baltagi et al. (2021), Fan et al. (2021) and Karabiyik and Westerlund (2021). Furthermore, factor models have been employed in finance to analyze stochastic volatility, Cipollini and Kapetanios (2008), market liquidity, Hallin et al. (2011), and market volatility, Barigozzi and Hallin (2016), and in energy economics to analyze electricity prices, Dordonnat et al. (2012), and Ergemen et al. (2016). On the other hand, the theoretical development of static and dynamic factor models has taken place in stationary $I(0)$ and nonstationary $I(1)$ setups (see e.g. Forni et al. (2000), Stock and Watson (2002a), Bai and Ng (2002) Bai (2003), Bai and Ng (2004), Forni et al. (2004), Forni et al. (2005), Bai and Ng (2008), Choi (2012) and Barigozzi et al. (2016, 2021) for a review) as well as under long-memory setups (see e.g. Ray and Tsay (2000), Chen and Hurvich (2006), Morana (2007), Luciani and Veredas (2015), Hartl (2020) and Cheung (2021)).

In this paper, we propose a dynamic factor model with arbitrary persistence in common factors, smoothly nesting the typical $I(0)$ and $I(1)$ cases as well as cases in which the integration orders are fractional thereby covering also long memory factor models. This strand of the literature allows for persistence in observed series to be generated by common latent sources, coupled with the empirical evidence that several economic and financial indicators exhibit long-range dependence of fractional orders; see, among others, Gil-Alaña and Robinson (1997), Michelacci and Zaffaroni (2000), and Bollerslev et al. (2013) as well as the theoretical evidence that aggregation can lead to long memory; see Robinson (1978), Granger (1980), Chambers (1998) and Pesaran and Chudik (2014).

In the setup, we allow for varying degrees of persistence in the analyzed series through common factors using fractionally integrated time series processes in which persistence levels are effectively unrestricted cf. Hualde and Robinson (2011) and thus the entire spectrum covering stationary, nonstationary, invertible and noninvertible cases can be studied, unlike the traditional static and dynamic factor models. Our approach also sets contrast to the existing long memory models in that our model has a greater deal of flexibility in the allowed range of memory values for factors; memory estimates can be obtained at the standard $\sqrt{T}$ rate; and there is no dependence on a user-chosen bandwidth parameter, unlike the models employing semiparametric memory estimation, see Hartl and Weigand (2019), Hartl (2020) and Cheung (2021), thus circumventing the problem of bandwidth selection to which memory estimates are sensitive.

In the methodology, we first estimate the factor loadings with first-differenced data, and then project the data in levels onto the space spanned by the estimated loadings to obtain the common factor estimates, which is essentially a least-squares procedure. The loadings and factor estimates are consistent at the rate $O_p\left(\max\left(N^{-1/2}, T^{-1/2}\right)\right)$. Based on these consistent factor estimates, we obtain the estimates of factor integration orders employing a CSS criterion. The dynamic parameter estimates are $\sqrt{T}$ consistent and, when $TN^{-2} \to 0$ as $(N, T) \to \infty$, have a centered asymptotic normal distribution, where N is the cross-section size and T is the time series length. For the asymptotic results, since we require that both N and T grow, we depart from short panel (fixed T) and multivariate time series (fixed N) setups as considered by Robinson and Velasco (2015, 2019). Further, we discuss the issues of estimation with nonstationary errors of general form, i.e. beyond the unit-root case we consider in this paper, and estimation of the number of factors. We explore the small-sample behavior of our estimates by means of Monte Carlo experiments and show that the estimates behave well even in small panels.

In the empirical application, we analyze the sources of common persistence in a panel of U.S. industry realized volatilities (RV's hereafter). We find cointegrating relationships between the estimated factor and industry RV's, and show how the estimated factor compares to other potential indices in terms of both explanatory and predictive power. We show that the estimated factor and a proxy obtained by cross-sectionally averaging the data perform equally well. We also present an out-of-sample study which shows that our factor estimate and models augmented with it significantly improve the forecasting performance over typical $ARFIMA(1, d, 0)$ alternatives.

The remainder of the paper proceeds as follows. Next section presents the model and the conditions imposed to study it. Section 3 details the estimation procedures for common factors and dynamic parameters, and contains the main results. Section 4 presents a finite-sample study based on Monte Carlo experiments. Section 5 presents the empirical application, and Section 6 concludes.

Throughout the paper, "$(N, T) \to \infty$" denotes asymptotics in which both N and T are growing; " $\to_p$ " denotes convergence in probability; " $\to_d$ " denotes convergence in distribution; $M < \infty$ denotes a generic positive number independent of T and N, and $\|A\| = trace(AA')^{1/2}$. All mathematical proofs are collected in the appendix.

2 The Model

For an observable array $\{x_{it}, i \geq 1, t \geq 1\}$, we consider

$$x_{it} = \gamma_i' f_t + e_{it} \tag{1}$$

where the common component, $\gamma_i' f_t$, consists of the $r \times 1$, with r fixed, vector of common factors, f_t, and the corresponding vector of factor loadings, γ_i, while the scalar e_{it} constitutes the idiosyncratic component. Adopting the model considered by Hualde and Robinson (2011), we assume that the dynamics in the vector of common factors are governed by

$$f_t = \lambda_t^{-1}(L; \theta_0)\varepsilon_t, \quad t = 1, \ldots, T, \tag{2}$$

so that the k-th component of f_t satisfies

$$f_{kt} = \lambda_t^{-1}(L; \theta_{k0})\varepsilon_{kt} \, , \tag{3}$$

for $k = 1, \ldots, r$, where ε_{kt} are factor-specific shocks, $\theta_{k0} \in \Theta_k \subset \mathbb{R}^{p+1}$ is a $(p+1) \times 1$ parameter vector, L is the lag operator, and

$$\lambda_t \left(L; \theta \right) = \sum_{j=0}^{t} \lambda_j \left(\theta \right) L^j \tag{4}$$

for $\theta \in \Theta$ and each $t > 0$ truncates $\lambda \left(L; \theta \right) = \lambda_\infty \left(L; \theta \right)$. We assume that $\lambda \left(L; \theta \right)$ takes the form

$$\lambda \left(L; \theta \right) = \Delta^\delta \psi \left(L; \xi \right) \, ,$$

where $\delta \geq 0$ is a scalar, ξ is a $p \times 1$ vector, and $\theta = (\delta, \xi')'$. Here, $\Delta = 1 - L$ is the first-difference operator, so that the fractional filter Δ^δ has the expansion

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

and we denote the truncated version as $\Delta_t^{\delta} = \sum_{j=0}^{t-1} \pi_j(\delta) L^j$, with $\Gamma(-\delta) = (-1)^{\delta} \infty$ for $\delta = 0, 1, \ldots,$ $\Gamma(0)/\Gamma(0) = 1$. Further, $\psi(L;\xi)$ is a known function, such that for complex-valued z, $|\psi(z;\xi)| \neq 0$, for $|z| \leq 1$, and in the expansion

$$\psi(L;\xi) = \sum_{j=0}^{\infty} \psi_j(\xi) L^j ,$$

the coefficients $\psi_j(\xi)$ satisfy

$$\psi_0(\xi) = 1 , \quad |\psi_j(\xi)| + \left\| \dot{\psi}_j(\xi) \right\| = O\left(\exp\left(-c(\xi) j\right)\right) , \tag{5}$$

for $\xi \in \Xi \subset \mathbb{R}^p$, where $\dot{\psi}_j(\xi) = (d/d\xi')\psi_j(\xi)$, and $c(\xi)$ is a positive-valued function of ξ. Note that

$$\lambda_j(\theta) = \sum_{k=0}^{j} \pi_{j-k}(\delta) \psi_k(\xi) , \quad j \geq 0 , \tag{6}$$

behaves asymptotically as $\pi_j(\delta)$,

$$\lambda_j(\theta) = \psi(1;\xi) \pi_j(\delta) + O\left(j^{-\delta-2}\right) , \quad \text{as } j \to \infty ,$$

where

$$\pi_j(\delta) = \frac{1}{\Gamma(-\delta)} j^{-\delta-1}(1 + O(j^{-1})) , \quad \text{as } \ j \to \infty ,$$

see Robinson and Velasco (2015). Thus, f_{kt} is integrated of order δ_{k0}, indicated as $f_{kt} \sim I(\delta_{k0})$, with the value of δ_{k0} determining whether it is asymptotically stationary $(\delta_{k0} < 1/2)$ or nonstationary $(\delta_{k0} \geq 1/2)$, and $\psi(L;\xi_{k0})$ describes the short memory dynamics of stable ARMA type. Important special cases included in the specification are stationary ARMA factors $(\delta_{k0} = 0)$, and factors of unit root type $(\delta_{k0} = 1)$. For example, if factors are truly AR(1), this will be picked up by the data, with the autocorrelation coefficient estimated as part of ξ_{k0}, jointly with the persistence δ_{k0}.

Temporarily suppressing the short-memory dynamics in (3) (so that $\xi_{k0} = 0$), we can write

$$f_{kt} = \sum_{s=0}^{t} \pi_s(-\delta_{k0}) \varepsilon_{kt-s}, \tag{7}$$

and when $\delta_{k0} = 1$,

$$f_{kt} = \sum_{s=0}^{t} \varepsilon_{kt-s}.$$

The latter results also by taking $\rho = 1$ in the autoregressive scheme popular in the dynamic factor model literature:

$$f_{kt} = \sum_{s=0}^{t} \rho^s \varepsilon_{kt-s}. \tag{8}$$

The typical alternatives to $\rho = 1$ covered by (8) are the stationary ones $\rho \in (-1, 1)$ or the explosive ones $\rho > 1$. These and other related versions of the dynamic (autoregressive) factor models are available in the literature, see e.g. Barigozzi et al. (2016). In contrast, note that in the fractional specification in (7), the weights $\pi_s(-\delta_{k0})$ have decay or growth that is, unlike in (8), not exponential but hyperbolic, since, for any δ,

$$\pi_s(\delta) = \frac{1}{\Gamma(-\delta)} s^{-\delta-1}(1 + O(s^{-1})) \quad \text{as} \quad s \to \infty,$$

see Robinson and Velasco (2015). Additionally, the fractional class described by $\pi_s(-\delta_{k0})$ has a smoothness at $\delta_{k0} = 1$ and elsewhere, thus the dynamic parameter estimates and the related test statistics have standard asymptotic distributions with the usual parametric rates, see also Robinson and Velasco (2015).

We impose the following conditions to study the model in (1).

Assumption A

A.1. The idiosyncratic errors e_{it} are governed by $(1 - \rho_i L)e_{it} = a_i(L)u_{it}$ for all i, where $a_i(L) = \sum_{k=0}^{\infty} a_{ik} L^k$ with $\sum_{k=0}^{\infty} k\,|a_{ik}| \leq M$, and $|\rho_i| \leq 1$. The vector of innovations, $u_t = (u_{1t}, \ldots, u_{Nt})'$, satisfy $u_t \sim iid\,(0, \Omega_u)$, $\Omega_u > 0$, with $E\,\|u_t\|^4 < \infty$, and $E\,(u_{it}u_{jt}) = \tau_{ij}$ with $\sum_{j=1}^{N} |\tau_{ij}| \leq M$ uniformly in i. The factor-specific idiosyncratic shocks vector, $\varepsilon_t \sim iid(0, \Omega_\varepsilon)$, $\Omega_\varepsilon > 0$, with $E\,\|\varepsilon_t\|^4 < \infty$, $\text{rank}\big(E\left[\Delta_t^{\delta_0} f_t \Delta_t^{\delta_0} f_t'\right]\big) = r$, with r fixed, and $E\left[\left(\Delta_t^{\delta_{k0}} f_{kt}\right)^2\right] > E\left[\left(\Delta_t^{\delta_{l0}} f_{lt}\right)^2\right] > 0$ for all $k, l = 1, \ldots, r$ with $k < l$. Also, for all i and $t \leq 0$, $u_{it} = 0$ and $\varepsilon_t = 0$.

A.2. The innovations u_{it} and ε_{ks}, and loadings γ_i are mutually independent groups, for $i = 1, \ldots, N$, $k = 1, \ldots, r$, and $s, t = 1, \ldots, T$.

A.3. Factor loadings γ_i are either nonrandom and satisfy $\|\gamma_i\| \leq M$, or random and satisfy $E\|\gamma_i\|^4 \leq M$, and as $N \to \infty$, $N^{-1}\sum_{i=1}^{N}\gamma_i\gamma_i' \to_p I_r$.

A.4. Let the parameter space for ξ_k be a compact subset Ξ_k of $\mathbb{R}^p$. For $\xi_k \in \Xi_k$, $\psi(z;\xi_k)$ is twice continuously differentiable in ξ_k. For all $\xi_k \neq \xi_{i0}$, $|\psi(z;\xi_k)| \neq |\psi(z;\xi_{k0})|$ on a subset of $\{z : |z| = 1\}$ of positive Lebesgue measure, and (5) holds, for all $\xi_k \in \Xi_k$, with $c_k(\xi_k)$ satisfying

$$\inf_{\Xi_k} c_k(\xi_k) = c_k^* > 0\,,$$

for $k = 1, \ldots, r$.

Assumption A.1 allows for $I(1)$ idiosyncratic components in (1) requiring only fourth-order moments as in Barigozzi et al. (2021), and unlike Bai and Ng (2004) and Cheung (2021) who require eighth-order moments. Further, the idiosyncratic errors are allowed to have weak cross correlation in the sense that even as T and N increase, the column sum of the error covariance matrix remains bounded, which makes (1) an approximate factor model in the sense of Chamberlain and Rothschild (1983). Also unlike Cheung (2021) and Ergemen and Velasco (2017), there is no restriction on the allowed range of memory values, cf. Hualde and Robinson (2011), for the common factors. The condition $E\left[\left(\Delta_t^{\delta_{k0}} f_{kt}\right)^2\right] > E\left[\left(\Delta_t^{\delta_{l0}} f_{lt}\right)^2\right] > 0$ for all $k, l = 1, \ldots, r$ with $k < l$, implies that $I(0)$ versions of all r factors contribute to the variance of x_{it} in a decreasing order of importance, and is a common identifying assumption in factor models, see Barigozzi et al. (2021) and Bai and Ng (2013) for analogous versions of this condition on the factors and loadings, respectively. The assumption on the initial conditions is common in nonstationary analysis, see also Barigozzi et al. (2021). Although, among others, Robinson and Velasco (2015, 2019) study the consequences of initial conditions under long memory setups, (1) is essentially a heterogeneous panel data model and Ergemen and Velasco (2017) show that the initial condition bias is asymptotically negligible for this class of models with N and T jointly growing. Assumption A.2 is a standard assumption in the common factor literature to allow for common and idiosyncratic components to be independent sources of variation. Assumption A.3 allows for cases of both random and nonrandom γ_i, and implies that the r factors are not redundant. The condition $N^{-1}\sum_{i=1}^{N}\gamma_i\gamma_i' \to_p I_r$ together with the identifying condition in Assumption A.1 are common identifying assumptions, see also Stock and Watson (2002b) and Barigozzi et al. (2021). Finally, Assumption A.4 ensures smoothness of the lag polynomials with the given parameters and the weights lead to short-memory dynamics, as also assumed by Robinson and Velasco (2015). The parameter spaces can involve stationarity and invertibility restrictions on the lag polynomials.

3 Estimation

The main interest in this paper is in the estimation of persistence induced by the common factors while the dynamics in idiosyncratic components are treated as nuisance. For each $k = 1, \ldots, r$, the parameter vector θ_{k0} in (3) bestows long- and short-memory dynamics on x_{it} through the common factor structure since for $i = 1, \ldots, N$, and $t = 1, \ldots, T$,

$$x_{it} = \sum_{k=1}^{r} \gamma_{ik} \lambda_t^{-1}(L; \theta_{k0}) \varepsilon_{kt} + e_{it}. \tag{9}$$

However, the common factor structure in (9) is not observable and needs to be extracted so that the dynamic parameter vector θ_{k0} can be estimated. In the factor estimation literature, PCA is the usual choice. Since we allow e_{it} to be nonstationary under Assumption A.1, we follow the approach by Barigozzi et al. (2021) to ensure the consistency of PCA. After extracting the common factors, we estimate their stochastic dynamics parametrically based on (2) without restricting persistence levels. Like us, Hartl and Weigand (2019), Hartl (2020) and Cheung (2021) consider factor memory estimation but unlike ours, their methods are based on local Whittle approaches using which the convergence rate is slower and depends on the bandwidth choice. In contrast, our estimates enjoy a faster, $\sqrt{T}$, convergence rate and is free of bandwidth choice.

We describe the estimation steps in turn, and also discuss the estimation of the number of factors.

Estimating the factor structure. For PCA, we follow Barigozzi et al. (2021) in that we estimate the factor loadings with first-differenced data, and then project the data in levels onto the space spanned by the estimated loadings to obtain the common factor estimates.

We can write down the first-differenced N-dimensional version of (1) as

$$\Delta \mathbf{x}_t = \Gamma \Delta f_t + \Delta \mathbf{e}_t, \quad t = 2, \ldots, T,$$

where Γ is $N \times r$, and $\Delta \mathbf{e}_t$ is $N \times 1$. Defining the $N \times T_1$, with $T_1 = T - 1$, first-differenced data matrix $\Delta\mathbf{X} = (\Delta\mathbf{X}_2, \ldots, \Delta\mathbf{X}_T)$, the estimated loadings matrix, $\hat{\Gamma}$, is obtained by $\sqrt{N}$-times the first r normalized eigenvectors of the $N \times N$ sample covariance matrix $T_1^{-1}\Delta\mathbf{X}\Delta\mathbf{X}'$. This choice of the sample covariance matrix is motivated by the fact that $N < T$ in many macroeconomic and financial applications. Then, the common factors are estimated by projecting the data vector in levels, $\mathbf{x}_t$, onto the space spanned by the estimated loadings, at any given point in time $t = 1, \ldots, T$,

$$\hat{f}_t = N^{-1}\hat{\Gamma}'\mathbf{x}_t = \frac{1}{N}\sum_{i=1}^{N}\hat{\gamma}_i x_{it} \tag{10}$$

under the normalization $N^{-1}\hat{\Gamma}'\hat{\Gamma} = I_r$.

For any $r \times r$ invertible matrix H, it is possible to write the N-dimensional version of (1) as

$$\mathbf{x}_t = \left(\Gamma H^{-1}\right)\left(H' f_t\right) + \mathbf{e}_t = \Gamma^* f_t^* + \mathbf{e}_t, \quad t = 1, \ldots, T,$$

from which it is clear that $\Gamma f_t = \Gamma^* f_t^*$, but the factors and loadings are not separately identified. However, imposing the conditions $N^{-1}\sum_{i=1}^{N}\gamma_i\gamma_i' \to_p I_r$ in Assumption A.3 and $E\left[\left(\Delta_t^{\delta_{k0}} f_{kt}\right)^2\right] > E\left[\left(\Delta_t^{\delta_{l0}} f_{lt}\right)^2\right] > 0$ for all $k, l = 1, \ldots, r$ with $k < l$ in Assumption A.1, the factors and loadings are identified up to a sign. These conditions are common identifying assumptions in the literature, see e.g. Barigozzi et al. (2021), and the theory presented in what follows can be adapted to different identifying conditions, such as those proposed by Bai and Ng (2013) and Fan et al. (2013). For the purpose of estimating factor dynamics, these assumptions are innocuous.

Since the main interest of this paper is in estimating the factor dynamics, only the convergence rates at which the factors and loadings are consistent are required for the subsequent analysis. In the next lemma, we collect these consistency results for the factors and loadings.

Lemma 1. *Let J be an $r \times r$ diagonal matrix with entries ± 1 depending on N and T. Then, under Assumption A and as $(N, T) \to \infty$,*

(i) for all i, $\|\hat{\gamma}_i' - \gamma_i' J\| = O_p\left(\max\left(N^{-1/2}, T^{-1/2}\right)\right)$,

(ii) for fixed t, $T^{-1/2}\left\|\hat{f}_t - J f_t\right\| = O_p\left(\max\left(N^{-1/2}, T^{-1/2}\right)\right)$.

These results are readily established in Lemma 1 of Barigozzi et al. (2021). The result for the loadings in (i) is uniform in i, using which the convergence rate in (*ii*) is established cf. (10). For the purpose of this paper, the consistency result for the factors in (*ii*) is sufficient. Finally, the dependence on matrix J and the resulting sign indeterminacy do not pose any problem for the estimation of factor dynamics in what follows.

Estimating the factor dynamics. For a given parameter vector $\theta_k = (\delta_k, \xi_k')'$, we can write (3) based on the estimated factors as

$$\hat{f}_{kt} = \lambda_t^{-1}(L; \theta_k)\hat{\varepsilon}_{kt}, \tag{11}$$

for $k = 1, \ldots, r$, from which the factor-specific-error estimates can be represented as

$$\hat{\varepsilon}_{kt}(\theta_k) = \lambda_t(L; \theta_k)\hat{f}_{kt}. \tag{12}$$

Denoting by $\hat{\theta}_k$ the estimate of the unknown true parameter vector θ_{k0},

$$\hat{\theta}_k = \arg \min_{\theta_k \in \Theta_k} L_{k,T}(\theta_k), \tag{13}$$

where the feasible CSS criterion to be minimized, based on (12), is

$$L_{k,T}(\theta_k) = \frac{1}{T}\hat{\varepsilon}_k(\theta_k)\hat{\varepsilon}_k(\theta_k)', \tag{14}$$

where $\hat{\varepsilon}_k = (\hat{\varepsilon}_{k1}, \ldots, \hat{\varepsilon}_{kT})$, and the dependence on $\hat{f}_{kt}$ cf. (12) is suppressed. The estimator $\hat{\theta}_k$ in (13) is implicitly defined and entails optimization over the compact set $\Theta_k = \mathcal{D}_k \times \Xi_k$, where $\mathcal{D}_k = [\underline{\delta}_k, \overline{\delta}_k]$ for suitable $\underline{\delta}_k$ and $\overline{\delta}_k$ such that $\underline{\delta}_k < \overline{\delta}_k$ and $\theta_{k0} \in \Theta_k$ for $k = 1, \ldots, r$. The set of admissible values of $\delta_{k0} \in \mathcal{D}_k$ is effectively unrestricted cf. Hualde and Robinson (2011), covering stationary, non-stationary, invertible, and non-invertible cases, without requiring knowledge on the whereabouts of δ_{k0}. This is particularly useful since f_{kt} are unobservable. The optimization in (14) can be conducted based on an appropriate numerical optimizer in practice. For the typical case in which the factors are assumed to have AR(1) short-memory dynamics, $\hat{\varepsilon}_{kt}(\theta)$ takes the simple form $\hat{\varepsilon}_{kt}(\theta) = \Delta_t^{\delta}\hat{f}_{kt} - \xi\Delta_t^{\delta}\hat{f}_{kt-1}$.

We define

$$\chi(L; \xi) = \frac{\partial}{\partial\theta}\log\lambda(L; \theta) = (\log\Delta, (\partial/\partial\xi')\log\psi(L; \xi))' = \sum_{j=1}^{\infty}\chi_j(\xi)L^j, \tag{15}$$

and introduce the $(p+1) \times (p+1)$ matrix

$$B(\xi) = \sum_{j=1}^{\infty} \chi_j(\xi) \chi_j'(\xi) = \begin{bmatrix} \pi^2/6 & -\sum_{j=1}^{\infty} \chi_{2j}'(\xi)/j \\ -\sum_{j=1}^{\infty} \chi_{2j}(\xi)/j & \sum_{j=1}^{\infty} \chi_{2j}(\xi) \chi_{2j}'(\xi) \end{bmatrix},$$

assuming that $B(\xi_0)$ is non-singular. The following result establishes the asymptotic behavior of the CSS estimates.

Theorem 1. *Under Assumption A, $\theta_{k0} \in \Theta_k$, and as $(N,T) \to \infty$,*

$$\hat{\theta}_k \to_p \theta_{k0}$$

for $k = 1, \dots, r$. If, further, $\theta_{k0} \in Int(\Theta_k)$, and $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$,

$$\sqrt{T}\left(\hat{\theta}_k - \theta_{k0}\right) \to_d \mathcal{N}\left(0, B(\xi_{k0})^{-1}\right).$$

This result is akin to those obtained by Hualde and Robinson (2011) and Nielsen (2015) for fractional time series models but unlike their setups, estimation herein is carried out on the estimated factors instead of observed series. Asymptotic elimination of the estimation effect requires N to grow for consistency and grow faster than $\sqrt{T}$ for asymptotic normality. This rate requirement is also common in slope estimation of heterogeneous panel data models, see, for example, Ergemen and Velasco (2017) and Pesaran (2006). The result in Theorem 1 contrasts to those in the long memory factor literature, see e.g. Cheung (2021), in that inference herein is parametric with a faster ($\sqrt{T}$) convergence rate, and there is no dependence on a user-chosen bandwidth parameter. Further, the requirement on the relative growth rates of N and T is standard and does not depend on the unknown memory parameters.

In practice, we may simplify the standard errors on $\hat{\delta}_k$ to $(T^{-1}(\pi^2/6)^{-1})^{1/2}$, and retrieve those on $\hat{\xi}_k'$ from estimation of an ARMA model for $\Delta_t^{\hat{\delta}_k} \hat{f}_{kt}$, although the long- and short-memory parameters are not asymptotically independent, as seen from the variance-covariance matrix, $B(\xi_{k0})^{-1}$, in Theorem 1.

Estimation under general idiosyncratic error dynamics. Alternative to the specification for idiosyncratic errors in Assumption A.1, e_{it} can be allowed to have pure fractional, short memory only or both long and short memory dynamics if we write $e_{it} = \varphi_t^{-1}(L; \zeta_{i0})\epsilon_{it}$ with $\epsilon_{it} \sim iid(0, \sigma_\epsilon^2)$ and $\zeta_{i0} = (\vartheta_{i0}, \varsigma_{i0}')'$ containing the long-memory parameters ϑ_{i0} and the vector of short-memory parameters ς_{i0}.

Cheung (2021) considers semiparametric estimation of ϑ_{i0} while being agnostic about ς_{i0} cf. Abadir et al. (2007). In our parametric setup, both ϑ_{i0} and ς_{i0} can be inferred from the CSS estimation based on some consistent estimate $\tilde{e}_{it}$, in a similar fashion to estimating the factor dynamics. While in practice, our method can be readily applied after estimating idiosyncratic error dynamics and then transforming the observed series to obtain iid residuals for PCA estimation, establishing asymptotic results in this case is made more difficult due to the convolution of two dynamic filters, in particular $\varphi_t(L;\zeta_i)\lambda_t^{-1}(L;\theta)$, on the factor errors and is therefore left for future research.

Estimating the number of factors. Under the dynamic specification in (2), we have so far assumed the number of factors, r, to be known. In practice, r is unknown and needs to be estimated. In the time domain, Onatski (2010) develops an estimation strategy for the number of factors, in which idiosyncratic and systematic eigenvalues of the sample covariance matrix are distinguished. This approach also handles substantial serial correlation in the idiosyncratic terms, which could arise due to model misspecifications. The required r_{max} value for the test can be determined following Ahn and Horenstein (2013)'s eigenvalue and growth ratio statistics. In Section 5, we employ these criteria since they are suitable under our setup, particularly allowing for strong dependence in the idiosyncratic errors.

4 Simulations

In this section we carry out Monte Carlo experiments to study the small-sample performance of the factor and memory estimates in the general case in which short and long memory dynamics are allowed in the factors, and cross-sectional and time dependence are allowed in the idiosyncratic errors. In particular, we generate the idiosyncratic errors $e_{it} = \rho_i e_{it-1} + u_{it}$, where $u_{it} \sim iidN(0,1)$ over t and fix $Corr(u_{it}u_{jt}) = 0.5$, while considering $\rho_i \in \{0.5, 1\}$. We draw the factor loadings γ_i as standard normals not to restrict the sign. We then generate the serially correlated common factors, $f_t = \xi_0 f_{t-1} + \Delta_t^{-\delta_0}\varepsilon_t$, based on the *iid* shocks ε_t drawn as standard normals and then fractionally integrated to the order δ_0. We focus on different cross-section sizes and time-series lengths, N and T, as well as different values of δ_0 and ξ_0. Simulations are based on 5,000 replications.

We investigate the finite-sample properties of our estimate of δ_0 based on both the factor estimates $\hat{f}_t$, i.e. $\hat{\delta}_{\hat{f}}$, and the true factors f_t, i.e. $\hat{\delta}_f$, as well as the consistency of factor estimates based on average R^2 measures. We set

$N = 50, 100, 250, 400$ and $T = 50, 100, 250, 400$ for values of $\delta_0 = 0, 0.3, 0.6, 1$ thus covering the stationary $I(0)$ case, a stationary long memory case, a slightly non-stationary case and the $I(1)$ (unit-root) case, respectively. Further, we consider $\xi_0 = 0, 0.5$ (as known) to cover pure fractional and serially correlated factor cases, and $r = 2$ to study the performance in the more challenging two-factor case: the performance in the $r = 1$ case is slightly better, especially when $\rho_i = 1$.

Tables 1 - 4 collect the simulation results. In all cases, an increasing N and T reduces the amount of bias in memory estimation, with T playing a more important role in bias reduction as expected. Factor serial correlation leads to slightly more biased estimates $\hat{\delta}_{\hat{f}}$, which is more pronounced for stationary values of δ_0. This is expected since short and long memory parameters are not independent of each other cf. Theorem 1. When $\rho_i = 1$ resembling the $I(1)$ errors case, the memory estimates $\hat{\delta}_{\hat{f}}$ suffer from some bias when $\delta_0 < 0.6$ but when factor serial correlation is additionally introduced, there is some bias cancellation. Also, a comparison of the two memory estimates $\hat{\delta}_{\hat{f}}$ and $\hat{\delta}_f$ reveals that $\hat{\delta}_{\hat{f}}$ typically requires a larger (N, T), especially when $\rho_i = 1$ and $\delta_0 < 0.6$. Figure 1 plots the histogram of the memory estimates based on estimated factors, $\hat{\delta}_{\hat{f}}$, for the empirically relevant case of $\rho_i = 1$ and $\xi_0 = 0.5$, and demonstrates a good approximation to the normal distribution for $\delta_0 = 1$ with $N = 400$ and $T = 400$.

In terms of consistency of the factor estimates, the factors are approximated well in all cases, and $\bar{R}^2$ increases in N and typically in δ_0 as a result of the increased signal strength.

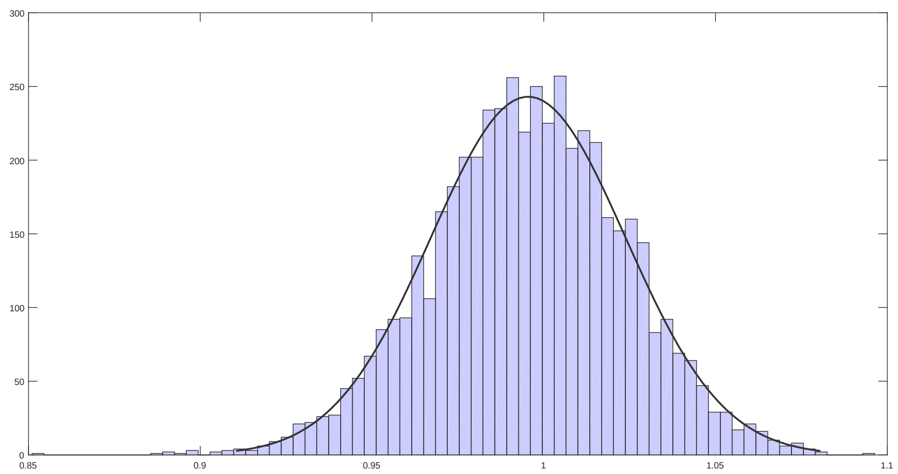

Figure 1: Histogram of estimated factor memories *vs.* the normal curve

5 Empirical Application

The study of financial volatility and its impact on economic activity and uncertainty has found considerable interest in the literature, see for example Chauvet et al. (2015) for a review. As an illustrative example of our methodology, we analyze the common persistence sources of U.S. industry RV's in relation to the market RV. Earlier, Ergemen and Velasco (2017) have disclosed cointegrating relationships between industry RV's and the market RV for 30 U.S. industries after projecting out the latent factor structure based on cross-sectionally averaged data. In this application, we estimate the common factors of industry RV's and their dynamics in order to conclude whether there is a significant difference between these two approaches in practice. Further in a predictive regression setup, we study how the industry RV common factors fare to the market RV and cross-sectionally averaged industry RV's as well as the news-based U.S. economic policy uncertainty (EPU) index due to Baker et al. (2016) for predicting industry RV's one-step ahead. Finally, we compare the out-of-sample forecasting performance of our factor model against $ARFIMA(1,d,0)$ specifications as in Luciani and Veredas (2015).

In order to calculate the monthly RV measures, we use daily average-value-weighted returns data for the time period January 3, 2000 – September 30, 2021 (T=261 months) from Kenneth French's Data Library for 49 industries in the U.S. economy. As for the composite market returns, we use a weighted average of daily returns of NYSE, NASDAQ and AMEX since the companies considered in industry returns trade in one of these markets. Using the composite index returns of NYSE, NASDAQ and AMEX, i.e. $r_{m,t}$, we calculate

$$RVM_t = \left(\sum_{s \in t}^{N_t} r_{m,s}^2 \right)^{1/2}, \qquad t = 1, 2, \ldots, T,$$

where N_t is the number of trading (typically 22) days in a month. Next, for each industry $i = 1, \ldots, 49$, we calculate

$$RVI_{i,t} = \left(\sum_{s \in t}^{N_t} e_{i,s}^2 \right)^{1/2}, \qquad t = 1, 2, \ldots, T,$$

where $e_{i,s} = r_{i,s} - r_{m,s}$, cf. Chauvet et al. (2015). While we choose these RV measures due to their simplicity, we note that several other RV measures are readily available in the literature and they could also be employed instead.

In Figure 2, we plot the estimated integration orders of industry RV's based on the CSS estimation. The estimated values are in the 0.5143-0.7124 range while the integration orders of the market RV, cross-sectionally averaged industry RV's and U.S. EPU index are 0.6264, 0.6458 and 0.6444, respectively, with the standard deviation of 0.0482.

Next, we estimate the latent common factors based on

$$RVI_{i,t} = \gamma_i' \hat{f}_t + e_{i,t}, \qquad t = 1, 2, \ldots, T.$$

We note that a drift term in (1), which is generally included to study level series in practice, can also be included here simply by augmenting $\hat{f}_t$ with a constant. Alternatively, it would also be possible to perform the estimation on the demeaned series.

To determine the number of factors, we employ the test proposed by Onatski (2010) and determine the required r_{max} value based on Ahn and Horenstein (2013), which suggests $r = 1$. We then perform diagnostic checks on the regression residuals. The estimated integration orders of the regression residuals are not significantly different from zero at the 5% significance level (with the highest estimated value of 0.0885) and an autoregressive fitting to the residuals yields an average $AR(1)$ parameter of 0.41, which is suitable given the conditions under Assumption A.1. Since the regression residuals are $I(0)$ in this application, it is more efficient to estimate the parameters based on level data instead of taking differences so we carry out the estimation steps in Section 3 in levels.

In Figure 3, we plot the PCA estimate of industry RV common factor whose integration order estimate, $\hat{\delta}_{\hat{f}}$, equals 0.6450 and the $AR(1)$ parameter estimate, $\hat{\xi}$, is 0.02 thus suggesting predominant long-range dependence characteristics. The estimated common factor and other indices all show similar long-memory dynamics based on their integration orders. Given the estimated integration orders of industry RV's in Figure 2 and the $I(0)$ regression residuals, the regressions of the industry RV's on the estimated factor as well as other indices are cointegrating regressions (with slope parameters positive and significantly different from zero). In Figure 4, we plot the explained variation in each industry RV based on the estimated common factor, cross-sectionally averaged industry RV's, the market RV and U.S. EPU index. The estimated common factor and cross-sectionally averaged industry RV's are comparable in terms of the R^2 measures while the market RV and EPU index have lower explanatory power on average.

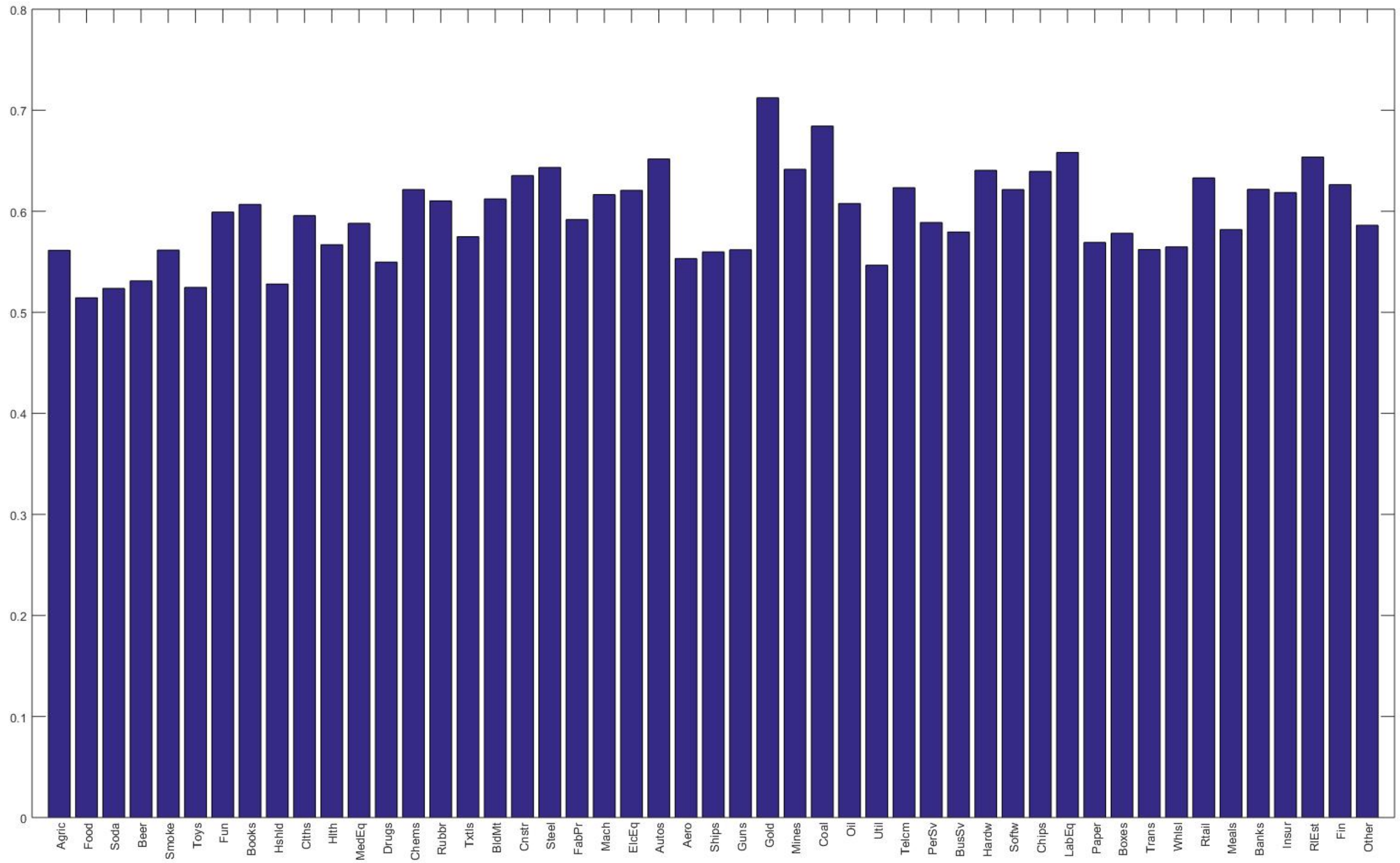

Figure 2: CSS estimates of industry RV integration orders. Standard error of these estimates is 0.0482.

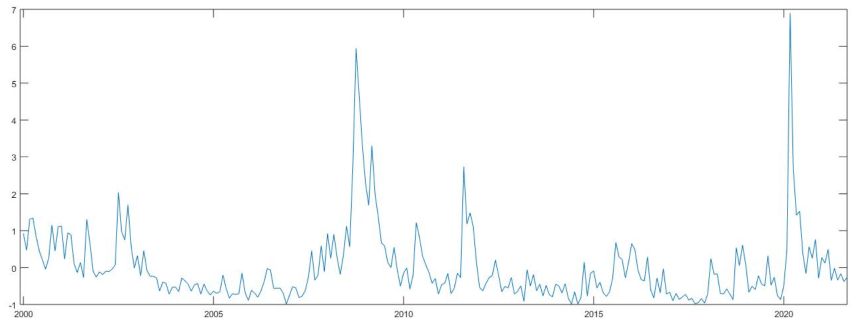

Figure 3: PCA estimate of industry RV common factor ($\hat{f}_t$).

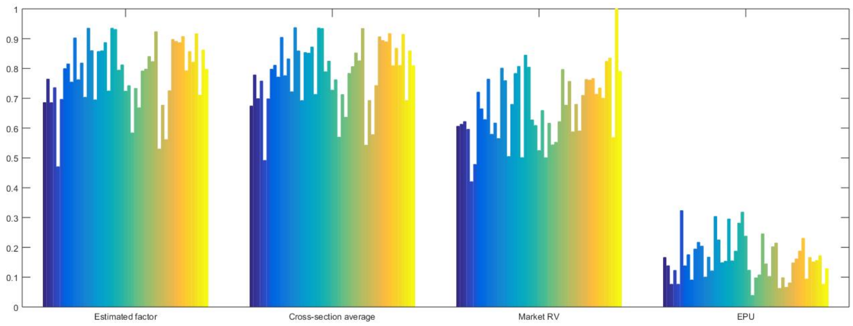

Figure 4: Explained variation in industry RV's across different indices.

Predicting industry RV's. We study the ability of the estimated factor for predicting industry RV's in comparison to the other indices based on the following regression:

$$RVI_{i,t} = \beta_i \hat{f}_{t-1} + v_{i,t}, \qquad t = 2, \ldots, T, \tag{16}$$

where $v_{i,t}$ are assumed to satisfy the conditions in Assumption A.1. The estimates of β_i obtained from (16) are all positive and significantly different from zero at the 5% significance level, and the R^2 measures are reported together with the predictive performances of other indices in Figure 5. According to these results, the estimated factor, cross-sectionally averaged data and the market RV all show comparable predictive ability with the average R^2 measures of 0.3787, 0.3715 and 0.3356, respectively, while the EPU index performs significantly worse with an average R^2 measure of 0.1212.

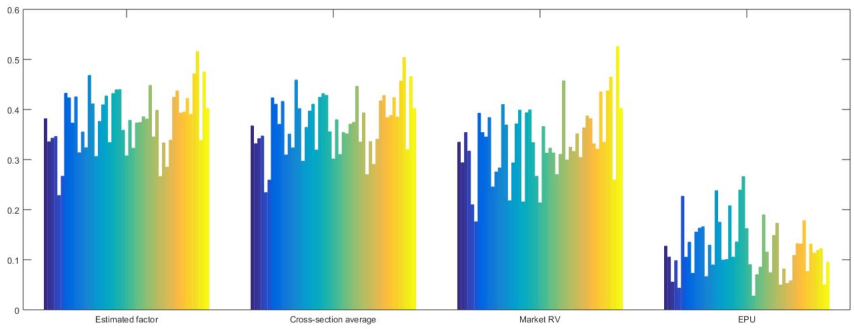

Figure 5: Predictive ability for industry RV's across different indices.

Overall, the estimated factor and its proxy based on the cross-section average of data, as used, among others, by Pesaran (2006) and Ergemen and Velasco (2017), show equal performance in terms of explaining the industry RV's as well as predicting them in sample.

Out-of-sample forecasts for industry RV's. Given the outstanding in-sample predictive ability of the estimated factor and its proxy based on the cross-section average of data, we analyze the out-of-sample forecasting performance for different specifications and forecast horizons. We consider forecasts based on autoregressive fractionally integrated moving average ($ARFIMA(1, d, 0)$) specifications as in Luciani and Veredas (2015), the estimated factor ($\hat{f}_t$), $\hat{f}_t$ augmented with $ARFIMA(1, d, 0)$, cross-sectionally averaged data (*C-S avg*), and *C-S avg* augmented with $ARFIMA(1, d, 0)$. The forecasts are obtained for one- and four-month horizons in an expanding window scheme. The models are then evaluated in terms of their forecast performance based on the model confidence set (MCS) approach by Hansen et al. (2011).

We consider the January 2000 - April 2017 in-sample period, corresponding to 80% of the data, and obtain forecasts for the May 2017 - September 2021 out-of-sample period, to stress-test the method by also including the abrupt RV changes in the recent COVID-19 period. We collect the 95% MCS, $\widehat{\mathcal{M}}_{0.95}$, p-values over different horizons in Table 5. The results show that in one-month-ahead forecasts, $ARFIMA(1, d, 0)$ is outperformed for all industries, and the models including $\hat{f}_t$ and *C-S avg* are consistently contained in $\widehat{\mathcal{M}}_{0.95}$. Further, the specifications including $\hat{f}_t$ outperform those only with the *C-S avg* for most industries. For the four-month horizon, models including $\hat{f}_t$ are still dominating although the results are heterogeneous: $ARFIMA(1, d, 0)$ is outperformed for 40 out of 49 industries mainly by the specifications including $\hat{f}_t$, which perform the best for 24 out of 49 industries. Models based on *C-S avg* performs the best for 16 out of 49 industries.

Overall, we can conclude that the estimated factor $\hat{f}_t$ and its augmented specifications can be used as benchmarks for forecasting industry RV's while in several cases, also a measure as simple as *C-S avg* can be a powerful forecasting tool.

6 Final Remarks

We have considered a dynamic factor model with arbitrary persistence in the common factors. Under a large N, T setup, we estimate the common factors based on PCA and employ a CSS criterion to estimate factor memories, in both of which cases the estimates are consistent at the standard parametric rates and the factor memory estimates have centered normal asymptotics. Simulations based on Monte Carlo experiments indicate that the method works even in small panels, and then we apply it to a panel of U.S. industry realized volatilities. While the setup allows for general dynamics in the factors, it can be extended nontrivially in the following directions: 1) asymptotic results can be established under more general stochastic dynamics in the idiosyncratic errors, e_{it}; 2) time-varying loadings, $\gamma_{i,t}$, can be introduced; 3) general factor dynamics under nonlinear setups can be investigated; 4) spatial dependence can be incorporated into the model for more robust insights in empirical applications; 5) in a frequency-domain approach, singularity in the spectrum taking place at a frequency away from zero or at seasonal frequencies can be considered.

References

Abadir, K. M., W. Distaso, and L. Giraitis (2007): "Nonstationarity-extended local Whittle estimation," *Journal of Econometrics*, 141, 1353–1384.

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, 191–221.

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

——— (2008): "Large Dimensional Factor Analysis," *Foundations and Trends (R) in Econometrics*, 3, 89–163.

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

Baker, R., N. Bloom, and S. J. Davis (2016): "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131(4), 1593–1636.

Baltagi, B., C. Kao, and F. Wang (2021): "Estimating and Testing High Dimensional Factor Models with Multiple Structural Changes," *Journal of Econometrics*, 220(2), 349–365.

Barigozzi, M. and M. Hallin (2016): "Generalized dynamic factor models and volatilities: recovering the market volatility shocks," *The Econometrics Journal*, 19.

Barigozzi, M., M. Lippi, and M. Luciani (2016): "Non-Stationary Dynamic Factor Models for Large Datasets," *Working Paper*, arXiv:1602.02398.

——— (2021): "Large-dimensional Dynamic Factor Models: Estimation of Impulse-Response Functions with I(1) cointegrated factors," *Journal of Econometrics*, 221, 455–482.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach," *The Quarterly Journal of Economics*, 120, 387–422.

Bollerslev, T., D. Osterrieder, N. Sizova, and G. Tauchen (2013): "Risk and Return: Long-Run Relationships, Fractional Cointegration, and Return Predictability," *Journal of Financial Economics*, 108(2), 409–424.

Chamberlain, G. and M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica: Journal of the Econometric Society*, 1281–1304.

Chambers, M. J. (1998): "Long Memory and Aggregation in Macroeconomic Time Series," *International Economic Review*, 39(4), 1053–1072.

Chauvet, M., Z. Senyuz, and E. Yoldas (2015): "What Does Realized Volatility Tell Us About Macroeconomic Fluctuations?" *Journal of Economic Dynamics and Control*, 52, 340–360.

Chen, W. W. and C. M. Hurvich (2006): "Semiparametric Estimation of Fractional Cointegrating Subspaces," *Annals of Statistics*, 34(6), 2939–2979.

Cheung, Y. L. (2021): "Long Memory Factor Model: On Estimation of Factor Memories," *Journal of Business and Economic Statistics*, DOI: 10.1080/07350015.2020.1867559.

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

Cipollini, A. and G. Kapetanios (2008): "A stochastic variance factor model for large datasets and an application to S&P data," *Economics Letters*, 100, 130–134.

Cristadoro, R., M. Forni, L. Reichlin, and G. Veronese (2005): "A core inflation indicator for the euro area," *Journal of Money, Credit, and Banking*, 37, 539–560.

Dordonnat, V., S. J. Koopman, and M. Ooms (2012): "Dynamic factors in periodic time-varying regressions with an application to hourly electricity load mo-delling," *Computational Statistics & Data Analysis*, 56, 3134–3152.

Ergemen, Y. E., N. Haldrup, and C. V. Rodríguez-Caballero (2016): "Common long-range dependence in a panel of hourly Nord Pool electricity prices and loads," *Energy Economics*, 60, 79–96.

Ergemen, Y. E. and C. Velasco (2017): "Estimation of Fractionally Integrated Panels with Fixed-Effects and Cross-Section Dependence," *Journal of Econometrics*, 196(2), 248–258.

Fan, J., Y. Ke, and Y. Liao (2021): "Augmented Factor Models with Applications to Validating Market Risk Factors and Forecasting Bond Risk Premia," *Journal of Econometrics*, 222(1), 269–294.

Fan, J., Y. Liao, and M. Mincheva (2013): "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society – Series B*, 75, 603–680.

Favero, C. A., M. Marcellino, and F. Neglia (2005): "Principal components at work: the empirical analysis of monetary policy with large data sets," *Journal of Applied Econometrics*, 20, 603–620.

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

——— (2004): "The generalized dynamic factor model consistency and rates," *Journal of Econometrics*, 119, 231–255.

——— (2005): "The generalized dynamic factor model: one-sided estimation and forecasting," *Journal of the American Statistical Association*, 100, 830–840.

Gil-Alaña, L. and P. Robinson (1997): "Testing of Unit Root and Other Nonstationary Hypotheses in Macroeconomic Time Series," *Journal of Econometrics*, 80(2), 241–268.

Granger, C. (1980): "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227–238.

Hallin, M., C. Mathias, H. Pirotte, and D. Veredas (2011): "Market liquidity as dynamic factors," *Journal of econometrics*, 163, 42–50.

Hansen, P. R., A. Lunde, and J. M. Nason (2011): "The model confidence set," *Econometrica*, 79, 453–497.

Hartl, T. (2020): "Macroeconomic Forecasting with Fractional Factor Models," *arXiv preprint arXiv:2005.04897*.

Hartl, T. and R. Weigand (2019): "Multivariate fractional components analysis," *arXiv preprint arXiv:1812.09149*.

Hosoya, Y. (2005): "Fractional Invariance Principle," *Journal of Time Series Analysis*, 26, 463–486.

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.

Kapetanios, G. (2004): "A note on modelling core inflation for the UK using a new dynamic factor estimation method and a large disaggregated price index dataset," *Economics Letters*, 85, 63–69.

Karabiyik, H. and J. Westerlund (2021): "Forecasting using cross-section averageâaugmented time series regressions," *The Econometrics Journal*, 24(2), 315–333.

Luciani, M. and D. Veredas (2015): "Estimating and Forecasting Large Panels of Volatilities with Approximate Dynamic Factor Models," *Journal of Forecasting*, 34(3), 163–176.

Marinucci, D. and P. Robinson (2000): "Weak Convergence of Multivariate Fractional Processes," *Stochastic Processes and their Applications*, 86, 103–120.

Michelacci, C. and P. Zaffaroni (2000): "(Fractional) Beta Convergence," *Journal of Monetary Economics*, 45, 129–153.

Morana, C. (2007): "On the Macroeconomic Causes of Exchange Rates Volatility," *ICER Working Papers*.

Nielsen, M. Ø. (2015): "Asymptotics for the Conditional-Sum-of-Squares Estimator in Multivariate Fractional Time Series Models," *Journal of Time Series Analysis*, doi: 10.1111/jtsa.12100.

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

Pesaran, H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012.

Pesaran, M. H. and A. Chudik (2014): "Aggregation in Large Dynamic Panels," *Journal of Econometrics*, 178, 273–285.

Ray, B. K. and R. S. Tsay (2000): "Long-Range Dependence in Daily Stock Volatilities," *Journal of Business and Economic Statistics*, 18(2), 254–262.

Robinson, P. M. (1978): "Statistical Inference for a Random Coefficient Autoregressive Model," *Scandinavian Journal of Statistics*, 5, 163–168.

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

——— (2019): "Estimation for dynamic panel data with individual effects," *Econometric Theory*, 36, 185–222.

Stock, J. H. and M. W. Watson (2002a): "Forecasting using principal components from a large number of predictors," *Journal of the American statistical association*, 97, 1167–1179.

——— (2002b): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

Uematsu, Y. and T. Yamagata (2022): "Estimation of Sparsity-Induced Weak Factor Models," *Journal of Business and Economic Statistics*, https://doi.org/10.1080/07350015.2021.2008405.

7 Appendix

7.1 Proof of Lemma 1

For the proof, we only show that our assumptions match those imposed by Barigozzi et al. (2021) for proving their Lemma 1 although in (1) we do not consider deterministic components, so the proof under our setup follows easier steps. Under our Assumption A.1, $\varepsilon_t \sim iid(0, \Omega_\varepsilon)$, $\Omega_\varepsilon > 0$, with $E\left\|\varepsilon_t\right\|^4 < \infty$, $\text{rank}\big(E\left[\Delta_t^{\delta_0} f_t \Delta_t^{\delta_0} f_t'\right]\big) = r$, with r fixed, and $E\left[\left(\Delta_t^{\delta_{k0}} f_{kt}\right)^2\right] > E\left[\left(\Delta_t^{\delta_{l0}} f_{lt}\right)^2\right] > 0$ for all $k, l = 1, \ldots, r$ with $k < l$. These conditions match their Assumption 1, except our factor DGP is assumed to be given by (2). However, we can write cf. (3),

$$\Delta_t^{\delta_{k0}} f_{kt} = \psi(L; \xi_{k0})^{-1} \varepsilon_{kt}, \quad t = 1, \ldots, T,$$

and under Assumption A.4, the RHS induces short-memory dynamics resembling the conditions in Barigozzi et al. (2021)'s Assumption 1(b). The assumptions regarding loadings given in our Assumption A.3 match their Assumption 2.

The conditions $(1 - \rho_i L)e_{it} = a_i(L)u_{it}$ for all i, where $a_i(L) = \sum_{k=0}^{\infty} a_{ik} L^k$ with $\sum_{k=0}^{\infty} k\left|a_{ik}\right| \le M$, and $|\rho_i| \le 1$; $u_t = (u_{1t}, \ldots, u_{Nt})'$, satisfy $u_t \sim iid\left(0, \Omega_u\right)$, $\Omega_u > 0$, with $E\left\|u_t\right\|^4 < \infty$, and $E\left(u_{it}u_{jt}\right) = \tau_{ij}$ with $\sum_{j=1}^{N} |\tau_{ij}| \le M$ uniformly in i in Assumption A.1, together with our Assumption A.2 match those in Assumption 3 of Barigozzi et al. (2021). Finally, their Assumption 4 is stated within our Assumption A.1 as for all i and $t \le 0$, $u_{it} = 0$ and $\varepsilon_t = 0$.

Therefore, the uniform result on the loadings readily applies in our case. For obtaining the convergence rates of the factor estimates, our setup is not affected by the estimation of deterministic components, which is why the rate $N^{-(1-\eta)}$, $\eta < 1$, in Lemma 1 of Barigozzi et al. (2021) is not required here. □

7.2 Proof of Theorem 1

We begin by writing from (12)

$$\begin{aligned}\hat{\varepsilon}_{kt}(\theta_k) =& \lambda_t(L;\theta_k)f_{kt} + \lambda_t(L;\theta_k)\left(\hat{f}_{kt} - f_{kt}\right),\\ =& \varepsilon_{kt}(\theta_k) + \lambda_t(L;\theta_k)\left(\hat{f}_{kt} - f_{kt}\right)\end{aligned}$$

suppressing dependence on the matrix J in Lemma 1 since the factor dynamics are not affected by the signs of the factors. Then, making the dependence on the estimation effect $\hat{f}_{kt} - f_{kt}$ explicit in (14),

$$\begin{aligned}L_{k,T}(\theta_k;\hat{f}_{kt}) =& \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{kt}(\theta_k)^2,\\ =& \frac{1}{T}\sum_{t=1}^{T}\left[\varepsilon_{kt}(\theta_k) + \lambda_t(L;\theta_k)\left(\hat{f}_{kt} - f_{kt}\right)\right]^2,\\ =:& L_T(\theta_k;f_t) + O_p\left(T^{-1} + (NT)^{-1/2} + N^{-1}\right)\end{aligned}$$

by the result in Lemma 1. Then, as $(N,T) \to \infty$, $L_{k,T}(\theta_k;\hat{f}_{kt}) = L_{k,T}(\theta_k;f_{k,t}) + o_p(1)$, and $\sqrt{T}\frac{\partial}{\partial\theta_k}L_{k,T}(\theta_k;\hat{f}_{kt}) = \sqrt{T}\frac{\partial}{\partial\theta_k}L_{k,T}(\theta_k;f_{kt}) + O_p\left(T^{1/2}N^{-1}\right)$, which is why $\sqrt{T}/N \to 0$ is required for asymptotic normality in Theorem 1. Given these, for the rest of the proof, we show the results based on $L_{k,T}(\theta_k;f_{kt})$ and suppress the dependence on f_{kt} to write $L_{k,T}(\theta_k)$.

7.2.1 Proof of Consistency

As $(N,T) \to \infty$, the CSS criterion can be written as

$$L_{k,T}(\theta_k) = \frac{1}{T}\sum_{t=1}^{T}\left(\lambda_t^0\left(L;\theta_k\right)\varepsilon_{kt}\right)^2 \tag{17}$$

where

$$\lambda_t^0\left(L;\theta_k\right) = \lambda_t\left(L;\theta_k\right)\lambda_t^{-1}\left(L;\theta_{k0}\right) = \sum_{j=0}^{t}\lambda_j^0\left(\theta_k\right)L^j.$$

Following Hualde and Robinson (2011), we give the proof for the most general case where possibly $\underline{\delta}_k \leq \delta_{k0} - 1/2$. While δ_k may take arbitrary values from $[\underline{\delta}_k, \overline{\delta}_k]$ for given $\underline{\delta}_k < \overline{\delta}_k$, ensuring uniform convergence of $L_{k,T}(\theta_k)$ requires the study of cases depending on $\delta_{k0} - \delta_k$, noting that under Assumption A.4 short memory dynamics are dominated by long memory properties, see Robinson and Velasco (2015). We analyze each case separately in the following.

In the proof, we adopt the arguments in Hualde and Robinson (2011). For $\epsilon_k > 0$, define $Q_{k\epsilon} = \{\theta_k : |\delta_k - \delta_{k0}| < \epsilon_k\}$, $\overline{Q}_{k\epsilon} = \{\theta_k : \theta_k \notin Q_{k\epsilon}, \delta_k \in \mathcal{D}_k\}$. For small enough ϵ_k,

$$Pr(\hat{\theta}_k \in \overline{Q}_{k\epsilon}) \leq Pr\left(\inf_{\Theta_k \in \overline{Q}_{k\epsilon}} S_{k,T}(\theta_k) \leq 0\right)$$

where $S_{k,T}(\theta_k) = L_{k,T}(\theta_k) - L_{k,T}(\theta_{k0})$. In the rest of the proof, we will show that $L_{k,T}(\theta_k)$, and thus $S_{k,T}(\theta_k)$, converges in probability to a well-behaved function when $\delta_{k0} - \delta_k < 1/2$ and diverges when $\delta_{k0} - \delta_k \geq 1/2$. In order to analyze the asymptotic behavior of $S_{k,T}(\delta_k)$ in a neighborhood of $\delta_k = \delta_{k0} - 1/2$, a special treatment is required. For arbitrarily small $\zeta_k > 0$, such that $\zeta_k < \delta_{k0} - 1/2 - \underline{\delta}_k$, define the disjoint sets $\Theta_{k1} = \{\theta_k : \underline{\delta}_k \leq \delta_k \leq \delta_{k0} - 1/2 - \zeta_k\}$, $\Theta_{k2} = \{\theta_k : \delta_{k0} - 1/2 - \zeta_k < \delta_k < \delta_{k0} - 1/2\}$, $\Theta_{k3} = \{\theta_k : \delta_{k0} - 1/2 \leq \delta_k \leq \delta_{k0} - 1/2 + \zeta_k\}$, and $\Theta_{k4} = \left\{\theta_k : \delta_{k0} - 1/2 + \zeta_k < \delta_k \leq \overline{\delta}_k\right\}$, so that $\Theta_k = \cup_{l=1}^{4} \Theta_{kl}$. We will show

$$Pr\left(\inf_{\theta_k \in \overline{Q}_{k\epsilon} \cap \Theta_{kl}} S_{k,T}(\delta_k) \leq 0\right) \to 0 \text{ as } T \to \infty,\ l = 1, \ldots, 4. \tag{18}$$

Proof of (18) **for** $l = 4$. We denote by σ_k^2 the k-th diagonal element of Ω_ε, i.e. $E(\varepsilon_{kt}^2) = \sigma_k^2$, and show that

$$\sup_{\theta_k \in \Theta_{k4}} \left| \frac{1}{T} \sum_{t=1}^{T} \left[(\lambda_t^0 (L; \theta_k) \varepsilon_{kt})^2 - \sigma_k^2 \sum_{j=0}^{\infty} \lambda_j^0 (\theta_k)^2 \right] \right| = o_p(1). \tag{19}$$

We first show, following Hualde and Robinson (2011), that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T} (\lambda_t^0 (L; \theta_k) \varepsilon_{kt})^2 &= \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{j=0}^{t} \lambda_j^0 (\theta_k) \varepsilon_{kt-j} \right)^2 \\ &\to_p \sigma_k^2 \sum_{j=0}^{\infty} \lambda_j^0 (\theta_k)^2, \end{aligned}$$

uniformly in δ_k by Assumption A.1 as $T \to \infty$ since $-1/2 + \zeta_k < \delta_k - \delta_{k0}$ for some $\zeta_k > 0$. Since the limit is uniquely minimized at $\theta_k = \theta_{k0}$, as it is positive for all $\theta_k \neq \theta_{k0}$, (18) holds for $l = 4$ if (19) holds.

To check (19), we show

$$\sup_{\theta_k \in \Theta_{k4}} \left| \frac{1}{T} \sum_{t=1}^{T} \left[\left(\sum_{j=0}^{t} \lambda_j^0 (\theta_k) \varepsilon_{kt-j} \right)^2 - \mathbb{E} \left(\left(\sum_{j=0}^{t} \lambda_j^0 (\theta_k) \varepsilon_{kt-j} \right)^2 \right) \right] \right| = o_p(1),$$

where the term in absolute value is

$$\frac{1}{T}\sum_{j=0}^{T}\lambda_j^0\left(\theta_k\right)^2\sum_{l=1}^{T-j}(\varepsilon_{kl}-\sigma_k^2)$$
$$+\frac{2}{T}\sum_{j=0}^{T}\lambda_j^0\left(\theta_k\right)^2\sum_{l=s-j+1}^{T-j}\varepsilon_{kl}\varepsilon_{kl-(s-j)}=(a)+(b). \tag{20}$$

We have

$$\mathbb{E}\sup_{\Theta_{k4}}|(a)|\leq\left(\frac{1}{T}\sum_{j=0}^{T}\sup_{\Theta_{k4}}\lambda_j^0\left(\theta_k\right)^2\mathbb{E}\left|\sum_{l=1}^{T-j}(\varepsilon_{kl}^2-\sigma_k^2)\right|\right).$$

Further, $\mathrm{Var}(\sum_{l=1}^{T-j}\varepsilon_{kl}^2)=O(T)$, uniformly in j, so using $-1/2+\zeta_k<\delta_k-\delta_{k0}$,

$$\sup_{\Theta_{k4}}|(a)|=O_p\left(T^{-1/2}\sum_{j=1}^{\infty}j^{-2\zeta_k-1}\right)=O_p(T^{-1/2}).$$

By summation by parts, the term (b) in (20) is

$$\frac{2\lambda_T^0\left(\theta_k\right)}{T}\sum_{j=0}^{T-1}\sum_{s=j+1}^{T-1}\sum_{l=s-j+1}^{T-j-1}\lambda_j^0\left(\theta_k\right)'\varepsilon_{kl}\varepsilon_{kl-(s-j)}$$
$$-\frac{2}{T}\sum_{j=0}^{T-1}\lambda_j^0\left(\theta_k\right)\sum_{s=j+1}^{T-1}\left[\lambda_{s+1}^0\left(\theta_k\right)-\lambda_s^0\left(\theta_k\right)\right]\sum_{r=j+1}^{s}\sum_{l=r-j+1}^{T-j-1}\varepsilon_{kl}\varepsilon_{kl-(r-j)}$$
$$=\ (b_1)+(b_2)\,.$$

Then, using that $\mathrm{Var}\left(\sum_{s=j+1}^{T-2}\sum_{l=k-j+1}^{T-j-1}\left\{\varepsilon_{kl}\varepsilon_{kl-(s-j)}\right\}\right)=O\left(T^2\right)$ uniformly in j,

$$\mathbb{E}\sup_{\Theta_{k4}}|(b_1)|\leq T^{-\zeta_k-3/2}\sum_{j=1}^{T}j^{-\zeta_k-1/2}\mathrm{Var}\left(\sum_{s=j+1}^{T-2}\sum_{l=s-j+1}^{T-j-1}\left\{\varepsilon_{kl}\varepsilon_{kl-(s-j)}\right\}\right)^{1/2}\leq T^{-2\zeta_k},$$

while

$$\begin{aligned}\mathbb{E}\sup_{\Theta_{k4}}|(b_2)| &\leq T^{-1}\sum_{j=1}^{T}j^{-\zeta_k-1/2}\sum_{s=j+1}^{T-1}s^{-\zeta_k-3/2}\mathrm{Var}\left(\sum_{r=j+1}^{s}\sum_{l=r-j+1}^{T-j}\left\{\varepsilon_{kl}\varepsilon_{kl-(r-j)}\right\}\right)^{1/2}\\ &\leq T^{-1/2}\sum_{j=1}^{T}j^{-\zeta_k-1/2}\sum_{s=j+1}^{T-1}s^{-\zeta_k-3/2}\left(s-j\right)^{1/2}\leq KT^{-2\zeta_k},\end{aligned}$$

and therefore $(b)=O_p(T^{-2\zeta_k})=o_p(1)$. This proves (19) and thus (18) for $l=4$.

Proof of (18) **for** $l = 3, 2$. The uniform convergence of the idiosyncratic component in the proof of (18) follows the same steps as for $l = 4$, based on the arguments in Hualde and Robinson (2011).

Proof of (18) **for** $l = 1$. Noting that

$$L_{k,T}(\theta_k) = \frac{1}{T}\sum_{t=1}^{T}\left(\lambda_t^0(L;\theta_k)\,\varepsilon_{kt}\right)^2 \geq \frac{1}{T^2}\left(\sum_{t=1}^{T}\lambda_t^0(L;\theta_k)\,\varepsilon_{kt}\right)^2,$$

we write

$$Pr\left(\inf_{\Theta_{k1}} L_T(\theta_k) > K\right) \geq Pr\left(T^{2\zeta_k}\inf_{\Theta_{k1}}\left(\frac{1}{T^{\delta_{k0}-\delta_k+1/2}}\sum_{t=1}^{T}\lambda_t^0(L;\theta_k)\,\varepsilon_{kt-j}\right)^2 > K\right)$$

since $\delta_k - \delta_{k0} \leq -1/2 - \zeta_k$. For arbitrarily small $\epsilon_k > 0$, we show

$$\begin{aligned} & Pr\left(T^{2\zeta_k}\inf_{\Theta_{k1}}\left(\frac{1}{T^{\delta_{k0}-\delta_k+1/2}}\sum_{t=1}^{T}\lambda_t^0(L;\theta_k)\,\varepsilon_{kt-j}\right)^2 > K\right) \\ \geq\ & Pr\left(\inf_{\Theta_{k1}}\left(\frac{1}{T^{\delta_{k0}-\delta_k+1/2}}\sum_{t=1}^{T}\lambda_t^0(L;\theta_k)\,\varepsilon_{kt-j}\right)^2 > \epsilon_k\right) \to 1 \end{aligned}$$

as $T \to \infty$. Define $h_{kT}(\delta_k) = T^{-\delta_{k0}+\delta_k-1/2}\lambda_t^0(L;\theta_k)\,\varepsilon_{kt-j} = T^{-1/2}\sum_{j=0}^{T}\frac{\lambda_j^0(\theta_k)}{T^{\delta_{k0}-\delta_k}}\varepsilon_{kt-j}$. By the weak convergence results in Marinucci and Robinson (2000),

$$h_{kT}(\delta_k) \Rightarrow \lambda_\infty^0(1;\theta_k)\int_0^1\frac{(1-s)^{\delta_{k0}-\delta_k}}{\Gamma(\delta_{k0}-\delta_k+1)}\delta_k B_k(s)$$

as $T \to \infty$, where $B_k(s)$ is a scalar Brownian motion, and by $\Rightarrow$ we mean convergence in the space of continuous functions in Θ_{k1} with uniform metric. Tightness and finite dimensional convergence follows from the fractional invariance property presented in Theorem 1 in Hosoya (2005) as well as $\sup_T \mathbb{E}\left[h_{kT}(\delta_k)^2\right] < \infty$. Then, as $T \to \infty$,

$$\begin{aligned} & \left(\frac{1}{T^{\delta_{k0}-\delta_k+1/2}}\sum_{t=1}^{T}\lambda_t^0(L;\theta_k)\,\varepsilon_{kt-j}\right)^2 \\ \to_p\ & \lambda_\infty^0(1;\theta_k)^2\,\mathrm{Var}\left(\int_0^1\frac{(1-s)^{\delta_{k0}-\delta_k}}{\Gamma(\delta_{k0}-\delta_k+1)}\delta_k B_k(s)\right) \\ =\ & \frac{\sigma_k^2\lambda_\infty^0(1;\theta_k)^2}{(2(\delta_{k0}-\delta_k)+1)\,\Gamma^2(\delta_{k0}-\delta_k+1)}, \end{aligned}$$

uniformly in $\theta_k \in \Theta_{k1}$, where

$$\inf_{\Theta_{k1}} \lambda_\infty^0 (1;\theta)^2 \operatorname{Var}\left(\int_0^1 \frac{(1-s)^{\delta_{k0}-\delta_k}}{\Gamma(\delta_{k0}-\delta_k+1)}\delta_k B_k(s)\right) = \frac{\sigma_k^2}{(2(\delta_{k0}-\underline{\delta}_k)+1)\,\Gamma^2(\delta_{k0}-\underline{\delta}_k+1)} > 0,$$

so that

$$Pr\left(\inf_{\Theta_{k1}}\left(\frac{1}{T^{\delta_{k0}-\delta_k+1/2}}\sum_{t=1}^{T}\lambda_t^0\left(L;\theta_k\right)\varepsilon_{kt-j}\right)^2 > \epsilon_k\right) \to 1 \quad \text{as } T\to\infty$$

and (18) follows for $l=1$ as ϵ_k is arbitrarily small. □

7.2.2 Proof of Asymptotic Normality

We first analyze the $\sqrt{T}$-normalized first derivative of $L_{k,T}(\theta_k)$ evaluated at $\theta_k = \theta_{k0}$,

$$\sqrt{T}\frac{\partial}{\partial\theta_k}L_{k,T}(\theta_k)\bigg|_{\theta_k=\theta_{k0}} = \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{kt}\left(\chi_t\left(L;\xi_{k0}\right)\varepsilon_{kt}\right) \tag{21}$$

where $\chi_t\left(L;\xi_{k0}\right)$ is the truncated version of (15) evaluated at $\xi_k = \xi_{k0}$.

Then, applying Proposition 2 in Robinson and Velasco (2015), we have as $T\to\infty$,

$$\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{kt}\left(\chi_t\left(L;\xi_{k0}\right)\varepsilon_{kt}\right) \to_d \mathcal{N}\left(0, 4\sigma_k^4 B\left(\xi_{k0}\right)\right)$$

under Assumptions A.1 and A.4.

Finally, we analyze the second derivative of $L_{k,T}\left(\theta_k\right)$ evaluated at $\theta_k=\theta_{k0}$, $\left(\partial^2/\partial\theta_k\partial\theta_k'\right)L_{k,T}(\theta_k)|_{\theta_k=\theta_{k0}}$, which equals

$$\frac{2}{T}\sum_{t=1}^{T}\left(\chi_t\left(L;\xi_{k0}\right)\varepsilon_{kt}\right)\left(\chi_t\left(L;\xi_{k0}\right)\varepsilon_{kt}\right)' + \frac{2}{T}\sum_{t=1}^{T}\varepsilon_{kt}\left(b_t^0(L)\varepsilon_{kt}\right)$$

where $b_t^0\left(L\right) = \dot{\chi}_t\left(L;\xi_{k0}\right) + \chi_t\left(L;\xi_{k0}\right)\chi_t\left(L;\xi_{k0}\right)'$ and $\dot{\chi}_t\left(L;\xi_k\right) = \left(\partial/\partial\theta'\right)\chi_t\left(L;\xi_k\right)$. Therefore, as $T\to\infty$,

$$\frac{\partial^2}{\partial\theta_k\partial\theta_k'}L_{k,T}(\theta_k)|_{\theta_k=\theta_{k0}} \to_p 2\sigma_k^2 B\left(\xi_{k0}\right).$$

The convergence of the Hessian of $L_{k,T}(\theta_k)$,

$$\ddot{L}_{k,T}(\theta_k) \to_p \ddot{L}_{k,T}(\theta_{k0})$$

can be shown as in Theorem 2 of Hualde and Robinson (2011) under Assumptions A.1 and A.4 since $\hat{\theta}_k \to_p \theta_{k0}$. The proof is then complete. $\square$

Table 1: **Memory Estimation Bias and Factor Estimate Consistency with** $r = 2$, $\xi_0 = 0$, **and** $\rho_i = 0.5$

This table reports the memory bias based on $\hat{f}_t$ (i.e. $\hat{\delta}_{\hat{f}} - \delta_0$), memory bias based on f_t (i.e. $\hat{\delta}_f - \delta_0$), and the average R^2 measure from the regression of f_t on $\hat{f}_t$ (i.e. $\bar{R}^2$) across $\delta_0 \in \{0, 0.3, 0.6, 1\}$, cross-section dimensions $N \in \{50, 100, 250, 400\}$, and time series lengths $T \in \{50, 100, 250, 400\}$. Standard error of the memory estimates are 0.1103, 0.0779, 0.0493 and 0.0389, respectively for $T = 50, 100, 250, 400$.

	$\hat{\delta}_{\hat{f}} - \delta_0$				$\hat{\delta}_f - \delta_0$				$\bar{R}^2$			
$\delta_0 = 0$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	0.0084	0.0078	0.0075	0.0074	0.0055	0.0054	0.0054	0.0054	0.9888	0.9948	0.9972	0.9983
100	0.0041	0.0038	0.0036	0.0035	0.0028	0.0028	0.0028	0.0028	0.9849	0.9913	0.9967	0.9978
250	0.0017	0.0015	0.0014	0.0013	0.0012	0.0012	0.0012	0.0011	0.9832	0.9913	0.9965	0.9974
400	0.0011	0.0010	0.0009	0.0008	0.0007	0.0007	0.0007	0.0007	0.9812	0.9896	0.9955	0.9973
$\delta_0 = 0.3$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	-0.0010	-0.0013	-0.0014	-0.0017	-0.0028	-0.0028	-0.0030	-0.0031	0.9900	0.9960	0.9975	0.9985
100	-0.0006	-0.0005	-0.0007	-0.0007	-0.0013	-0.0009	-0.0011	-0.0011	0.9871	0.9936	0.9977	0.9986
250	-0.0001	-0.0001	-0.0002	-0.0002	-0.0003	-0.0002	-0.0003	-0.0003	0.9859	0.9921	0.9969	0.9979
400	0.0000	0.0000	-0.0001	-0.0002	-0.0002	-0.0001	-0.0001	-0.0002	0.9861	0.9919	0.9970	0.9978
$\delta_0 = 0.6$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	-0.0041	-0.0039	-0.0039	-0.0041	-0.0028	-0.0030	-0.0030	-0.0032	0.9959	0.9978	0.9990	0.9994
100	-0.0018	-0.0014	-0.0014	-0.0014	-0.0012	-0.0009	-0.0011	-0.0011	0.9956	0.9983	0.9992	0.9995
250	-0.0005	-0.0004	-0.0004	-0.0004	-0.0003	-0.0002	-0.0002	-0.0002	0.9959	0.9980	0.9991	0.9995
400	-0.0003	-0.0002	-0.0002	-0.0002	-0.0002	-0.0001	-0.0001	-0.0002	0.9968	0.9982	0.9993	0.9995
$\delta_0 = 1$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	-0.0057	-0.0052	-0.0050	-0.0051	-0.0031	-0.0032	-0.0033	-0.0034	0.9995	0.9996	0.9999	0.9999
100	-0.0025	-0.0018	-0.0018	-0.0017	-0.0013	-0.0010	-0.0011	-0.0012	0.9996	0.9998	0.9999	0.9999
250	-0.0008	-0.0006	-0.0005	-0.0004	-0.0003	-0.0003	-0.0003	-0.0003	0.9998	0.9999	1.0000	1.0000
400	-0.0005	-0.0003	-0.0002	-0.0003	-0.0002	-0.0001	-0.0001	-0.0002	0.9999	0.9999	1.0000	1.0000

Table 2: **Memory Estimation Bias and Factor Estimate Consistency with** $r = 2$, $\xi_0 = 0$, **and** $\rho_i = 1$
This table reports the memory bias based on $\hat{f}_t$ (i.e. $\hat{\delta}_{\hat{f}} - \delta_0$), memory bias based on f_t (i.e. $\hat{\delta}_f - \delta_0$), and the average R^2 measure from the regression of f_t on $\hat{f}_t$ (i.e. $\bar{R}^2$) across $\delta_0 \in \{0, 0.3, 0.6, 1\}$, cross-section dimensions $N \in \{50, 100, 250, 400\}$, and time series lengths $T \in \{50, 100, 250, 400\}$. Standard error of the memory estimates are 0.1103, 0.0779, 0.0493 and 0.0389, respectively for $T = 50, 100, 250, 400$.

	$\hat{\delta}_{\hat{f}} - \delta_0$				$\hat{\delta}_f - \delta_0$				$\bar{R}^2$			
$\delta_0 = 0$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	0.0318	0.0225	0.0146	0.0123	0.0055	0.0054	0.0054	0.0054	0.9752	0.9893	0.9952	0.9967
100	0.0278	0.0212	0.0144	0.0111	0.0028	0.0028	0.0028	0.0028	0.9627	0.9772	0.9922	0.9942
250	0.0213	0.0170	0.0125	0.0102	0.0012	0.0012	0.0012	0.0011	0.9087	0.9366	0.9775	0.9886
400	0.0181	0.0149	0.0112	0.0095	0.0007	0.0007	0.0007	0.0007	0.8734	0.9317	0.9553	0.9756
$\delta_0 = 0.3$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	0.0102	0.0048	0.0009	-0.0003	-0.0028	-0.0028	-0.0030	-0.0031	0.9822	0.9898	0.9958	0.9967
100	0.0101	0.0062	0.0029	0.0014	-0.0013	-0.0009	-0.0011	-0.0011	0.9634	0.9829	0.9908	0.9943
250	0.0084	0.0056	0.0031	0.0021	-0.0003	-0.0002	-0.0003	-0.0003	0.9295	0.9561	0.9811	0.9882
400	0.0074	0.0051	0.0029	0.0020	-0.0002	-0.0001	-0.0001	-0.0002	0.8812	0.9432	0.9698	0.9782
$\delta_0 = 0.6$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	0.0006	-0.0010	-0.0019	-0.0023	-0.0028	-0.0030	-0.0030	-0.0032	0.9938	0.9964	0.9986	0.9987
100	0.0015	0.0005	-0.0003	-0.0005	-0.0012	-0.0009	-0.0011	-0.0011	0.9840	0.9939	0.9975	0.9977
250	0.0015	0.0008	0.0002	0.0000	-0.0003	-0.0002	-0.0002	-0.0002	0.9768	0.9863	0.9934	0.9965
400	0.0014	0.0007	0.0002	0.0000	-0.0002	-0.0001	-0.0001	-0.0002	0.9639	0.9817	0.9912	0.9954
$\delta_0 = 1$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	-0.0031	-0.0032	-0.0032	-0.0034	-0.0031	-0.0032	-0.0033	-0.0034	0.9987	0.9995	0.9998	0.9998
100	-0.0013	-0.0010	-0.0011	-0.0011	-0.0013	-0.0010	-0.0011	-0.0012	0.9989	0.9995	0.9998	0.9998
250	-0.0004	-0.0002	-0.0003	-0.0003	-0.0003	-0.0003	-0.0003	-0.0003	0.9991	0.9994	0.9998	0.9999
400	-0.0002	-0.0001	-0.0001	-0.0002	-0.0002	-0.0001	-0.0001	-0.0002	0.9989	0.9995	0.9998	0.9999

Table 3: **Memory Estimation Bias and Factor Estimate Consistency with** $r = 2$, $\xi_0 = 0.5$, **and** $\rho_i = 0.5$
This table reports the memory bias based on $\hat{f}_t$ (i.e. $\hat{\delta}_{\hat{f}} - \delta_0$), memory bias based on f_t (i.e. $\hat{\delta}_f - \delta_0$), and the average R^2 measure from the regression of f_t on $\hat{f}_t$ (i.e. $\bar{R}^2$) across $\delta_0 \in \{0, 0.3, 0.6, 1\}$, cross-section dimensions $N \in \{50, 100, 250, 400\}$, and time series lengths $T \in \{50, 100, 250, 400\}$. Standard error of the memory estimates are 0.1103, 0.0779, 0.0493 and 0.0389, respectively for $T = 50, 100, 250, 400$.

	$\hat{\delta}_{\hat{f}} - \delta_0$				$\hat{\delta}_f - \delta_0$				$\bar{R}^2$			
$\delta_0 = 0$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	0.0065	0.0066	0.0066	0.0066	0.0055	0.0054	0.0054	0.0054	0.9915	0.9963	0.9982	0.9987
100	0.0031	0.0032	0.0033	0.0033	0.0028	0.0028	0.0028	0.0028	0.9881	0.9944	0.9975	0.9985
250	0.0012	0.0013	0.0012	0.0013	0.0012	0.0012	0.0012	0.0011	0.9864	0.9933	0.9972	0.9982
400	0.0007	0.0008	0.0008	0.0008	0.0007	0.0007	0.0007	0.0007	0.9861	0.9924	0.9967	0.9979
$\delta_0 = 0.3$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	-0.0044	-0.0040	-0.0038	-0.0040	-0.0028	-0.0028	-0.0030	-0.0031	0.9960	0.9979	0.9990	0.9994
100	-0.0021	-0.0016	-0.0016	-0.0015	-0.0013	-0.0009	-0.0011	-0.0011	0.9951	0.9974	0.9991	0.9995
250	-0.0007	-0.0005	-0.0004	-0.0004	-0.0003	-0.0002	-0.0003	-0.0003	0.9931	0.9967	0.9987	0.9991
400	-0.0005	-0.0003	-0.0002	-0.0003	-0.0002	-0.0001	-0.0001	-0.0002	0.9934	0.9967	0.9987	0.9990
$\delta_0 = 0.6$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	-0.0062	-0.0056	-0.0053	-0.0054	-0.0028	-0.0030	-0.0030	-0.0032	0.9985	0.9993	0.9997	0.9998
100	-0.0028	-0.0020	-0.0019	-0.0018	-0.0012	-0.0009	-0.0011	-0.0011	0.9987	0.9995	0.9998	0.9999
250	-0.0010	-0.0006	-0.0005	-0.0005	-0.0003	-0.0002	-0.0002	-0.0002	0.9989	0.9994	0.9998	0.9998
400	-0.0006	-0.0004	-0.0003	-0.0003	-0.0002	-0.0001	-0.0001	-0.0002	0.9991	0.9995	0.9998	0.9999
$\delta_0 = 1$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	-0.0071	-0.0061	-0.0057	-0.0057	-0.0031	-0.0032	-0.0033	-0.0034	0.9998	0.9999	1.0000	1.0000
100	-0.0033	-0.0022	-0.0020	-0.0019	-0.0013	-0.0010	-0.0011	-0.0012	0.9999	0.9999	1.0000	1.0000
250	-0.0013	-0.0008	-0.0005	-0.0005	-0.0003	-0.0003	-0.0003	-0.0003	0.9999	1.0000	1.0000	1.0000
400	-0.0008	-0.0005	-0.0003	-0.0003	-0.0002	-0.0001	-0.0001	-0.0002	1.0000	1.0000	1.0000	1.0000

Table 4: **Memory Estimation Bias and Factor Estimate Consistency with** $r = 2$, $\xi_0 = 0.5$, **and** $\rho_i = 1$

This table reports the memory bias based on $\hat{f}_t$ (i.e. $\hat{\delta}_{\hat{f}} - \delta_0$), memory bias based on f_t (i.e. $\hat{\delta}_f - \delta_0$), and the average R^2 measure from the regression of f_t on $\hat{f}_t$ (i.e. $\bar{R}^2$) across $\delta_0 \in \{0, 0.3, 0.6, 1\}$, cross-section dimensions $N \in \{50, 100, 250, 400\}$, and time series lengths $T \in \{50, 100, 250, 400\}$. Standard error of the memory estimates are 0.1103, 0.0779, 0.0493 and 0.0389, respectively for $T = 50, 100, 250, 400$.

	$\hat{\delta}_{\hat{f}} - \delta_0$				$\hat{\delta}_f - \delta_0$				$\bar{R}^2$			
$\delta_0 = 0$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	0.0170	0.0120	0.0084	0.0078	0.0055	0.0054	0.0054	0.0054	0.9875	0.9934	0.9961	0.9978
100	0.0157	0.0114	0.0074	0.0058	0.0028	0.0028	0.0028	0.0028	0.9632	0.9819	0.9915	0.9947
250	0.0136	0.0101	0.0069	0.0053	0.0012	0.0012	0.0012	0.0011	0.9218	0.9526	0.9826	0.9893
400	0.0122	0.0095	0.0065	0.0052	0.0007	0.0007	0.0007	0.0007	0.8993	0.9420	0.9709	0.9858
$\delta_0 = 0.3$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	0.0014	-0.0006	-0.0018	-0.0023	-0.0028	-0.0028	-0.0030	-0.0031	0.9901	0.9958	0.9985	0.9988
100	0.0031	0.0014	0.0002	-0.0003	-0.0013	-0.0009	-0.0011	-0.0011	0.9838	0.9904	0.9962	0.9976
250	0.0035	0.0020	0.0009	0.0005	-0.0003	-0.0002	-0.0003	-0.0003	0.9604	0.9824	0.9912	0.9955
400	0.0033	0.0021	0.0010	0.0005	-0.0002	-0.0001	-0.0001	-0.0002	0.9407	0.9744	0.9837	0.9911
$\delta_0 = 0.6$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	-0.0026	-0.0029	-0.0029	-0.0032	-0.0028	-0.0030	-0.0030	-0.0032	0.9978	0.9987	0.9995	0.9996
100	-0.0009	-0.0007	-0.0010	-0.0010	-0.0012	-0.0009	-0.0011	-0.0011	0.9957	0.9979	0.9993	0.9994
250	-0.0001	-0.0001	-0.0002	-0.0002	-0.0003	-0.0002	-0.0002	-0.0002	0.9936	0.9956	0.9981	0.9991
400	0.0001	0.0000	-0.0001	-0.0002	-0.0002	-0.0001	-0.0001	-0.0002	0.9872	0.9949	0.9977	0.9984
$\delta_0 = 1$												
T/N	50	100	250	400	50	100	250	400	50	100	250	400
50	-0.0053	-0.0048	-0.0045	-0.0048	-0.0031	-0.0032	-0.0033	-0.0034	0.9996	0.9999	0.9999	0.9999
100	-0.0023	-0.0017	-0.0017	-0.0017	-0.0013	-0.0010	-0.0011	-0.0012	0.9997	0.9999	0.9999	1.0000
250	-0.0009	-0.0006	-0.0005	-0.0004	-0.0003	-0.0003	-0.0003	-0.0003	0.9997	0.9998	0.9999	1.0000
400	-0.0006	-0.0003	-0.0002	-0.0003	-0.0002	-0.0001	-0.0001	-0.0002	0.9997	0.9998	0.9999	1.0000

Table 5: MCS p-values for May 2017 - September 2021 out-of-sample period using the January 2000 - April 2017 in-sample period. Bold indicates the models included in $\widehat{\mathcal{M}}_{0.95}$, and h indicates the forecast horizon.

	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books	Hshld	Clths	Hlth	MedEq	Drugs	Chems
h=1:														
ARFIMA(1,d,0)	0.217	0.202	0.871	0.177	0.041	1.000	0.307	0.208	0.254	0.012	0.232	0.333	0.051	0.078
$\hat{f}_t$	0.217	**1.000**	**1.000**	**1.000**	**1.000**	0.037	0.567	0.562	**1.000**	**1.000**	0.539	**0.986**	**1.000**	0.678
$\hat{f}_t$ + ARFIMA(1,d,0)	**1.000**	0.567	0.518	0.431	0.364	0.037	**1.000**	0.383	0.340	0.878	**1.000**	0.876	0.179	0.930
C-S avg.	0.217	0.567	0.518	0.431	0.155	0.037	0.492	0.546	0.340	0.113	0.539	0.876	0.179	0.713
C-S avg. + ARFIMA(1,d,0)	0.169	0.566	0.136	0.373	0.786	0.037	0.246	**1.000**	0.323	0.180	0.539	**1.000**	0.159	**1.000**
h=4:														
ARFIMA(1,d,0)	0.146	0.408	0.150	0.117	0.185	0.134	0.092	0.233	0.280	0.183	0.424	0.124	**1.000**	0.183
$\hat{f}_t$	0.149	0.408	0.129	0.117	**1.000**	0.134	0.092	0.159	0.280	**1.000**	0.424	**1.000**	0.138	**1.000**
$\hat{f}_t$ + ARFIMA(1,d,0)	0.149	0.408	0.150	**1.000**	0.185	**1.000**	**1.000**	0.233	0.280	0.133	0.424	0.124	0.046	0.170
C-S avg.	**1.000**	**1.000**	**1.000**	0.117	0.185	0.134	0.092	**1.000**	**1.000**	0.143	**1.000**	0.124	0.138	0.183
C-S avg. + ARFIMA(1,d,0)	0.128	0.408	0.129	0.117	0.185	0.134	0.092	0.150	0.280	0.121	0.424	0.124	0.032	0.170

	Rubbr	Txtls	BldMt	Cnstr	Steel	FabPr	Mach	ElcEq	Autos	Aero	Ships	Guns	Gold	Mines
h=1:														
ARFIMA(1,d,0)	0.093	0.088	0.082	0.528	0.761	0.422	**1.000**	0.036	0.693	0.235	0.225	0.215	0.445	0.100
$\hat{f}_t$	0.413	0.857	**1.000**	**1.000**	**1.000**	0.208	0.893	**1.000**	0.262	0.235	**1.000**	**1.000**	0.000	0.279
$\hat{f}_t$ + ARFIMA(1,d,0)	0.397	**1.000**	0.153	0.763	0.650	0.080	0.658	0.162	0.693	0.299	0.643	0.417	0.000	0.613
C-S avg.	0.081	0.138	0.880	0.763	0.761	**1.000**	0.590	0.274	**1.000**	0.304	0.643	0.329	**1.000**	0.279
C-S avg. + ARFIMA(1,d,0)	**1.000**	0.138	0.105	0.627	0.229	0.422	0.590	0.070	0.135	**1.000**	0.599	0.303	0.035	**1.000**
h=4:														
ARFIMA(1,d,0)	0.188	**1.000**	0.177	0.426	0.223	0.198	0.111	0.297	**1.000**	0.277	0.467	0.192	0.200	0.220
$\hat{f}_t$	0.188	0.367	0.177	0.426	**1.000**	0.198	0.111	**1.000**	0.176	**1.000**	**1.000**	0.192	0.047	**1.000**
$\hat{f}_t$ + ARFIMA(1,d,0)	0.188	0.401	**1.000**	0.426	0.223	**1.000**	0.111	0.297	0.176	0.206	0.467	0.192	**1.000**	0.220
C-S avg.	**1.000**	0.245	0.177	1.000	0.223	0.198	**1.000**	0.297	0.176	0.277	0.467	**1.000**	0.200	0.220
C-S avg. + ARFIMA(1,d,0)	0.188	0.145	0.162	0.426	0.107	0.164	0.092	0.297	0.143	0.194	0.467	0.192	0.025	0.213

Table 5 continued from previous page

	Coal	Oil	Util	Telcm	PerSv	BusSv	Hardw	Softw	Chips	LabEq	Paper	Boxes	Trans	Whlsl
h=1:														
ARFIMA(1,d,0)	0.844	0.096	0.241	0.331	0.602	0.234	0.614	0.069	0.272	0.074	0.461	0.113	0.033	0.092
$\hat{f}_t$	0.306	0.151	**1.000**	0.128	**1.000**	**1.000**	0.569	0.511	0.387	0.754	0.791	0.494	**1.000**	0.493
$\hat{f}_t$ + ARFIMA(1,d,0)	0.045	0.096	0.273	0.128	0.999	0.787	0.492	0.468	0.272	0.372	**1.000**	0.295	0.393	**1.000**
C-S avg.	0.326	**1.000**	0.273	0.128	0.999	0.603	0.613	0.468	0.272	0.754	0.704	**1.000**	0.860	0.396
C-S avg. + ARFIMA(1,d,0)	**1.000**	0.151	0.273	**1.000**	0.989	0.562	**1.000**	**1.000**	**1.000**	**1.000**	0.700	0.471	0.324	0.395
h=4:														
ARFIMA(1,d,0)	0.240	**1.000**	0.206	0.053	0.276	0.110	0.128	**1.000**	**1.000**	**1.000**	0.065	0.192	0.244	0.250
$\hat{f}_t$	**1.000**	0.108	0.206	0.161	0.276	0.142	0.128	0.279	0.177	0.185	0.241	0.192	0.244	0.250
$\hat{f}_t$ + ARFIMA(1,d,0)	0.240	0.141	0.206	0.084	0.276	0.142	**1.000**	0.279	0.177	0.185	0.241	0.192	**1.000**	**1.000**
C-S avg.	0.240	0.108	**1.000**	**1.000**	**1.000**	**1.000**	0.128	0.279	0.177	0.185	**1.000**	**1.000**	0.244	0.250
C-S avg. + ARFIMA(1,d,0)	0.224	0.108	0.206	0.050	0.276	0.110	0.128	0.279	0.177	0.185	0.061	0.192	0.244	0.250

	Rtail	Meals	Banks	Insur	RlEst	Fin	Other
h=1:							
ARFIMA(1,d,0)	0.079	0.111	0.472	0.091	0.152	0.101	0.425
$\hat{f}_t$	0.677	**1.000**	0.584	0.104	**1.000**	0.441	**1.000**
$\hat{f}_t$ + ARFIMA(1,d,0)	0.381	0.113	0.584	0.091	0.320	**1.000**	0.420
C-S avg.	0.566	0.113	**1.000**	0.104	0.272	0.441	0.420
C-S avg. + ARFIMA(1,d,0)	**1.000**	0.113	0.452	**1.000**	0.272	0.083	0.272
h=4:							
ARFIMA(1,d,0)	0.141	**1.000**	0.108	0.078	0.232	0.066	**1.000**
$\hat{f}_t$	**1.000**	0.123	**1.000**	0.080	**1.000**	0.163	0.076
$\hat{f}_t$ + ARFIMA(1,d,0)	0.141	0.123	0.108	**1.000**	0.232	0.163	0.076
C-S avg.	0.141	0.123	0.108	0.080	0.232	**1.000**	0.076
C-S avg. + ARFIMA(1,d,0)	0.141	0.123	0.108	0.078	0.232	0.066	0.069

Research Papers

2021-10: Søren Johansen and Anders Ryghn Swensen: Adjustment coefficients and exact rational expectations in cointegrated vector autoregressive models

2021-11: Bent Jesper Christensen, Mads Markvart Kjær and Bezirgen Veliyev: The incremental information in the yield curve about future interest rate risk

2021-12: Mikkel Bennedsen, Asger Lunde, Neil Shephard and Almut E. D. Veraart: Inference and forecasting for continuous-time integer-valued trawl processes and their use in financial economics

2021-13: Anthony D. Hall, Annastiina Silvennoinen and Timo Teräsvirta: Four Australian Banks and the Multivariate Time-Varying Smooth Transition Correlation GARCH model

2021-14: Ulrich Hounyo and Kajal Lahiri: Estimating the Variance of a Combined Forecast: Bootstrap-Based Approach

2021-15: Salman Huseynov: Long and short memory in dynamic term structure models

2022-01: Jian Kang, Johan Stax Jakobsen, Annastiina Silvennoinen, Timo Teräsvirta and Glen Wade: A parsimonious test of constancy of a positive definite correlation matrix in a multivariate time-varying GARCH model

2022-02: Javier Haulde and Morten Ørregaard Nielsen: Fractional integration and cointegration

2022-03: Yue Xu: Spillovers of Senior Mutual Fund Managers' Capital Raising Ability

2022-04: Morten Ørregaard Nielsen, Wonk-ki Seo and Dakyung Seong: Inference on the dimension of the nonstationary subspace in functional time series

2022-05: Kristoffer Pons Bertelsen: The Prior Adaptive Group Lasso and the Factor Zoo

2022-06: Ole Linnemann Nielsen and Anders Merrild Posselt: Betting on mean reversion in the VIX? Evidence from ETP flows

2022-07: Javier Hualde and Morten Ørregaard Nielsen: Truncated sum-of-squares estimation of fractional time series models with generalized power law trend

2022-08: James MacKinnon and Morten Ørregaard Nielsen: Cluster-Robust Inference: A Guide to Empirical Practice

2022-09: Mikkel Bennedsen, Eric Hillebrand and Sebastian Jensen: A Neural Network Approach to the Environmental Kuznets Curve

2022-10: Yunus Emre Ergemen: Parametric Estimation of Long Memory in Factor Models