International Journal of Forecasting 39 (2023) 405–430

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Estimation of a dynamic multi-level factor model with possible long-range dependence

Yunus Emre Ergemen [a,b], C. Vladimir Rodríguez-Caballero [a,c,*]

[a] CREATES, Denmark
[b] Aarhus University, Department of Economics and Business Economics, Denmark
[c] Department of Statistics, ITAM, Mexico

ARTICLE INFO

Keywords:
Multi-level factors
Long-range dependence
Fractional cointegration
Nord Pool power market
Electricity price forecasting

ABSTRACT

A dynamic multi-level factor model with possible stochastic time trends is proposed. In the model, long-range dependence and short memory dynamics are allowed in global and local common factors as well as model innovations. Estimation of global and local common factors is performed on the prewhitened series, for which the prewhitening parameter is estimated semiparametrically from the cross-sectional and local average of the observable series. Employing canonical correlation analysis and a sequential least-squares algorithm on the prewhitened series, the resulting multi-level factor estimates have centered asymptotic normal distributions under certain rate conditions depending on the bandwidth and cross-section size. Asymptotic results for common components are also established. The selection of the number of global and local factors is discussed. The methodology is shown to lead to good small-sample performance via Monte Carlo simulations. The method is then applied to the Nord Pool electricity market for the analysis of price comovements among different regions within the power grid. The global factor is identified to be the system price, and fractional cointegration relationships are found between local prices and the system price, motivating a long-run equilibrium relationship. Two forecasting exercises are then discussed.

1. Introduction

Factor models are extensively used as a dimension-reduction tool in the analysis of large data sets and have been used in several applications. Kapetanios (2004) and Cristadoro et al. (2005), among others, used factor models to construct economic indicators. Meanwhile, Bernanke et al. (2005) and Favero et al. (2005) used factor models for policy purposes. On the other hand, some others have focused on the use of factors in forecasting; see Artis et al. (2005), Banerjee and Marcellino (2006), Stock and Watson (2006), Ludvigson and Ng (2007), and Giannone et al. (2008). Furthermore, factor models have recently been used in other fields, as in finance to analyze stochastic volatility (Cipollini & Kapetanios, 2008), market liquidity (Hallin et al., 2011), and market volatility (Barigozzi & Hallin, 2016), and in energy economics to analyze electricity prices Alonso et al. (2016), Dordonnat et al. (2012), and Ergemen et al. (2016). For estimation and inferential theory under different setups, see e.g. Forni et al. (2000), Stock and Watson (2002), Bai and Ng (2002) (Bai, 2003), Bai and Ng (2004), Forni et al. (2004, 2005), and Bai and Ng (2008). While there is a vast literature on estimation of the common factors and the number of common factors when both the cross-section and time-series dimensions are large, most available methodologies rely only on the existence of pervasive factors.

* Correspondence to: Río Hondo No.1, Col. Progreso Tizapán, Álvaro Obregón, CDMX, 01080, Mexico.
E-mail addresses: ergemen@econ.au.dk (Y.E. Ergemen), vladimir.rodriguez@itam.mx (C.V. Rodríguez-Caballero).

https://doi.org/10.1016/j.ijforecast.2021.12.004

More recently, there has been some interest in decomposing common factor structures into different levels. The intuition behind a multi-level factor structure is intrinsically related to Tobler's well-known first law of geography, "everything is related to everything else, but near things are more related than distant things", which is the foundation of many ideas embodied in spatial statistics. In this sense, a multi-level factor structure is based on a decomposition of the factor space into global and local components. While the global factors capture commonalities between all regions, the local components capture only those that are unique to specific regions. Standard (one-level) factor analysis is quite limited when there is also an interest in local rather than just global dynamics. Furthermore, ignoring such factors that affect only a subset of variables may result in highly correlated idiosyncratic components and thus poor, or possibly inconsistent, principal component (PC) estimates of the factors that affect all variables; see Wang (2010) and Choi et al. (2018).

In this paper, we propose a multi-level factor model with possible long-range dependence in the factors and idiosyncratic errors. The setup differs from conventional one-level factor setups in that the model implies lots of zero restrictions in the associated loadings matrix. Wang (2010) and Choi et al. (2018) consider multi-level factor structures under stationary $I(0)$ (non-integrated) setups for which identification is discussed and inferential theory is developed. On a related but different note, Moench et al. (2013) and Diebold et al. (2008), among others, consider $I(0)$ hierarchical multi-level factor structures in which each block of data is divided into sub-blocks to characterize within- and between-block variations. However, as discussed by Wang (2010), the interpretation of these factors is different. For example, Moench et al. (2013) consider a larger dimension for block-specific factors for each block than the dimension of pervasive factors, in contrast to the methodology proposed in this paper that does not impose such a restriction; see also (Wang, 2010), Breitung and Eickmeier (2016), and Choi et al. (2018). Our setup also differs from the existing multi-level factor models because we allow pervasive (or global) common factors, nonpervasive (local) common factors, and idiosyncratic components to exhibit long-range dependence and short memory dynamics without imposing $I(0)$ stationarity or $I(1)$ nonstationarity restrictions as traditionally imposed in the literature; see e.g. Choi et al. (2018). This way, the model has a great deal of flexibility, allowing for extensive cointegration analysis between the observed panel series and factors. Furthermore, nonstationary cases are nested smoothly, leading to asymptotically normally distributed estimates and thus chi-squared distributed test statistics. Also, from an empirical perspective, the necessity for allowing for long-range dependence has been well justified. For example, many economic and financial indicators, such as production, price, interest, and exchange rate series, may exhibit fractional long-range dependence; see e.g. Gil-Alana and Robinson (1997), Michelacci and Zaffaroni (2000), Bollerslev et al. (2013), and Pesaran and Chudik (2014). Our model is well suited for, but not limited to, the study of such indicators.

Estimation follows sequential steps. Given the allowance for general long-range dependence and short memory dynamics in both factors and idiosyncratic errors, we first prewhiten the observed series by obtaining a prewhitening parameter estimate from the cross-sectionally and locally averaged data, adopting the exact local Whittle method of Shimotsu and Phillips (2005). We establish the asymptotic behavior of the prewhitening parameter estimate under different cases. We then employ canonical correlation analysis (CCA) to obtain starting values for factors that are in a suitable vicinity of the global minima and then use a sequential least-squares algorithm, through which we apply an orthogonalization (or projection) procedure, to consistently estimate the spaces of prewhitened global and local factors separately. We establish asymptotic results for the prewhitened factors as well as their corresponding loadings and the common components. We also discuss how to determine the number of global and local factors, adopting the methods of Hallin and Liška (2011).

We assess the finite-sample performance of our methodology via Monte Carlo simulations and show that the method works well even in relatively small panels. We then apply the methodology to study the complex price dynamics of the Nord Pool power market in a large panel of hourly observations, for which the global and local factors drive the commonality overall and among bidding areas, respectively. We find that the global factor can be interpreted as the system price and that there are fractional cointegrating relationships between local prices and the system price, thus justifying a long-run equilibrium type of analysis. Finally, by two separate forecasting exercises, we show that pervasive and local factor estimates can be useful to predict system and regional prices.

The next section introduces the model along with the model assumptions, and contains the estimation details as well as the inferential theory. Section 3 discusses the selection of the number of global and local factors. Section 4 presents a finite-sample study based on Monte Carlo simulations. Section 5 provides an empirical application to the Nord Pool energy market. Section 6 presents two forecasting exercises, and finally, Section 7 concludes the paper.

Throughout the paper, $\|A\| = (trace(A'A))^{1/2}$ for a matrix A; $\mathbf{x}_n = O_p(\mathbf{y}_n)$ states that the vector of random variables, $\mathbf{x}_n$, is at most of order $\mathbf{y}_n$ in probability, and $\mathbf{x}_n = o_p(\mathbf{y}_n)$ is of smaller order in probability than $\mathbf{y}_n$; $\rightarrow_p$ denotes convergence in probability, and $\rightarrow_d$ denotes (pointwise) convergence in distribution; and e.g. $(N, T) \rightarrow \infty$ denotes joint dynamics in which both the cross-section size and time-series length are growing. All mathematical proofs and intermediate technical explanations are collected in an appendix at the end of the paper.

2. Two-level factor model with possible long-range dependence

2.1. Model

We consider a two-level factor model in that the unobserved common factors are classified into two groups:

the first group contains global factors, which are pervasive top-level factors that affect all sectors or regions; and the second group contains local or sector-specific factors, which are the nonpervasive sub-level factors and affect only a particular sector or region. In this paper, we use the terms "global factor" and "pervasive factor" interchangeably, and likewise the terms "local factor," "nonpervasive factor," and "sector-specific factor."

Let $y_{r,it}$ be the observation in region r of cross-section unit i at time t for $r = 1, \ldots, R$; $i = 1, \ldots, N_r$; and $t = 1, \ldots, T$ that is generated by

$$y_{r,it} = \gamma_{r,i}' G_t + \lambda_{r,i}' F_{r,t} + \epsilon_{r,it}. \tag{1}$$

In the model, the total number of observations across all regions is $N = N_1 + N_2 + \cdots + N_R$. We take the number of regions R to be fixed, since our methodology does not require $R \to \infty$ for consistent estimation and in practice R is almost always small. The $\mathbf{r}_G \times 1$ vector $G_t = \left(G_t^1, \ldots, G_t^{r_G}\right)'$ contains the $\mathbf{r}_G$ unobservable global factors, and the $\mathbf{r}_{F_r} \times 1$ vector $F_{r,t} = \left(F_{r,t}^1, \ldots, F_{r,t}^{r_{F_r}}\right)'$ consists of the $\mathbf{r}_{F_r}$ unobservable local factors in region r. Naturally, the number of local factors can be different in each region. $\gamma_{r,i}$ and $\lambda_{r,i}$ are $\mathbf{r}_G$- and $\mathbf{r}_{F_r}$-dimensional factor loadings, showing how each unit i in region r is affected by G_t and $F_{r,t}$, respectively.

The intuition behind a multi-level factor model is that each process $y_{r,it}$ is the sum of a global common component, a local common component, and an idiosyncratic component. Common components of region r are driven by the respective $\mathbf{r}_G$ and $\mathbf{r}_F$ vectors of common factors (global and local), which are possibly loaded differently. For example, there may be an interest in measuring certain comovements between countries employing multi-level factors. In that case, the global component would capture common movements in all groups of countries, and the local component would capture common movements with the country's neighbors, whereas the specific country component would capture movements that are unique to that specific country. Comovements between countries as captured by these multi-level factors can then be used to measure the connectivity of the countries analyzed. For instance, if the local component of a specific country weighs more than the global component, the country would seem to be more connected with its neighbors than with all the countries as a whole.

In (1), for $s = 1, \ldots, r_G$ and $j = 1, \ldots, r_{F_r}$,

$$\begin{aligned} G_t^s &= \Delta_t^{-\delta_0^s} w_t^s, \\ F_{r,t}^j &= \Delta_t^{-\vartheta_{r,0}^j} v_{r,t}^j, \text{ and} \\ \epsilon_{r,it} &= \Delta_t^{-d_{r,i0}} u_{r,it}, \end{aligned}$$

where w_t^s, $v_{r,t}^j$, and $u_{r,it}$ are stationary $I(0)$ processes (see Hendry (1995) for a definition) with spectral densities $f_{w^s}(\omega) \sim \Psi_w^s$, $f_{v_r^j}(\omega) \sim \Psi_{v_r}^j$ and $f_{u_{r,i}}(\omega) \sim \Psi_{u_{r,i}}$ when $\omega \sim 0$ for $r = 1, \ldots, R$ and $i = 1, \ldots, N_r$.

With $\Delta = 1 - L$, and L such that $L^k x_t = x_{t-k}$, $\Delta^{-\zeta}$ has the expansion

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

for $\zeta > 0$, with $\Gamma(\cdot)$ denoting the gamma function such that $\Gamma(\tau) = \infty$ for $\tau = 0, -1, \ldots$, and $\Gamma(0)/\Gamma(0) = 1$. $\Delta_t^{-\zeta}$ truncates this filter and introduces a type-II fractional process, $\Delta_t^{-\zeta} = \sum_{j=0}^{t} \pi_j(-\zeta) L^j$, and this truncation allows for the study of both asymptotically stationary ($\zeta < 1/2$) and asymptotically nonstationary ($\zeta \geq 1/2$) cases, unlike the untruncated filter $\Delta^{-\zeta}$, which does not converge when $\zeta \geq 1/2$; see Davidson and Hashimzade (2009).

We can write (1) for all R regions in the system form as

$$\begin{pmatrix} y_{1,\cdot t} \\ \vdots \\ y_{R,\cdot t} \end{pmatrix} = \begin{pmatrix} \Gamma_1 & \Lambda_1 & 0 & \cdots & 0 \\ \Gamma_2 & 0 & \Lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Gamma_R & 0 & 0 & \cdots & \Lambda_R \end{pmatrix} \times \begin{pmatrix} G_t \\ F_{1,t} \\ F_{2,t} \\ \vdots \\ F_{R,t} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,\cdot t} \\ \vdots \\ \epsilon_{R,\cdot t} \end{pmatrix},$$

where Γ_r and Λ_r denote the stacked versions of $\gamma_{r,i}$ and $\lambda_{r,i}$, respectively, over i. Then, further stacking over r, we can write

$$y_t = \Lambda^* F_t^* + \epsilon_t, \tag{2}$$

where $F_t^* = \left(G_t', F_{1,t}', \ldots, F_{R,t}'\right)'$ and $\Lambda^* = [\Gamma, \Lambda]$ with $\Gamma = (\Gamma_1', \ldots, \Gamma_R')'$ and $\Lambda = diag(\Lambda_1, \ldots, \Lambda_R)$. The entire system can also be written in matrix form, if we further stack over t, as

$$Y = F^* \Lambda^{*'} + E,$$

with the dimensions of Y, F^*, and Λ^* being $T \times N$, $T \times \left(\mathbf{r}_{F_1} + \cdots + \mathbf{r}_{F_R} + \mathbf{r}_G\right)$, and $N \times \left(\mathbf{r}_{F_1} + \cdots + \mathbf{r}_{F_R} + \mathbf{r}_G\right)$, respectively. From these representations, we can immediately see that there is a large number of zero restrictions and the number of factors can grow with the number of regions, in contrast to factor models that consider only pervasive factors.

We introduce the following conditions to study (1), letting $\mathcal{M}$ denote a generic positive constant.

Assumption A. Long-range dependence and short memory dynamics:

A_1 For $s = 1, \ldots, r_G$, $f_{w^s}(\omega) \sim \Psi_w^s \in (0, \infty)$ as $\omega \to 0+$. Furthermore, in a neighborhood $(0, \kappa)$ of the origin, $f_{w^s}(\omega)$ is differentiable, and $d/d\omega \log f_{w^s}(\omega) = O(\omega^{-1})$ as $\omega \to 0+$. These conditions also hold for $f_{v_r^j}(\omega)$ with $\Psi_{v_r}^j$ for $j = 1, \ldots, r_{F_r}$, and $f_{u_{r,i}}(\omega)$ with $\Psi_{u_{r,i}}$.

A_2 For $s = 1, \ldots, r_G$, $w_t^s = A(L) m_t^s = \sum_{k=0}^{\infty} a_k m_{t-k}^s$ with $\sum_{k=0}^{\infty} a_k^2 < \infty$, where $E(m_t^s|\mathcal{F}_{t-1}) = 0$, $E((m_t^s)^2|\mathcal{F}_{t-1}) = 1$, $E((m_t^s)^3|\mathcal{F}_{t-1}) = \mu_3$, $E((m_t^s)^4|\mathcal{F}_{t-1}) = \mu_4$, with finite constants μ_3, μ_4 almost surely, $t = 0, 1, \ldots$, in which $\mathcal{F}_t$ is the $\sigma-$field generated by m_l^s, $l \leq t$, and there exists a random variable m such that $E(m^2) < \infty$ and for all $\eta > 0$ and some $K > 0$, $Pr(\left|m_t^s\right| > \eta) \leq KPr(|m| > \eta)$. Similarly, for

$v^j_{r,t} = B_r(L)z^j_{r,t} = \sum_{k=0}^{\infty} b_{r,k} z^j_{r,t-k}$ with $\sum_{k=0}^{\infty} b^2_{r,k} < \infty$, $j = 1, \ldots, r_{F_r}$, and for $u_{r,it} = C_{r,i}(L)\varepsilon_{r,it} = \sum_{k=0}^{\infty} c_{r,ik}\varepsilon_{r,it-k}$ with $\sum_{k=0}^{\infty} c^2_{r,ik} < \infty$ for each r and i, the same conditions hold.

A_3 In a neighborhood $(0, \kappa)$ of the origin, $A(e^{i\omega})$, $B_r(e^{i\omega})$, and $C_{ri}(e^{i\omega})$, with i in the exponent s.t. $i^2 = -1$, are differentiable, $(d/d\omega)A(e^{i\omega}) = O(\omega^{-1})$, $(d/d\omega) B_r(e^{i\omega}) = O(\omega^{-1})$, and $(d/d\omega)C_{ri}(e^{i\omega}) = O(\omega^{-1})$ as $\omega \to 0+$.

A_4 Denoting m a bandwidth parameter, as $T \to \infty$, $m^{-1} + m(\log m)^{1/2}T^{-1} + m^{-\gamma}\log T \to 0$ for any $\gamma > 0$.

A_5 $\delta^s_0 \in \mathcal{G}^s = [\underline{\delta}^s, \bar{\delta}^s]$ and $\bar{\delta}^s - \underline{\delta}^s \leq 9/2$ for $s = 1, \ldots, r_G$. Also, for $r = 1, \ldots, R$, with R fixed, $\vartheta^j_{r,0} \in \mathcal{V}^j_r = [\underline{\vartheta}^j_r, \bar{\vartheta}^j_r]$ and $\bar{\vartheta}^j_r - \underline{\vartheta}^j_r \leq 9/2$ for $j = 1, \ldots, r_{F_r}$, and for $i = 1, \ldots, N_r$, $d_{r,i0} \in \mathcal{D}_{r,i} = [\underline{d}_{r,i}, \bar{d}_{r,i}]$ and $\bar{d}_{r,i} - \underline{d}_{r,i} \leq 9/2$.

Assumption A imposes the restrictions used by Shimotsu and Phillips (2005) for exact local Whittle estimation, which allows for being agnostic about the underlying short-run dynamics. The allowed range of memory values greatly relaxes the $I(0)$—$I(1)$ restrictions vastly imposed in the factor literature. The model in (1) simultaneously admits combinations of persistence levels in factors as well as in the idiosyncratic terms, while allowing for different integration orders within G_t and $F_{r,t}$. Hence, Assumption A permits extensive fractional cointegrating relationships in the model and can be useful in understanding the behavior of co-persistent indicators involved in the dynamics of a complex system. As a result, most factor models in the literature, such as those proposed by Stock and Watson (2002), Bai and Ng (2002, 2004), Forni et al. (2004, 2005), Wang (2010), and Breitung and Eickmeier (2016), are readily nested under (1). Furthermore, this allowance for general fractional dynamics contributes to the growing literature on fractional factor models; see e.g. Ray and Tsay (2000), Chen and Hurvich (2006), Morana (2007), Luciani and Veredas (2015), and Hartl and Weigand (2018). For further discussion on the conditions in Assumption A, readers are referred to Shimotsu and Phillips (2005).

Assumption B. Factors:

Denoting the $I(0)$ global and local factors as G^0_t and $F^0_{r,t}$, respectively, with $\delta^s_0 = \vartheta^j_{r,0} = 0$ for $s = 1, \ldots, r_G$ and $j = 1, \ldots, r_{F_r}$, define $H^0_{r,t} = \left(G^{0'}_t, F^{0'}_{r,t}\right)'$. $T^{-1}\sum_{t=1}^{T}\sum_{r=1}^{R} H^0_{r,t} H^{0'}_{r,t} \xrightarrow{p} \sum_H$ for some positive-definite matrix Σ_H as $T \to \infty$ with rank $\mathbf{r}_G + \mathbf{r}_{F_r}$.

Assumption B is a standard condition allowing for the fully whitened $I(0)$ factors to be stationary autoregressive processes in line with Assumption A_2. The rank condition states that different factors are not perfectly correlated.

Assumption C. Factor loadings:

C_1 $\lambda_{r,i}$ is either deterministic, such that $\|\lambda_{r,i}\| \leq \mathcal{M} < \infty$ and $\bar{\bar{\lambda}} = N^{-1}\sum_{r=1}^{R}\sum_{i=1}^{N_r}\lambda_{r,i} \neq 0$, or it is stochastic, such that $E\left(\lambda_{r,i}\right) \neq 0$ and $E\|\lambda_{r,i}\|^4 \leq \mathcal{M} < \infty$. In the latter case, $N_r^{-1}\Lambda_r'\Lambda_r \xrightarrow{p} \Sigma_{\Lambda_r}$ for an $\mathbf{r}_F \times \mathbf{r}_F$, with $\mathbf{r}_F = r_{F_1} + \cdots + r_{F_r}$, positive-definite non-random matrix Σ_{Λ_r}, as $N_r \to \infty$ for all $r = 1, \ldots, R$ with R fixed.

C_2 $\gamma_{r,i}$ is either deterministic, such that $\|\gamma_{r,i}\| \leq \mathcal{M}$ and $\bar{\bar{\gamma}} = N^{-1}\sum_{r=1}^{R}\sum_{i=1}^{N_r}\gamma_{r,i} \neq 0$, or it is stochastic, such that $E\left(\gamma_{r,i}\right) \neq 0$ and $E\|\gamma_{r,i}\|^4 \leq \mathcal{M} < \infty$ with $N^{-1}\mathbf{\Gamma}'\mathbf{\Gamma} \xrightarrow{p} \Sigma_{\mathbf{\Gamma}}$ for an $\mathbf{r}_G \times \mathbf{r}_G$ positive-definite non-random matrix $\Sigma_{\mathbf{\Gamma}}$, as $N_r \to \infty$ for all $r = 1, \ldots, R$ with R fixed.

C_3 Rank $([\Gamma\ \Lambda_r]) = \mathbf{r}_G + \mathbf{r}_{F_r}$.

Assumption C_1 ensures that the global factor G^s_t has a nontrivial contribution to the variance of y_t, $s = 1, \ldots, r_G$, while Assumption C_2 ensures that each local factor $F^j_{r,t}$ has a nontrivial contribution to the variance of $y_{r,t}$, $j = 1, \ldots, r_{F_r}$. The latter means that G^s_t pervades all variables, whereas the local factor $F^j_{r,t}$ pervades only within region r. The non-zero mean conditions imposed for global and local factor loadings ensure that the maximal memory in the panel can be estimated from the cross-sectional and local average of the data, which we use for prewhitening. The condition imposed on R being fixed can be removed, and $R \to \infty$ can be allowed, as in Wang (2010). In practice, however, R is always fixed (and small), and our estimation method does not require a growing R for the consistency of the estimates, unlike in the sequential principal-component method pursued by Wang (2010). The rank condition in Assumption C_3 guarantees enough heterogeneity among individual series within region r when responding to both factors. For estimation purposes, we assume the number of factors $\mathbf{r}_G$ and $\mathbf{r}_{F_r}$, $r = 1, \ldots, R$, to be known and fixed at this step, although we discuss this issue in Section 3. Formal tests or information criteria for the number of factors in a multi-level setup are, to our knowledge, not yet available, even discarding long-range dependence in the factors.

Assumption D. Model error innovations:

D_1 $E(u_{r,it}) = 0$ and $E(u^8_{r,it}) < \infty$.

D_2 $\left|E\left(N^{-1}\sum_{r=1}^{R}\sum_{i=1}^{N_r} u_{r,ik}u_{r,it}\right)\right| \leq \mathcal{M}$ for all t, and $T^{-1}\sum_{k=1}^{T}\sum_{t=1}^{T}\left|E\left(N^{-1}\sum_{r=1}^{R}\sum_{i=1}^{N_r} u_{r,ik}u_{r,it}\right)\right| \leq \mathcal{M}$.

D_3 $E\left(u_{r_1,it}u_{r_2,jt}\right) = \tau_{r_1r_2,ijt}$ with $\left|\tau_{r_1r_2,ijt}\right| \leq \tau_{r_1r_2,ij}$ for some $\tau_{r_1r_2,ij}$ and for all t. Furthermore, $N^{-1}\sum_{r_1=1}^{R}\sum_{r_2=1}^{R}\sum_{i=1}^{N_{r_1}}\sum_{i=1}^{N_{r_2}}\tau_{r_1r_2,ij} \leq \mathcal{M}$.

D_4 $E\left(u_{r_1,ik}u_{r_2,jt}\right) = \tau_{r_1r_2,ijkt}$ and $(NT)^{-1}\sum_{r_1=1}^{R}\sum_{r_2=1}^{R}\sum_{i=1}^{N_{r_1}}\sum_{i=1}^{N_{r_2}}\sum_{k=1}^{T}\sum_{t=1}^{T}\left|\tau_{r_1r_2,ijkt}\right| \leq \mathcal{M}$.

D_5 For every (k, t), $E\left|N^{-1/2}\sum_{r=1}^{R}\sum_{i=1}^{N_r}\left[u_{r,ik}u_{r,it} - E(u_{r,ik}u_{r,it})\right]\right|^4 \leq \mathcal{M}$.

Assumption D imposes conditions on the model error innovations that allow for weak time-series and cross-section dependence as well as heteroskedasticity in both time and cross-section dimensions. This way, the model has an approximate factor structure along the lines of Chamberlain and Rothschild (1983), but further allowing for heteroskedasticity in the time dimension.

Assumption E. Relationship between components:

Processes $\{u_{r,it}\}$, $\{v^j_{r,t}\}$, $\{w^s_t\}$, $\{\lambda_{r,i}\}$, and $\{\gamma_{r,i}\}$ are mutually independent groups.

Assumption E implies that unobservable factors, factor loadings, and error components are assumed to be independent of each other. This assumption could be relaxed to allow for weak dependence between factors and errors, as is done in the $I(0)$ factor literature, but this would not provide further insights. Under our setup, local factors from different regions can still be correlated, though they are uncorrelated with other model components.

Assumption F. Identification:

F_1 $F^{0'}_r F^0_r/T = I_{\mathbf{r}_{F_r}}$ and $\Lambda'_r\Lambda_r$ diagonal (within-region identification).
F_2 $G^{0'}G^0/T = I_{\mathbf{r}_G}$ and $\Gamma'\Gamma$ diagonal (between-region identification).
F_3 $E(G^0_t F^{0'}_{r,t}) = 0$ for $r = 1, \dots, R$.

We impose the three conditions on the fully whitened $I(0)$ factors in Assumption F for identification. Assumptions F1 and F2 are standard in factor analysis and are imposed to identify the model under such normalizations. Assumption F3 rules out any possibility of correlation between the $I(0)$ global and local factors, implying that the global factors do not contain information about local factors and vice versa. This assumption enables us to separately identify $I(0)$ local factors and global factors. Readers are referred to Wang (2010) for further discussion on the restrictions involved in both assumptions. Note that although the $I(0)$ global and local factors are assumed to be block orthogonal, the factors in (1) can be correlated after prewhitening, due to similarly evolving persistence characteristics as induced by long memory dynamics. However, it is still possible to orthogonalize the prewhitened factor estimates via regression, as we discuss in the next section.

Under these assumptions, following Breitung and Eickmeier (2016) and Proposition 1 in Wang (2010), the factor loadings corresponding to the $I(0)$ versions of the factors in (2) are identified up to a linear transformation of the loading matrix that preserves the same zero restrictions of the model given by $\Lambda^* Q$ with

$$Q = \begin{pmatrix} Q_{00} & 0 & 0 & \cdots & 0 \\ Q_{10} & Q_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Q_{R0} & 0 & 0 & \cdots & Q_{RR} \end{pmatrix}, \qquad (3)$$

where orthonormal global and local factors within each of the $R+1$ blocks are given by $Q_{00} = \left(T^{-1}\sum_{t=1}^T G^0_t G^{0'}_t\right)^{-1/2}$ and $Q_{rr} = \left(T^{-1}\sum_{t=1}^T F^0_{r,t}F^{0'}_{r,t}\right)^{-1/2}$ for all r. Matrix Q in (3) imposes that the R blocks of local factors are uncorrelated with the blocks of global factors.

2.2. Estimation

In the estimation, we follow a sequential approach. First, we semiparametrically estimate the integration order of the cross-sectionally and locally averaged data. We prewhiten the observable series by this prewhitening parameter estimate, following a similar reasoning to first differencing used by Bai and Ng (2004) but allowing for a wider range of persistence characteristics. We then estimate the prewhitened global and local factors based on CCA and sequential least squares, as also discussed by Breitung and Eickmeier (2016) and Choi et al. (2018) in the case of $I(0)$ multi-level factors. We also establish asymptotics for the factor loadings and common components. Once the unobservable prewhitened factors are estimated, they are integrated back by the initial prewhitening order so that the original factor estimates are obtained. Individual integration orders of these factor estimates can then also be determined by, e.g., the extended local Whittle (ELW) method proposed by Abadir et al. (2007), which consistently estimates the memory parameter, allowing for a wide range of values. We discuss these steps in detail as follows.

Memory estimation. To obtain an estimate of the prewhitening parameter from the cross-sectionally and locally averaged data, both parametric (see e.g. Ergemen and Velasco (2017)) and semiparametric (see e.g. Shimotsu and Phillips (2005)) methods can be employed. The advantage of semiparametric methods is that they can handle underlying short memory dynamics in an agnostic way and are robust to possible misspecification, which proves to be important under our setup, since both the multi-level factor structure and idiosyncratic errors are unobservable. On the other hand, a parametric method requires conditions reflecting prior knowledge on the model form and the short memory dynamics therein, so that, e.g., a conditional-sum-of-squares criterion can be correctly parametrized. With this in mind, we opt for semiparametric estimation, employing the exact local Whittle estimation due to Shimotsu and Phillips (2005).

Note that it is also possible to estimate the memory of each individual series and whiten each $y_{r,it}$ by its own estimated memory. In practice, this is generally feasible for several application-size data sets. However, our main aim in estimating the maximal memory, as provided by the averaged data, is to offer a general framework for cases in which it can be computationally costly, such as for population-wide register data sets providing information at an individual level.

To formalize the discussion, let us first denote

$$\delta_{max} = \max_s \delta^s_0, \quad \vartheta_{max} = \max_{j,r} \vartheta^j_{r,0} \quad \text{and} \quad d_{max} = \max_{i,r} d_{r,i0}.$$

Then, we denote the cross-sectional and local average of (1),

$$\bar{\bar{y}}_t := \frac{1}{N}\sum_{r=1}^R \sum_{i=1}^{N_r} y_{r,it} = \bar{\bar{\gamma}}' G_t + \bar{\bar{\lambda}}'\bar{F}_t + \bar{\bar{\epsilon}}_t,$$

where $\bar{F}_t = R^{-1}\sum_{r=1}^R F_{r,t}$ with R fixed and the quantities $\bar{\bar{\gamma}} \neq 0$ and $\bar{\bar{\lambda}} \neq 0$ under Assumption C do not depend on i or r. So $\bar{\bar{y}}_t$ is a pure time series that is integrated of order θ, such that

$$\theta = \max\{\delta_{max}, \vartheta_{max}, d_{max}\}$$

under Assumption C, or $\bar{\bar{y}}_t \sim I(\theta)$, so long as $\lambda_{r,i} \neq 0$ and $\gamma_{r,i} \neq 0$ for all i, r. Along this line, if $\lambda_{r,i}$ and/or $\gamma_{r,i}$ are zero for a small subset of indices, then this situation can be seen as a restricted version of (1) in which none of the remaining factor loadings are zero, so θ still captures the maximal available memory. For ease of exposition, we motivate the case in which $\lambda_{r,i} \neq 0$ and $\gamma_{r,i} \neq 0$ for all i, r.

Given that the model components are orthogonal to each other and R is fixed,

$$\bar{\bar{y}}_t = O_p\left(1 + \left(T^{\delta_{max}-1/2} + T^{\vartheta_{max}-1/2} + N^{-1/2}T^{d_{max}-1/2}\right)\log T\right),$$

as we show in Appendix 8.1. Therefore, the averaged error term is dominated by G_t and $F_{r,t}$ if $d_{max} \leq \min\{\delta_{max}, \vartheta_{max}\}$ or $\min\{\delta_{max}, \vartheta_{max}\} \leq d_{max} \leq \max\{\delta_{max}, \vartheta_{max}\}$, or if $N^{-1/2}T^{d_{max}-\min\{\delta_{max},\vartheta_{max}\}} \to 0$ when $d_{max} > \max\{\delta_{max}, \vartheta_{max}\}$, as $N \to \infty$. In such a case, an estimate of θ, say $\hat{\theta}$, estimates $\max\{\delta_{max}, \vartheta_{max}\}$ corresponding to the fractional cointegration case. Otherwise, if $N^{-1/2}T^{d_{max}-\min\{\delta_{max},\vartheta_{max}\}} \to \infty$ when $d_{max} > \min\{\delta_{max}, \vartheta_{max}\}$, as $N \to \infty$, $\hat{\theta}$ estimates d_{max}. In most empirical applications, the case in which global or regional factors are at least as persistent as the noise is the only relevant case, for which consistent estimation does not require further rate conditions on the relative growth rates of N and T. Therefore, we introduce the following condition before studying the asymptotic behavior of $\hat{\theta}$:

Assumption G. $\max\{\delta_{max}, \vartheta_{max}\} \geq d_{max}$.

The first condition basically states that there should be at least one factor whose persistence is higher than or equal to the maximal error memory; see also e.g. Bai and Ng (2004), who allow for $I(0)-I(1)$ possibilities for factors and errors.

We present the consistency of $\hat{\theta}$ in the following theorem, based on the exact local Whittle estimation results by Shimotsu and Phillips (2005).

Theorem 2.1. *Under Assumptions A, C, E, and G, as $T \to \infty$,*

$$\hat{\theta} = \max\{\delta_{max}, \vartheta_{max}\} + O_p(m^{-1/2}),$$

where m is a bandwidth for which conditions have been imposed in Assumption A_4.

This result shows that $\hat{\theta}$ constitutes a $\sqrt{m}$-consistent prewhitening parameter that can be used to difference the panel components to stationarity. The convergence rate in Theorem 2.1 depends on the bandwidth parameter m, which is usual in semiparametric memory estimation. It is typical in the semiparametric memory estimation literature to consider $m = T^{\kappa}$ with $1/2 < \kappa < 4/5$. Nevertheless, the optimal selection of m, based on the approximate mean squared error, can be done heuristically in the same manner as in Henry and Robinson (1996) or Andrews and Sun (2004).

Remark 1. Note that $\bar{\bar{y}}_t$ is a linear combination of fractional processes. When all integration orders are equal, $\bar{\bar{y}}_t$ can be considered as the DGP employed by Shimotsu and Phillips (2005) for memory estimation. In a more realistic scenario in which the integration orders are different, $\bar{\bar{y}}_t$ can be posited as a fractional process of order θ, and under Assumption A, the results apply for the exact local Whittle (ELW) estimator.

Remark 2. The cases in which deterministic components are incorporated into (1) pose challenges to the estimation of θ, requiring further assumptions on model components as well as different estimation techniques. Among others, Shimotsu (2010), Iacone (2010), McCloskey and Perron (2013), Arteche (2020), and Hualde and Nielsen (2020) study memory estimation in the presence of deterministic components. In this paper, we estimate θ in order to use it only as a prewhitening parameter below. Therefore, the extensions of (1), covering the impact of deterministic components for memory estimation, are beyond the scope of this paper.

Prewhitening. To detail the prewhitening procedure, let us introduce the notation for $\tau \geq 0$,

$$x_t(\tau) = \Delta_t^{\tau} x_t$$

to indicate (fractional) differencing of order τ.

In order to estimate the global and local factors, we prewhiten the series by $\hat{\theta}$ and write the prewhitened version of (1) as

$$y_{r,it}(\hat{\theta}) = \gamma_{r,i}' G_t(\hat{\theta}) + \lambda_{r,i}' F_{r,t}(\hat{\theta}) + \epsilon_{r,it}(\hat{\theta}), \tag{4}$$

which can be written in matrix notation based on (2) as

$$y_t(\hat{\theta}) = \Lambda^* F_t^*(\hat{\theta}) + \epsilon_t(\hat{\theta}).$$

Note that this type of prewhitening is similar to first differencing, as commonly used in nonstationary factor literature. Just like Bai and Ng (2004), we allow for the possibility that the prewhitened stationary series are serially correlated with (possibly infinite) autoregressive representations. However, our approach is more flexible in the sense that integration orders of factors and error terms can take a range of values beyond $I(0)$ or $I(1)$ cases. Along this line, it is also important to note that if there is prior knowledge on the $I(1)$ nonstationarity of the factors and $I(0)$ stationarity of the errors, for instance, then it is much more efficient to directly estimate the factor structure from the raw panel data (without prewhitening), possibly obtaining super-consistent estimates. However, to the best of our knowledge, there is no available method in the literature to obtain such prior information for multi-level factor models.

Factor estimation. Since under (4), the series are asymptotically stationary, we adopt the sequential least-squares (SLS) procedure that is proposed by Breitung and Eickmeier (2016) to estimate the factors. We outline the steps of this algorithm, in which the main goal is to minimize the residual sum of squares (RSS) function,

$$S\left(F_t^*(\hat{\theta}), \Lambda^*\right) = \sum_{t=1}^{T}\left(y_t(\hat{\theta}) - \Lambda^* F_t^*(\hat{\theta})\right)'\left(y_t(\hat{\theta}) - \Lambda^* F_t^*(\hat{\theta})\right)$$

$$= \sum_{r=1}^{R}\sum_{i=1}^{N_r}\sum_{t=1}^{T}\left(y_{r,it}(\hat\theta) - \gamma_{r,i}'G_t(\hat\theta) - \lambda_{r,i}'F_{r,t}(\hat\theta)\right)^2 \quad (5)$$

by a sequence of two least-squares regressions until RSS achieves a minimum. The algorithm is executed as follows:

1. The algorithm is initialized by using initial estimates of the global and local factors, $\hat G^{(0)}(\hat\theta) = \left(\hat G_1^{(0)}(\hat\theta), \ldots, \hat G_T^{(0)}(\hat\theta)\right)'$ and $\hat F_r^{(0)}(\hat\theta) = \left(\hat F_{r,1}^{(0)}(\hat\theta), \ldots, \hat F_{r,T}^{(0)}(\hat\theta)\right)'$, which are obtained by CCA.
2. Once initial estimators are obtained, the corresponding factor loadings at the initial step are estimated from the time-series regression $y_{r,it} = \gamma_{r,i}'\hat G_t^{(0)}(\hat\theta) + \lambda_{r,i}'F_{r,t}^{(0)}(\hat\theta) + \widetilde\epsilon_{r,it}(\hat\theta)$ that construct the factor loadings matrix, $\hat\Lambda^{*(0)}$, as in (2).
3. The global and local factors in the next step, $\hat G^{(1)}(\hat\theta)$ and $\hat F_{r,1}^{(1)}(\hat\theta)$, are updated from the least-squares regression of $y_t(\hat\theta)$ on $\hat\Lambda^{*(0)}$ to obtain $F_t^{*(1)}(\hat\theta) = \left(\hat\Lambda^{*(0)'}\hat\Lambda^{*(0)}\right)^{-1}\hat\Lambda^{*(0)'}y_t(\hat\theta)$.
4. Next, the updated factors $F_t^{*(1)}(\hat\theta)$ are used to get the associated factor loading matrix, $\hat\Lambda^{*(1)}$, as in step 2.
5. Steps 3 and 4 are repeated until the RSS converges to a minimum, from which $\hat F_t^{*}(\hat\theta)$ and $\hat\Lambda^{*}$ are collected.

Since the previous estimates are contained in the parameter space of the subsequent least-squares estimates, the RSS cannot increase at the next estimation step. Solving for the condition,

$$\hat\Lambda^{*'}Y'Y\left(I_N - \hat\Lambda^*\left(\hat\Lambda^{*'}\hat\Lambda^*\right)^{-1}\hat\Lambda^{*'}\right) = 0,$$

we can characterize the fixed points, since the RSS no longer decreases when the estimated factors and their loadings are orthogonal to the residuals of the previous step, which is equivalent to the first-order condition for PC estimation, but under a multi-level factor structure the loadings matrix has many zero entries. Furthermore, there exists a set of fixed points associated with the space spanned by $\hat\Lambda^*Q^{pre}$, where Q^{pre} is constructed with the $\hat\theta$-prewhitened factors in a similar way as in (3), and imposing the conditions in Assumption F, the fixed point is a unique minimum, arguing as in Breitung and Eickmeier (2016).

At the first step of the estimation, the initial values for the factors are obtained via CCA to ensure that the estimation procedure starts in a suitable vicinity of the global minimum, ensuring the quick convergence of the iterative algorithm; see also Breitung and Eickmeier (2016) and Choi et al. (2018). CCA is carried out in two steps. At the first step, in each region, $\mathbf{r} = \mathbf{r}_G + \mathbf{r}_{F_r}$ principal components are estimated to obtain the vector of factors $\hat H_{r,t}(\hat\theta)$ that constitutes a consistent estimator for the space of $\left(G_t(\hat\theta)', F_{r,t}(\hat\theta)'\right)'$. Taking the principal components from any two different regions, $G_t(\hat\theta)$ is the global component, for which we detail (the second step of) the procedure as follows. Let $\mathscr{H}_t(\hat\theta) = \left(\hat H_{r,t}(\hat\theta), \hat H_{s,t}(\hat\theta)\right)'$, with $c^0\mathscr{H}_t(\hat\theta)$ denoting the canonical variables. The CCA solves the following maximization problem:

$$\max\left\{c^0\Sigma_{01}c^1/\left[c^0\Sigma_{00}c^0\cdot c^1\Sigma_{11}c^1\right]^{1/2}\right\}$$

$$\text{s.t. } c^0\Sigma_{00}c^0 = 1, \text{ and } c^1\Sigma_{11}c^1 = 1,$$

where $\Sigma_{00} = \mathrm{Var}\left(\mathscr{H}_t(\hat\theta)\right)$, $\Sigma_{11} = \mathrm{Var}\left(\mathscr{H}_{t-1}(\hat\theta)\right)$ and, $\Sigma_{01} = \mathrm{Cov}\left(\mathscr{H}_t(\hat\theta), \mathscr{H}_{t-1}(\hat\theta)\right)$. The resulting linear combination with the largest canonical correlation is an estimate of the global factor, $\hat G^{(0)}(\hat\theta)$. Subsequently, we regress the original principal components of region r, $\hat H_{r,t}(\hat\theta)$, on the estimated global factors in order to obtain $\hat F_r^{(0)}(\hat\theta)$ for all $r = 1, \ldots, R$. See Choi et al. (2018) for a detailed treatment.

The vector of common components $\xi_t^*(\hat\theta) = \Lambda^*F_t^*(\hat\theta)$ is identified and consistently estimated without imposing a particular normalization. However, $\Lambda^{*(0)}$ and $F_t^{*(0)}(\hat\theta)$ are consistently estimated up to an arbitrary rotation, and Assumption F is invoked as follows to impose the normalization: first, the final estimators of $\hat F_{r,t}(\hat\theta)$, $r = 1, \ldots, R$ are regressed on $\hat G_t(\hat\theta)$, from which the obtained residuals provide the orthogonalized local factors; and second, the normalized global factors are obtained as the PCs of the estimated common components resulting from the nonzero eigenvalues and corresponding eigenvectors of $\hat\Gamma\left(T^{-1}\sum_{t=1}^{T}\hat G_t(\hat\theta)\hat G_t(\hat\theta)'\right)\hat\Gamma'$, which adopts the same normalization as in PC analysis. Normalized local factors are obtained similarly, from $\hat\Lambda_r\left(T^{-1}\sum_{t=1}^{T}\hat F_{r,t}(\hat\theta)\hat F_{r,t}(\hat\theta)'\right)\hat\Lambda_r'$, for $r = 1, \ldots, R$.

In order to analyze the asymptotic behavior of the factor estimates, we impose the following conditions on the vector of $I(0)$ global and local factors, G_t^0 and $F_{r,t}^0$, and $I(0)$ errors $\epsilon_{r,it}^0 = \Delta^{d_{r,i0}}\epsilon_{r,it}$.

Assumption H. Moments, central limit theorem, and distinct eigenvalues:

- H_1 For fixed t and r, $E\left\|\frac{1}{\sqrt{N_rT}}\sum_{i=1}^{N_r}\sum_{s=1}^{T}F_{r,t}^0\left[\epsilon_{r,is}^0\epsilon_{r,it}^0 - E(\epsilon_{r,is}^0\epsilon_{r,it}^0)\right]\right\|^2 \le \mathcal{M}$. $E\left\|\frac{1}{\sqrt{NT}}\sum_{r=1}^{R}\sum_{i=1}^{N_r}\sum_{s=1}^{T}G_t^0\left[\epsilon_{r,is}^0\epsilon_{r,it}^0 - E(\epsilon_{r,is}^0\epsilon_{r,it}^0)\right]\right\|^2 \le \mathcal{M}$.
- H_2 $E\left\|\frac{1}{\sqrt{N_rT}}\sum_{i=1}^{N_r}\sum_{t=1}^{T}F_{r,t}^0\lambda_{r,i}'\epsilon_{r,it}^0\right\|^2 \le \mathcal{M}$. $E\left\|\frac{1}{\sqrt{NT}}\sum_{r=1}^{R}\sum_{i=1}^{N_r}\sum_{t=1}^{T}G_t^0\gamma_{r,i}'\epsilon_{r,it}^0\right\|^2 \le \mathcal{M}$.
- H_3 For fixed t and r, as $N_r \to \infty$ and $N \to \infty$, $\frac{1}{\sqrt{N_r}}\sum_{i=1}^{N_r}\lambda_{r,i}\epsilon_{r,it}^0 \to_d N(0, \Sigma_{r,t})$, and $\frac{1}{\sqrt{N}}\sum_{r=1}^{R}\sum_{i=1}^{N_r}\gamma_{r,i}\epsilon_{r,it}^0 \to_d N(0, \Sigma_t)$, where $\Sigma_{r,t} = \lim_{N_r\to\infty}(1/N_r)\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\lambda_{r,i}\lambda_{r,j}'E(\epsilon_{r,it}^0\epsilon_{r,jt}^0)$ and $\Sigma_t = \lim_{N\to\infty}(1/N)\sum_{r=1}^{R}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\gamma_{r,i}\gamma_{r,j}'E(\epsilon_{r,it}^0\epsilon_{r,jt}^0)$.
- H_4 The eigenvalues of $\Sigma_{\Lambda_r}\Sigma_{F_r}$ are distinct. The eigenvalues of $\Sigma_\Gamma\Sigma_G$ are distinct.

Define rotation matrices L_{F_r} for $r = 1, \ldots, R$ as

$$L_{F_r} = (\Lambda_r'\Lambda_r/N)(F_r(\theta)'\hat{F}_r(\theta)/T)V_{RT}^{-1}$$

where V_{RT} is a diagonal matrix consisting of the first r_{F_r} largest eigenvalues of Y_rY_r'/N_rT, and L_G as

$$L_G = (\Gamma'\Gamma/N)(G(\theta)'\hat{G}(\theta)/T)V_G^{-1}$$

where V_G is a diagonal matrix consisting of the first r_G largest canonical correlations from the CCA step described above. Also, define

$$\Xi_G = plim_{T\to\infty}V_G^{-1}\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(\theta)G_t(\theta)', \quad \text{and}$$

$$\Xi_{F_r} = plim_{T\to\infty}V_{RT}^{-1}\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{r,t}(\theta)F_{r,t}(\theta)'$$

and redefine the variance–covariance matrices in Assumption H.3 to reflect the effect of prewhitening as Σ_t^* and $\Sigma_{r,t}^*$, where

$$\Sigma_t^* = \lim_{N\to\infty}(1/N)\sum_{r=1}^{R}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\gamma_{r,i}\gamma_{r,j}'E(\epsilon_{r,it}(\theta)\epsilon_{r,jt}(\theta))$$

$$\Sigma_{r,t}^* = \lim_{N_r\to\infty}(1/N_r)\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\lambda_{r,i}\lambda_{r,j}'E(\epsilon_{r,it}(\theta)\epsilon_{r,jt}(\theta)).$$

Then, in the next result, we establish the asymptotic behavior of the prewhitened factor estimates.

Theorem 2.2. *Under Assumptions A–H, as $(N_r, N, T) \to \infty$,*

- *if $N/m \to 0$,*

  $$\sqrt{N}(\hat{G}_t(\hat{\theta}) - L_G'G_t(\theta)) \to_d N(0, \Xi_G\Sigma_t^*\Xi_G')$$

  for fixed t;
- *if $N_r/m \to 0$,*

  $$\sqrt{N_r}(\hat{F}_{r,t}(\hat{\theta}) - L_{F_r}'F_{r,t}(\theta)) \to_d N(0, \Xi_{F_r}\Sigma_{r,t}^*\Xi_{F_r}')$$

  for fixed r and t.

These results show that the global and local factor estimates have asymptotic normal distributions and the convergence rates are $\sqrt{N}$ and $\sqrt{N_r}$, respectively, if $N/m \to 0$ and $N_r/m \to 0$ as $(N_r, N, T) \to \infty$. These rate requirements under Assumption A.4 imply for $m = T^\kappa$ that $0 < \kappa < 1$, so a larger time series may now be required than would be needed if the memory of the cross-sectionally averaged series were estimated based on a correctly parametrized criterion, which is difficult to obtain given the latency of model components. In practice, the user-chosen bandwidth plays a role on the minimum required time-series length (for a given cross-section size). Typically, it is preferable to impose the lower bound $\kappa > 1/2$ to avoid short memory contamination; see e.g. Henry and Robinson (1996). The components of variance–covariance matrices can be estimated in practice as

$$\hat{\Xi}_G = V_G^{-1}\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(\hat{\theta})\hat{G}_t(\hat{\theta})',$$

$$\hat{\Xi}_{F_r} = V_{RT}^{-1}\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{r,t}(\hat{\theta})\hat{F}_{r,t}(\hat{\theta})',$$

$$\hat{\Sigma}_t^* = (1/N)\sum_{r=1}^{R}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\hat{\gamma}_{r,i}\hat{\gamma}_{r,j}'E(\hat{\epsilon}_{r,it}(\hat{\theta})\hat{\epsilon}_{r,jt}(\hat{\theta}))$$

$$\hat{\Sigma}_{r,t}^* = (1/N_r)\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\hat{\lambda}_{r,i}\hat{\lambda}_{r,j}'E(\hat{\epsilon}_{r,it}(\hat{\theta})\hat{\epsilon}_{r,jt}(\hat{\theta}))$$

where $\hat{\Xi}_G$ and $\hat{\Xi}_{F_r}$ simplify with the normalizations $\hat{G}(\hat{\theta})'\hat{G}(\hat{\theta})/T = I_{r_G}$ and $\hat{F}_r(\hat{\theta})'\hat{F}_r(\hat{\theta})/T = I_{r_{F_r}}$, while V_G and V_{RT} are both $O_p(1)$, due to stationarity and following the result in Lemma A.3 of Bai (2003), with their probability limits being the eigenvalues of $\Sigma_\Gamma\Sigma_G$ and $\Sigma_{\Lambda_r}\Sigma_{F_r}$ satisfying Assumption H.4, and

$$\hat{\epsilon}_{r,it}(\hat{\theta}) = y_{r,it} - \hat{\gamma}_{r,i}'\hat{G}_t(\hat{\theta}) + \hat{\lambda}_{r,i}'\hat{F}_{r,t}(\hat{\theta}) \quad (6)$$

where the factor loading estimates $\hat{\gamma}_{r,i}$ and $\hat{\lambda}_{r,i}$ are obtained by least squares as $\hat{\Lambda}^*(\hat{\theta}) = y_t(\hat{\theta})'\hat{F}_t^*(\hat{\theta})$, for which establishing consistency follows least-squares arguments. The consistency of the remaining components can be established similarly, by first using a mean-value argument to account for the estimation effect of the prewhitening parameter, and then using the time-domain alternatives of the results from (Robinson & Hidalgo, 1997) and Robinson (2005).

More than just the factor estimates themselves, there can be interest in the overall common components, especially in a forecasting context. Therefore, we show convergence rate results for the prewhitened common component estimates, $\hat{C}_G(\hat{\theta}) = \hat{\gamma}_{r,i}'\hat{G}_t(\hat{\theta})$ and $\hat{C}_F(\hat{\theta}) = \hat{\lambda}_{r,i}'\hat{F}_t(\hat{\theta})$.

Theorem 2.3. *Under Assumptions A–H, as $(N_r, N, T) \to \infty$,*

$$\hat{C}_G(\hat{\theta}) - C_G(\theta) = O_p\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T}\}}\right) + o_p(m^{-1/2}),$$

$$\hat{C}_F(\hat{\theta}) - C_F(\theta) = O_p\left(\frac{1}{\min\{\sqrt{N_r}, \sqrt{T}\}}\right) + o_p(m^{-1/2}).$$

These results are similar to the standard ones in the literature (see e.g. Choi et al. (2018)), except that the prewhitening parameter estimation effect is made explicit in the convergence rate. For this result, there is no rate requirement between N_r and T.

It should be remarked here that for these results, we maintained the assumption that R is fixed. The case in which $R \to \infty$, though not of practical interest, may be of independent theoretical interest. In such a case, showing asymptotic results is facilitated by the fact that the persistence of global factors becomes more dominant at the prewhitening step in general, and the sequential PC method of Wang (2010) can be adopted under our Assumptions A–H, such that our results still hold under extra rate conditions that involve R.

Recovering the original factors. From the estimates of prewhitened global and local factors obtained under Theorem 2.2, the original factor estimates can be recovered by integrating back by $\hat{\theta}$, as

$$\Delta^{-\hat{\theta}}\hat{G}_t(\hat{\theta}) = \hat{G}_t \quad \text{and} \quad \Delta^{-\hat{\theta}}\hat{F}_{r,t}(\hat{\theta}) = \hat{F}_{r,t},$$

omitting dependence on $\hat{\theta}$ and assuming away the initial conditions that are negligible under joint (N_r, N, T) asymptotics, following a similar discussion to Ergemen and Velasco (2017). Using these original factor estimates, true integration orders of the global and local factors can be estimated either parametrically, based on a conditional-sum-of-squares (CSS) criterion (see e.g. Ergemen and Velasco (2017)), or semiparametrically (see e.g. Abadir et al. (2007)). Other time-varying model components can be treated the same way, and their integration orders can be estimated.

The memory of prewhitened residuals can be estimated from (6) by resorting to CSS or Whittle methods, and the original residual memory parameters can be obtained after integrating back by $\hat{\theta}$.

Last but not least, uniform consistency results for the original global and local factors are difficult to justify with additional long memory dynamics requiring rate conditions on N and T that depend on unknown memory parameters. However, it should be made clear that such extra rate conditions leading to uniform consistency are not necessary for our results to hold, and such an investigation is left for future research.

3. Determining the number of prewhitened local and global factors

In the previous section, we assume the number of prewhitened local and global factors to be fixed and known, in a similar way to existing literature, in which the number of local and global factors is fixed to be one, or some alternative models considering more factors are analyzed without using formal information criteria.

Although there are many methodologies to estimate the number of static factors in one-level factor models—see e.g. Bai and Ng (2002), Alessi et al. (2010), Onatski (2010), Kapetanios (2010), and Ahn and Horenstein (2013) —a formal methodology to estimate the number of factors in a multi-level factor model is not yet available, to the best of our knowledge. The closest exception is the proposal of Hallin and Liška (2011), who allow for identifying and estimating joint and block-specific common factors in the context of dynamic factor models, providing a rigorous theoretical treatment.

We adopt the methodology proposed by Hallin and Liška (2011) to identify the number of local and global factors under our setup. We retain Assumption A–H imposed to study the model in (1), and further assume (Hallin & Liška, 2011)'s Assumptions A1' and B, under which a consistent lag-window estimator for spectral density based on the sample covariance matrix is obtained with the window size, M_T, increasing with but slower than the time-series length, T. In this case, the information criterion relying on the estimated eigenvalues of the sample spectral density matrices, $\psi(\varpi_l)$, with the frequencies $\varpi_l := \pi l/(M_T + 1/2)$,

$$IC(k) := \log\left(\frac{1}{n}\sum_{i=k+1}^{n}\frac{1}{2M_T+1}\sum_{l=-M_T}^{M_T}\psi(\varpi_l)\right) + k\,c\,p(N,T), \quad c > 0, \tag{7}$$

produces consistent estimates for the number of factors under (Hallin & Liška, 2011)'s Assumptions A1'', A2', A3', and B, which we also assume to hold, considering that the prewhitened data are stationary, as we show in the proof of Theorem 2.2, given a k_{max} such that $0 \leq k \leq k_{max}$ and if $p(N,T)$ is $o(1)$ and $p^{-1}(N,T)$ is $o\left(\min\{N, M_T^2, M_T^{-1/2}T^{1/2}\}\right)$. The intuitive explanation behind the information criteria in Eq. (7) is that since the cross-sectional average of the idiosyncratic component is a function of k (the number of estimated factors), the penalty function $p(N,T)$ is used to avoid under-fitting or over-fitting, as with the AIC or BIC in time-series analysis. Examples of penalty functions are PC1–PC3 and IC1–IC3 in Bai and Ng (2002), where criteria IC1 and IC2 are more often used in empirical applications. Alessi et al. (2010) also propose two modified information criteria to improve the estimation of the number of factors in the presence of large idiosyncratic disturbances. Further inspection of Eq. (7) indicates that the estimated number of factors is also a function of a tuning multiplicative constant defined by c, introduced by Hallin and Liška (2007). This constant should be determined when implementing the methodology in practice. See Proposition 7 and a detailed discussion of the optimal tuning of c in Hallin and Liška (2007).

After implementing this procedure distinctly for each region, we use the inclusion-exclusion principle from set theory to determine the number of global and local factors. To motivate the idea, simply consider only two regions or blocks (B_x, B_y) in (1), say with only one local factor in each region and one global factor. We can now divide our data into three different factor spaces. Call the marginal factor spaces as those two different spaces spanned by the individual blocks of data B_x and B_y, and call the joint factor space as that spanned by the complete block $B_{x\cup y}$. In these spaces, we see that $s_{B_x} = 2$, $s_{B_y} = 2$, and $s_{Bx\cup y} = 3$, given that we have only one local factor in each region and only one global factor. The latter means that both marginal factor spaces consist of two static factors, whereas the joint factor space consists of three static factors. The number of factors in each one of these three factor spaces is consistently estimated by using the information criteria in Hallin and Liška (2011) after fractional differencing by $\hat{\theta}$, as described above.

The simple Venn diagram in Fig. 1 displays the strategy discussed above for the case of two blocks. The green sector represents the part of the factor space that is shared by both regions, and consists of one factor (the global factor). The marginal factor space B_x is represented by blue + green sectors having two factors, whereas the marginal factor space B_y is the yellow + green sectors and also has two factors. Naturally, the number of local factors is directly obtained after computing the number of global factor by the inclusion-exclusion principle, i.e. $s_{B_x\cup B_y} = s_{B_x} + s_{B_y} - s_{B_x\cap B_y}$, from which we get $s_{B_x\cap B_y} = 1$ (the global factor).

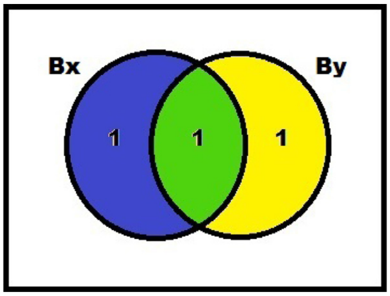

Fig. 1. Representation of the three factor spaces spanned by two regions or blocks of data.

The complexity of this methodology increases in the number of regions. Clearly, when we have R regions, the number of blocks to be analyzed will be the power set minus one, $2^R - 1$. Furthermore, we should compute the number of global factors by using each one of the number of factors (cardinalities) estimated in the individual, pairwise, triple-wise, etc. sets by the inclusion-exclusion principle. The number of local factors in region R would be determined by subtracting the number of factors previously estimated in each one of the intersections, where region R interacts with the number of factors previously estimated only in region R.

As an example, consider three regions now, such that we have the following blocks: B_x, B_y, B_z, $B_{x\cup y}$, $B_{x\cup z}$, $B_{y\cup z}$, and $B_{x\cup y\cup z}$. After fractionally differencing all the variables in the data set, we compute the number of factors that span each of the seven blocks. The number of factors in pairwise blocks will be given by $s_{B_{x\cap y}} = s_{B_x} + s_{B_y} - s_{B_{x\cup y}}$, for instance. The global factor will be given by $s_{B_{x\cap y\cap z}} = s_{B_{x\cup y\cup z}} - s_{B_x} - s_{B_y} - s_{B_z} + s_{B_{x\cap y}} + s_{B_{x\cap z}} + s_{B_{y\cap z}}$, and the number of local factors of region x by $s_{B_x} - s_{B_{x\cap y\cap z}} - s_{B_{x\cap y}} - s_{B_{x\cap z}}$, for instance.

Naturally, with this methodology, it is possible to specify not only the number of factors corresponding to the global and local levels but also the number of factors in each one of the three pairwise blocks of regions. This becomes especially useful in empirical work when the dependence between regions is of interest.

4. Simulations

In this section, we explore the finite-sample properties of the estimation procedure to investigate the performance of the model in (1) and the methodology proposed to estimate the number of global and local factors.[1]

4.1. Two-level factor model

In some cases, we generate a fractional cointegration relationship between $y_{r,it}$ and the global factor, G_t, since we believe such a relationship can be realistic in several empirical studies.

In these Monte Carlo studies, for which the results are presented in Tables 1–4, we analyze the performance of our model with $R = 2$ and $R = 4$, $N_r \in \{20, 80\}$, and different sample sizes with $T \in \{150, 1000, 5000\}$, respectively. One global factor and one local factor in each region are considered for simplicity. The global factor, both local factors, and all idiosyncratic terms are independently generated by ARFIMA(1, d_m,0) processes where d_m corresponds to δ, ϑ_r, or $d_{i,r}$, as appropriate.[2] Autoregressive parameters are taken as 0.5 for the unobservable factors and 0.1 for the idiosyncratic errors, following (Breitung & Eickmeier, 2016). In the first three experiments, we consider cross-sectionally independent idiosyncratic components; $u_{r,it} \overset{iid}{\sim} N(0, 2\phi)$ with ϕ controlling the signal-to-noise ratio with $\phi \in \{5, 2, 0.5\}$, corresponding to low, medium, and high signal-to-noise ratios. $w_t \overset{iid}{\sim} N(0, \sigma_w)$ and $v_{r,t} \overset{iid}{\sim} N\left(0, \sigma_{v_r}\right)$ control the ratio $\frac{\sigma_{v_r}}{\sigma_w}$ to study the relative impact of the factors on each other. Furthermore, all factor loadings are generated as $N(1, 1)$, following (Boivin & Ng, 2006). All results are based on 1000 replications of the model.

In each experiment, as explained above, we first estimate the fractional memory of $\bar{\bar{y}}_t$, denoted as $\hat{\theta}$. Second, the series are $\hat{\theta}$-fractionally differenced before estimating the model. Then, after collecting the estimated local and global factors, we integrate back the factors and estimate the memory parameters $\hat{\vartheta}_r$ and $\hat{\delta}$ using the extended local Whittle (ELW) procedure of Abadir et al. (2007), which consistently estimates fractionally integrated $I(d)$ processes for $d \in (-3/2, \infty)$, covering stationary and nonstationary regions. The number of Fourier frequencies used in this Monte Carlo study is $m = \lfloor T^{0.70} \rfloor$, with $T \in \{150, 1000, 5000\}$ corresponding to $m \in \{33, 125, 388\}$, respectively. Although there is no a general way to choose an optimal bandwidth for Whittle estimators, a bandwidth parameter that is too large is generally avoided. For instance, the finite-sample study of Abadir et al. (2007) suggests ruling out cases with $m = \lfloor T^{0.8} \rfloor$ for the estimation of d.

We also regress the actual factors (global or local) on the estimated ones in order to study the reliability of the procedure by computing the coefficient of determinations of the global factor and the average of the local factors, denoted as R^2_G and $\overline{R^2_R}$, respectively. Both coefficients can be considered a measure of consistency for all t; see Bai (2003). Finally, $\hat{d}_{R_{CSS}}$ and $\hat{d}_{R_{ELW}}$ denote the average of the estimated residual integration orders in the region r by CSS and ELW, respectively.

As seen from Tables 1–3, biases from memory estimates of the global and local factors are small, indicating that the fractional memory parameters of both levels of factors can be obtained accurately. In the same way, the fractional memory of the residuals is accurately estimated across all the sample sizes and the persistence levels allowed.

[1] We are deeply grateful to Sandra Eickmeier for sharing her Matlab code that helped us make our computations more efficient.

[2] Liu et al. (2017) provides a comprehensive survey and evaluation of ARFIMA codes available in different software platforms.

Table 1
$T = 150$, $N_r \in (20, 80)$, and $R = 2$. The fractional memory estimates of $\bar{\bar{y}}_t$ and $\hat{\theta}$, averages biases of the memory estimated in both levels of unobservable factors and in the residual, and the measure of the consistency of the unobservable factors estimated are presented in the report.

			$N_r = 20$							$N_r = 80$						
				BIAS				BIAS			BIAS				BIAS	
ϕ	σ_{v_r}	σ_w	$\hat{\theta}$	$\overline{\hat{\vartheta}_R}$	$\hat{\delta}$	$\overline{R^2_R}$	R^2_G	$\bar{\hat{d}}_{R_{CSS}}$	$\bar{\hat{d}}_{R_{ELW}}$	$\hat{\theta}$	$\overline{\hat{\vartheta}_R}$	$\hat{\delta}$	$\overline{R^2_R}$	R^2_G	$\bar{\hat{d}}_{R_{CSS}}$	$\bar{\hat{d}}_{R_{ELW}}$
			$\vartheta_{r0} = 0.2$	$\delta_0 = 0.4$		$d_{r,i0} = 0$										
0.5	2	1	0.364	0.016	0.021	0.984	0.986	−0.008	−0.002	0.361	0.019	0.037	0.990	0.997	−0.006	−0.003
0.5	1	1	0.322	0.072	0.023	0.962	0.985	−0.008	−0.004	0.354	0.020	0.007	0.984	0.997	−0.008	−0.004
0.5	1	2	0.373	0.018	0.014	0.959	0.996	−0.007	−0.003	0.372	0.009	0.013	0.984	0.999	−0.007	−0.003
2	2	1	0.372	0.007	0.039	0.959	0.943	−0.006	−0.001	0.382	0.010	0.022	0.984	0.986	−0.006	−0.003
2	1	1	0.337	0.028	0.023	0.878	0.944	−0.007	−0.001	0.350	0.013	0.021	0.963	0.986	−0.007	−0.004
2	1	2	0.375	0.032	0.008	0.876	0.985	−0.010	−0.006	0.377	0.023	0.012	0.961	0.997	−0.007	−0.003
5	2	1	0.368	0.055	0.026	0.736	0.965	−0.010	−0.006	0.380	0.036	0.011	0.917	0.991	−0.008	−0.004
5	1	1	0.353	0.059	0.032	0.737	0.964	−0.008	−0.004	0.365	0.034	0.024	0.919	0.991	−0.007	−0.002
5	1	2	0.368	0.051	0.020	0.743	0.965	−0.009	−0.006	0.364	0.042	0.021	0.921	0.991	−0.006	−0.003
			$\vartheta_{r0} = 0.4$	$\delta_0 = 0.8$		$d_{r,i0} = 0.2$										
0.5	2	1	0.790	0.013	0.045	0.963	0.996	−0.034	−0.026	0.788	0.015	0.036	0.966	0.999	−0.033	−0.026
0.5	1	1	0.797	0.007	0.042	0.938	0.996	−0.033	−0.024	0.794	0.002	0.035	0.959	0.999	−0.035	−0.027
0.5	1	2	0.739	0.001	0.039	0.945	0.999	−0.033	−0.025	0.758	0.002	0.017	0.958	1.000	−0.033	−0.027
2	2	1	0.804	−0.001	0.036	0.936	0.984	−0.040	−0.034	0.805	−0.002	0.045	0.960	0.996	−0.032	−0.025
2	1	1	0.793	0.017	0.049	0.864	0.984	−0.037	−0.030	0.792	0.010	0.047	0.942	0.996	−0.034	−0.025
2	1	2	0.741	0.029	0.035	0.866	0.996	−0.033	−0.027	0.757	−0.003	0.025	0.941	0.999	−0.031	−0.025
5	2	1	0.743	0.037	0.039	0.718	0.989	−0.037	−0.028	0.750	0.024	0.035	0.896	0.998	−0.036	−0.027
5	1	1	0.726	0.033	0.052	0.734	0.988	−0.034	−0.027	0.733	0.025	0.051	0.904	0.998	−0.036	−0.027
5	1	2	0.733	0.032	0.046	0.740	0.988	−0.038	−0.030	0.743	0.017	0.039	0.906	0.998	−0.035	−0.028
			$\vartheta_{r0} = 0.6$	$\delta_0 = 0.8$		$d_{r,i0} = 0.4$										
0.5	2	1	0.764	0.025	0.042	0.979	0.995	−0.007	−0.002	0.754	0.023	0.046	0.981	0.999	−0.007	−0.002
0.5	1	1	0.765	0.027	0.021	0.960	0.995	−0.008	−0.003	0.762	0.023	0.017	0.978	0.999	−0.007	−0.002
0.5	1	2	0.746	0.036	0.036	0.960	0.999	−0.007	−0.001	0.749	0.016	0.031	0.976	1.000	−0.006	−0.002
2	2	1	0.746	0.033	0.058	0.961	0.978	−0.009	−0.003	0.758	0.028	0.042	0.977	0.995	−0.007	−0.002
2	1	1	0.714	0.032	0.049	0.900	0.979	−0.009	−0.003	0.723	0.032	0.041	0.963	0.995	−0.007	−0.002
2	1	2	0.749	0.039	0.041	0.904	0.995	−0.008	−0.003	0.754	0.028	0.036	0.965	0.999	−0.006	−0.001
5	2	1	0.760	0.061	0.030	0.796	0.988	−0.010	−0.005	0.754	0.047	0.037	0.936	0.997	−0.007	−0.003
5	1	1	0.740	0.062	0.046	0.796	0.985	−0.009	−0.002	0.751	0.045	0.039	0.934	0.997	−0.007	−0.001
5	1	2	0.748	0.067	0.037	0.798	0.986	−0.009	−0.003	0.756	0.045	0.026	0.937	0.997	−0.007	−0.004

Notes: The DGP is $y_{r,it} = \gamma'_{r,i}G_t + \lambda'_{r,i}F_{r,t} + \Delta_t^{-d_{r,i0}}\epsilon_{r,it}$. $r = 1, 2$ and $i \in (20, 80)$ and $T = 150$. $\epsilon_{r,it} \overset{iid}{\sim} N(0, 2\phi)$ are generated independently with ϕ controlling the signal-to-noise ratio with $\phi = \{5, 2, 0.5\}$. Only one top-level factor and only one block-specific factor in each block are considered. $G_t = 0.5G_{t-1} + \Delta_t^{-\delta}w_t$, with $w_t^G \sim IIDN(0, \sigma_w)$ and $\sigma_w \in (1, 2)$. $F_{r,t} = 0.5F_{r,t-1} + \Delta_t^{-\vartheta_r}v_t$, with $v_t \sim IIDN(0, \sigma_v)$ and $\sigma_v \in (1, 2)$. $\overline{\hat{\vartheta}_R}$ is the average bias of the estimated memory of both local factors. $\hat{\delta}$ is the average bias of the estimated memory of the global factor. These memories are estimated with ELW with a bandwidth parameter of $m = \lfloor T^{0.70} \rfloor$ corresponding to $m = 33$. R^2_G is the R^2 of a regression of actual on estimates of the global factor, and $\overline{R^2_R}$ is the average of such R^2 on local factors. $\bar{\hat{d}}_{CSS}$ and $\bar{\hat{d}}_{ELW}$ are the average biases of the residual memory estimates using CSS and ELW, respectively. All experiments are based on 1000 replications.

The precision of global factor estimates is not distorted even in very small samples. Note that in all the cases, the global factor is more persistent than the regional factors as well as the idiosyncratic terms. Our findings indicate that the use of the canonical correlation procedures is sufficiently robust to estimate the global factor well. In this sense, Breitung and Pigorsch (2013) point out that CCA is useful for a wide range of stationary or mixing processes, and in particular, it works better than the usual PC methods if the variances of the factors are very different.

Changes in signal-to-noise ratios do not affect the estimated factors or residual integration orders. Note that the more the signal-to-noise ratio increases, the less accurate the regional factor is. The worst estimations are shown in rows with $\phi = 5$, where the precision of regional factors (see columns $\overline{R^2_R}$) decreases even below 0.80. Regional factor accuracy with high signal-to-noise ratios improves in two different scenarios: i) considerably, when the cross-sectional dimension increases (compare the columns with $N = 20$ and $N = 80$); and ii) slightly, when there is not much difference between the persistence of the regional and global factors (see the cases with $\vartheta_{r0} = 0.6$, and $\delta_0 = 0.8$).[3]

We study more realistic structures of the idiosyncratic components in Table 4 to have a better understanding of the finite-sample properties of the methodology. We simulate the same DGP as in Table 1 but with cross-correlated idiosyncratic components. The vector of idiosyncratic shocks $u_{r,t} = (u_{r,1t}, \ldots, u_{r,Nt})$ is simulated from a $N(0, \Sigma_u)$ distribution, independently across t, and with the $(i, j)th$ entry of the covariance matrix Σ_u given by $0.5^{|i-j|}$. We also increase the number of regions and varying the persistence of the factors. The first block of six rows shows simulations from standard cases as before, while the last six rows show special cases where the integration order of the idiosyncratic components is equal or even slightly greater than the memory parameter of the

[3] We also studied the same experiments using a smaller bandwidth, $m = \lfloor T^{0.6} \rfloor$. We did not find any relevant difference with respect to the tables that we analyze here. These tables are available upon request.

Table 2
$T = 1000$, $N_r \in (20, 80)$, and $R = 2$. The fractional memory estimates of $\bar{\bar{y}}_t$ and $\hat{\theta}$, averages biases of the memory estimated in both levels of unobservable factors and in the residual, and the measure of the consistency of the unobservable factors estimated are presented in the report.

			$N_r = 20$							$N_r = 80$						
				BIAS				BIAS			BIAS				BIAS	
ϕ	σ_{v_r}	σ_w	$\hat{\theta}$	$\overline{\hat{\vartheta}_R}$	$\hat{\delta}$	$\overline{R^2_R}$	R^2_G	$\bar{\hat{d}}_{R_{CSS}}$	$\bar{\hat{d}}_{R_{ELW}}$	$\hat{\theta}$	$\overline{\hat{\vartheta}_R}$	$\hat{\delta}$	$\overline{R^2_R}$	R^2_G	$\bar{\hat{d}}_{R_{CSS}}$	$\bar{\hat{d}}_{R_{ELW}}$
			$\vartheta_{r0} = 0.2$	$\delta_0 = 0.4$		$d_{r,i0} = 0$										
0.5	2	1	0.365	0.006	0.003	0.991	0.988	−0.002	−0.002	0.350	0.003	0.003	0.997	0.997	−0.002	0.000
0.5	1	1	0.360	0.009	−0.003	0.970	0.988	−0.002	−0.000	0.353	0.003	0.005	0.992	0.997	−0.002	−0.001
0.5	1	2	0.382	0.006	0.006	0.969	0.997	−0.002	−0.001	0.385	0.005	−0.001	0.992	0.999	−0.002	−0.001
2	2	1	0.395	0.007	0.003	0.969	0.954	−0.002	−0.002	0.352	0.004	−0.004	0.991	0.989	−0.002	−0.001
2	1	1	0.351	0.016	0.009	0.888	0.954	−0.002	−0.000	0.356	0.009	0.002	0.972	0.989	−0.002	−0.001
2	1	2	0.385	0.016	0.004	0.891	0.988	−0.002	−0.001	0.390	0.003	−0.001	0.972	0.997	−0.002	−0.001
5	2	1	0.379	0.028	0.006	0.768	0.970	−0.002	−0.002	0.382	0.011	0.006	0.934	0.993	−0.002	−0.001
5	1	1	0.382	0.027	0.004	0.768	0.971	−0.003	−0.001	0.388	0.015	0.000	0.934	0.993	−0.002	−0.001
5	1	2	0.376	0.034	0.009	0.768	0.970	−0.002	−0.001	0.392	0.010	−0.003	0.934	0.993	−0.002	−0.001
			$\vartheta_{r0} = 0.4$	$\delta_0 = 0.8$		$d_{r,i0} = 0.2$										
0.5	2	1	0.769	−0.001	0.009	0.978	0.999	−0.014	−0.018	0.762	−0.002	0.008	0.983	1.000	−0.014	−0.018
0.5	1	1	0.741	0.000	0.016	0.959	0.999	−0.015	−0.020	0.748	−0.005	0.004	0.982	1.000	−0.015	−0.020
0.5	1	2	0.780	−0.005	0.005	0.964	1.000	−0.014	−0.017	0.771	−0.002	0.016	0.981	1.000	−0.014	−0.017
2	2	1	0.756	−0.002	0.018	0.956	0.995	−0.017	−0.020	0.768	−0.002	0.008	0.979	0.999	−0.015	−0.020
2	1	1	0.742	−0.008	0.011	0.880	0.995	−0.016	−0.020	0.745	−0.004	0.009	0.963	0.999	−0.014	−0.017
2	1	2	0.783	0.001	0.004	0.896	0.999	−0.014	−0.017	0.778	−0.002	0.008	0.967	1.000	−0.015	−0.019
5	2	1	0.774	0.008	0.011	0.774	0.997	−0.015	−0.020	0.775	−0.002	0.013	0.932	0.999	−0.015	−0.019
5	1	1	0.772	0.011	0.012	0.775	0.997	−0.015	−0.019	0.772	0.002	0.015	0.933	0.999	−0.014	−0.017
5	1	2	0.773	0.008	0.012	0.762	0.997	−0.017	−0.023	0.773	0.005	0.014	0.934	0.999	−0.014	−0.018
			$\vartheta_{r0} = 0.6$	$\delta_0 = 0.8$		$d_{r,i0} = 0.4$										
0.5	2	1	0.762	0.008	0.010	0.990	0.998	−0.002	−0.004	0.766	0.004	0.009	0.994	1.000	−0.002	−0.005
0.5	1	1	0.747	0.012	0.011	0.981	0.998	−0.002	−0.004	0.750	0.000	0.015	0.992	1.000	−0.002	−0.005
0.5	1	2	0.780	0.006	0.010	0.981	1.000	−0.002	−0.006	0.788	0.009	0.000	0.991	1.000	−0.002	−0.005
2	2	1	0.791	0.006	0.013	0.979	0.992	−0.002	−0.006	0.765	0.004	0.009	0.991	0.998	−0.002	−0.005
2	1	1	0.745	0.020	0.020	0.933	0.993	−0.003	−0.006	0.754	0.004	0.010	0.982	0.998	−0.002	−0.005
2	1	2	0.779	0.015	0.007	0.945	0.998	−0.002	−0.006	0.777	0.009	0.010	0.982	1.000	−0.002	−0.005
5	2	1	0.771	0.029	0.014	0.865	0.996	−0.003	−0.007	0.785	0.010	0.005	0.964	0.999	−0.002	−0.005
5	1	1	0.774	0.036	0.012	0.865	0.995	−0.002	−0.005	0.782	0.017	0.007	0.963	0.999	−0.002	−0.005
5	1	2	0.774	0.035	0.012	0.862	0.995	−0.003	−0.006	0.778	0.011	0.011	0.963	0.999	−0.002	−0.005

Notes: The DGP is the same as that of Table 1, except that now, $T = 1000$. The bandwidth parameter is now $m = \lfloor T^{0.70} \rfloor$, corresponding to $m = 125$. All experiments are based on 1000 replications.

common factors. All results are based on 1000 replications of the model, with $T = 1000$, $Nr = 80$, and $R = 4$.

As seen in the first six rows, our findings from previous tables still hold up in this experiment. In the last six rows, those corresponding to special cases, we study the reliability of our procedure when Assumption G is no longer satisfied. We find interesting conclusions. First, the memory parameters of both levels of common factors, as well as that of residuals, are still well estimated. Second, the accuracy of regional factors depends on their level of persistence compared with that of global factors. Third, when $\vartheta > \delta$, the precision in estimating global factors decreases slightly, but it is still reasonable. These results seem to be independent of the signal-to-noise ratio.

To show graphically the precision of the projected estimators with the true ones, we simulate one replication of the model (1) with $R = 2$, $N_r = 300$, and $T = 150$, with the medium signal-to-noise ratio and $\vartheta_{r,0} = 0.6$, $\delta_0 = 1$, and $d_{r,i0} = 0.25$. Following (Wang, 2010), once we get $\hat{G}$, $\hat{F}$, $\hat{\Gamma}$, and $\hat{\Lambda}$, we project the true factors on the estimated factors to find the rotation matrix, $\hat{Q}_G = \left(\hat{G}'\hat{G}\right)^{-1} \hat{G}'G$. Then, we use $\left(\hat{Q}_G\right)^{-1}$ to rotate factor loadings. Fig. 2 displays such a simulation exercise.

4.2. Number of local and global factors

We now present a Monte Carlo experiment to show the reliability of the methodology proposed to estimate the number of local and global factors. We design our simulation study using the same framework as above.

Tables 5–7 use the information criteria of Bai and Ng (2002) and show that the number of factors is always consistently estimated when variables are fractionally differenced by $\theta = max\left(\delta_0, \vartheta_{r,i0}\right)$. Only in cases when $d_{r,i0} \leq 1$, the number of factors is accurately estimated taking the first differences of the variables. The original variables can be used only in the specific case when $d_{r,i0} = 0$, although the performance of the number of factors does not diminish considerably in cases when $d_{r,i0} < 0.5$ because the data satisfy all assumptions of Bai and Ng (2002). Since Tables 5 and 6 consider two regions, we have three different blocks of data, $B_{R_1 \cup R_2}$, B_{R_1}, and B_{R_2}, as explained above. Table 5 includes the case of only one global factor and one local factor in each region. Consequently, the actual number of static factors in each block is $s_{B_{R_1 \cup R_2}} = 3$, $s_{B_{R_1}} = 2$, and $s_{B_{R_2}} = 2$, as represented in Fig. 1. Table 6 considers the case of two global factors and two local factors, then $s_{B_{R_1 \cup R_2}} = 6$, $s_{B_{R_1}} = 4$, and $s_{B_{R_2}} = 4$. When considering three regions in Table 7,

Table 3
$T = 5000$, $N_r \in (20, 80)$, and $R = 2$. The fractional memory estimates of $\bar{\bar{y}}_t$ and $\hat{\theta}$, average biases of the memory estimated in both levels of unobservable factors and in the residual, and the measure of the consistency of the unobservable factors estimated are presented in the report.

			$N_r = 20$							$N_r = 80$						
				BIAS				BIAS			BIAS				BIAS	
ϕ	σ_{v_r}	σ_w	$\hat{\theta}$	$\overline{\hat{\vartheta}_R}$	$\hat{\delta}$	$\overline{R^2_R}$	R^2_G	$\overline{\hat{d}}_{R_{CSS}}$	$\overline{\hat{d}}_{R_{ELW}}$	$\hat{\theta}$	$\overline{\hat{\vartheta}_R}$	$\hat{\delta}$	$\overline{R^2_R}$	R^2_G	$\overline{\hat{d}}_{R_{CSS}}$	$\overline{\hat{d}}_{R_{ELW}}$
			$\vartheta_{r0} = 0.2$ $\delta_0 = 0.4$ $d_{r,i0} = 0$													
0.5	2	1	0.355	−0.001	0.003	0.992	0.989	−0.001	−0.001	0.355	0.001	−0.001	0.998	0.998	−0.001	0.000
0.5	1	1	0.360	0.003	0.003	0.971	0.989	−0.000	−0.000	0.367	0.000	0.000	0.993	0.998	−0.001	−0.001
0.5	1	2	0.391	0.003	−0.001	0.971	0.997	−0.001	−0.001	0.390	0.004	0.001	0.993	0.999	−0.001	0.000
2	2	1	0.360	0.003	0.006	0.970	0.958	−0.000	−0.000	0.361	0.003	0.002	0.993	0.990	0.000	0.000
2	1	1	0.359	0.010	0.008	0.891	0.960	−0.001	−0.001	0.364	0.004	0.002	0.973	0.990	−0.001	0.000
2	1	2	0.386	0.011	0.003	0.896	0.990	−0.001	−0.001	0.388	0.002	0.003	0.973	0.998	−0.001	−0.001
5	2	1	0.382	0.023	0.007	0.777	0.974	−0.001	−0.000	0.395	0.006	−0.006	0.935	0.994	−0.001	−0.001
5	1	1	0.385	0.020	0.005	0.774	0.973	−0.001	−0.001	0.387	0.005	0.003	0.936	0.994	0.000	0.000
5	1	2	0.384	0.020	0.005	0.771	0.974	−0.001	−0.001	0.387	0.008	0.003	0.935	0.994	−0.001	0.000
			$\vartheta_{r0} = 0.4$ $\delta_0 = 0.8$ $d_{r,i0} = 0.2$													
0.5	2	1	0.759	0.003	0.006	0.987	1.000	−0.007	−0.013	0.751	−0.005	0.007	0.990	1.000	−0.007	−0.014
0.5	1	1	0.766	−0.001	0.003	0.966	1.000	−0.007	−0.014	0.767	−0.003	0.001	0.987	1.000	−0.006	−0.012
0.5	1	2	0.787	−0.001	0.004	0.972	1.000	−0.006	−0.012	0.787	−0.003	0.004	0.990	1.000	−0.006	−0.012
2	2	1	0.763	−0.002	0.011	0.952	0.998	−0.008	−0.013	0.761	−0.003	0.006	0.987	1.000	−0.007	−0.013
2	1	1	0.758	−0.008	0.006	0.856	0.998	−0.007	−0.014	0.763	0.000	0.004	0.970	1.000	−0.006	−0.012
2	1	2	0.787	0.006	0.003	0.897	0.999	−0.006	−0.012	0.785	0.002	0.005	0.973	1.000	−0.006	−0.013
5	2	1	0.784	−0.007	0.006	0.734	0.999	−0.007	−0.014	0.787	0.003	0.004	0.938	1.000	−0.006	−0.013
5	1	1	0.782	−0.005	0.008	0.753	0.999	−0.007	−0.013	0.784	0.003	0.006	0.946	1.000	−0.006	−0.011
5	1	2	0.785	−0.007	0.005	0.743	0.999	−0.007	−0.014	0.792	0.003	−0.001	0.942	1.000	−0.007	−0.014
			$\vartheta_{r0} = 0.6$ $\delta_0 = 0.8$ $d_{r,i0} = 0.4$													
0.5	2	1	0.749	0.001	0.005	0.995	0.999	−0.001	−0.002	0.746	0.002	0.001	0.998	1.000	0.000	−0.001
0.5	1	1	0.763	0.001	0.004	0.987	0.999	−0.001	−0.002	0.766	0.000	0.002	0.995	1.000	0.000	−0.002
0.5	1	2	0.789	−0.000	0.001	0.988	1.000	−0.000	−0.001	0.788	−0.002	0.002	0.996	1.000	0.000	−0.002
2	2	1	0.717	−0.003	0.005	0.984	0.997	−0.001	−0.001	0.710	−0.002	0.006	0.996	0.999	0.000	−0.001
2	1	1	0.762	0.008	0.005	0.945	0.997	−0.000	−0.002	0.763	0.002	0.005	0.988	0.999	0.000	−0.002
2	1	2	0.786	0.011	0.004	0.953	0.999	−0.001	−0.002	0.788	0.005	0.002	0.989	1.000	0.000	−0.002
5	2	1	0.787	0.019	0.002	0.889	0.998	−0.001	−0.002	0.791	0.005	−0.001	0.975	1.000	−0.001	−0.002
5	1	1	0.786	0.019	0.001	0.881	0.998	−0.001	−0.002	0.783	0.004	0.006	0.974	1.000	0.000	−0.002
5	1	2	0.785	0.025	0.004	0.886	0.998	−0.001	−0.002	0.787	0.005	0.003	0.974	1.000	−0.001	−0.001

Notes: The DGP is the same as that of Table 1, except that now, $T = 5000$. The bandwidth parameter is now $m = \lfloor T^{0.70} \rfloor$, corresponding to $m = 388$. All experiments are based on 1000 replications.

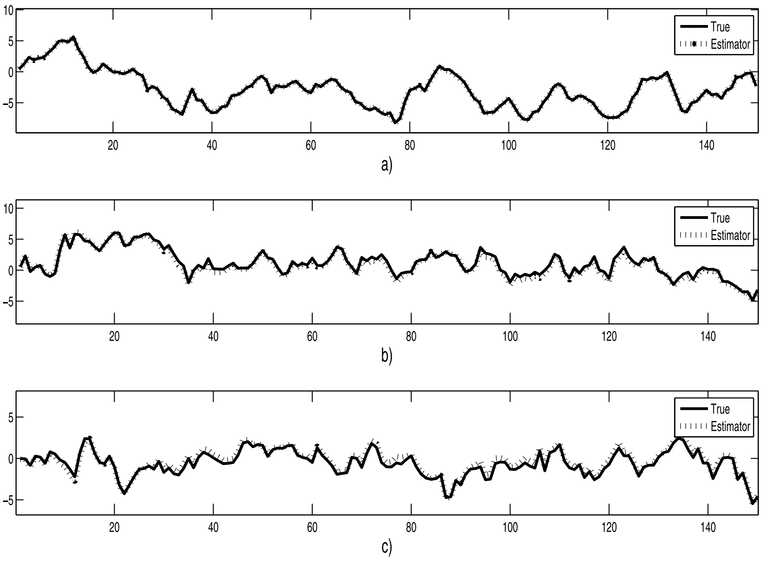

Fig. 2. The dashed lines are the estimators for factors projected onto the true ones (solid lines). The global factor is in panel (a), and the local factors are in panels (b) and (c).

with one global factor and one local factor in each region, we have seven different blocks with the number of static factors as follows: $s_{B_{R_1 \cup R_2 \cup R_3}} = 4$, $s_{B_{R_1}} = 2$, $s_{B_{R_2}} = 2$, $s_{B_{R_3}} = 2$, $s_{B_{R_1 \cup R_2}} = 3$, $s_{B_{R_1 \cup R_2}} = 3$, and $s_{B_{R_2 \cup R_3}} = 3$.

Table 4
$T = 1000$, $N_r = 80$, and $R = 4$. The fractional memory estimates of $\bar{\bar{y}}_t$ and $\hat{\theta}$, averages biases of the memory estimated in both levels of unobservable factors and in the residual, and the measure of the consistency of the unobservable factors estimated are presented in the report.

							BIAS				BIAS	
σ	ϑ_{10}	ϑ_{20}	ϑ_{30}	ϑ_{40}	δ_0	$d_{r,i0}$	$\overline{\hat{\vartheta}_R}$	$\overline{\hat{\delta}}$	$\overline{R_R^2}$	$\overline{R_G^2}$	$\overline{\hat{d}_{CSS}}$	$\overline{\hat{d}_{ELW}}$
Standard cases												
0.5	0.1	0.2	0.3	0.4	0.6	0	0.004	0.000	0.991	0.999	−0.017	−0.010
2	0.1	0.2	0.3	0.4	0.6	0	0.002	0.006	0.992	0.999	−0.016	−0.009
0.5	0.6	0.7	0.8	0.9	0.6	0.5	0.004	0.014	0.998	0.999	−0.001	0.001
2	0.6	0.7	0.8	0.9	0.6	0.5	0.005	0.005	0.998	0.999	−0.001	0.002
0.5	1.25	1.25	1.25	1.25	1.5	1	0.014	0.021	0.974	1.000	−0.002	0.006
2	1.25	1.25	1.25	1.25	1.5	1	0.012	0.026	0.969	1.000	−0.003	0.005
Special cases												
0.5	0.25	0.25	0.25	0.25	0.75	0.75	−0.061	0.010	0.895	0.993	0.008	0.013
2	0.25	0.25	0.25	0.25	0.75	0.75	−0.063	0.001	0.896	0.993	0.008	0.012
0.5	0.6	0.6	0.6	0.6	0.9	1	−0.017	0.007	0.761	0.990	0.004	0.013
2	0.6	0.6	0.6	0.6	0.9	1	−0.018	0.006	0.785	0.991	0.004	0.013
0.5	0.9	0.9	0.9	0.9	0.6	1	0.007	0.001	0.977	0.916	0.001	0.010
2	0.9	0.9	0.9	0.9	0.6	1	0.005	0.007	0.977	0.926	0.001	0.010

Notes: The DGP is the same as that of Table 1, except that now we consider cross-correlated idiosyncratic components. In this experiment, $R = 4$, $N_r = 80$, and $T = 1000$, with a bandwidth parameter of $m = \lfloor T^{0.70} \rfloor$ corresponding to $m = 125$. All experiments are based on 1000 replications.

Table 5
Number of common factors. Two regions. ($N = 40$, $T = 500$ and $k_{max} = 10$). One global factor and one local factor in each region.

	Neglected memory			First difference			Fractional differencing using δ_0		
	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$
$d_{r,i0} = 1.5$, $\delta_0 = 2$, and $\vartheta_{r0} = 1.8$.									
IC1	10	10	10	5.59	3.49	3.45	3.00	2.00	2.00
IC2	10	10	10	5.15	3.31	3.26	3.00	2.00	2.00
IC3	10	10	10	8.39	4.20	4.20	3.00	2.00	2.00
PC1	10	10	10	6.38	4.57	4.55	3.00	2.00	2.00
PC2	10	10	10	5.95	4.35	4.32	3.00	2.00	2.00
PC3	10	10	10	8.84	5.50	5.51	3.00	2.00	2.00
$d_{r,i0} = 0.4$, $\delta_0 = 1$, and $\vartheta_{r0} = 0.7$.									
IC1	5.74	3.51	3.54	3.00	2.00	2.00	3.00	2.00	2.00
IC2	5.30	3.35	3.37	3.00	2.00	2.00	3.00	2.00	2.00
IC3	8.55	4.28	4.31	3.00	2.00	2.00	3.00	2.00	2.00
PC1	6.50	4.63	4.63	3.00	2.00	2.00	3.00	2.00	2.00
PC2	6.06	4.41	4.42	3.00	2.00	2.00	3.00	2.00	2.00
PC3	8.95	5.53	5.56	3.00	2.00	2.00	3.00	2.00	2.00
$d_{r,i0} = 0.6$, $\delta_0 = 1$, and $\vartheta_{r0} = 0.8$.									
IC1	10	9.41	9.38	3.00	2.00	2.00	3.00	2.00	2.00
IC2	10	8.99	8.95	3.00	2.00	2.00	3.00	2.00	2.00
IC3	10	9.99	9.99	3.00	2.00	2.00	3.00	2.00	2.00
PC1	10	9.68	9.66	3.00	2.00	2.00	3.00	2.00	2.00
PC2	10	9.41	9.38	3.00	2.00	2.00	3.00	2.00	2.00
PC3	10	10	10	3.00	2.00	2.00	3.00	2.00	2.00

Notes: The DGP is the same as that of Table 1 but considering idiosyncratic errors to allow for cross-section correlation. $s_{B_{R_1 \cup R_2}}$, $s_{B_{R_1}}$, and $s_{B_{R_2}}$ are the averages of the number of factors estimated in each block. We compare three cases: (i) when memory is neglected, (ii) taking the first difference on $y_{r,it}$, and (iii) fractional differencing $y_{r,it}$ with $\delta_0 = max(\delta_0, \vartheta_{r0})$. IC1, IC2, IC3, PC1, PC2, and PC3 are the information criteria of Bai and Ng (2002). All experiments are based on 1000 replications.

Finally, the number of global and the local factors can be obtained by the inclusion-exclusion principle. Corona et al. (2017) perform a Monte Carlo study to evaluate the effects of several procedures to determine the number of factors after first-differencing the variables. Although it is not further explored here, a possible extension could be to evaluate the impact of fractional differencing the blocks of variables by the procedures of Onatski (2010) and Ahn and Horenstein (2013).

5. Application to Nord Pool power market

In this section, we provide an application of our methodology to study price comovements in the Nord Pool power spot market.

In liberalized power markets, power companies produce electricity from many different sources (hydro, thermal, nuclear, wind, and solar systems) to provide competitive prices. Such spot prices exhibit seasonality

Table 6
Number of common factors. Two regions. ($N = 100$, $T = 500$ and $k_{max} = 10$). Two global factors and two local factors in each region.

	Neglected memory			First difference			Fractional differencing using δ_0		
	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$
$d_{r,i0} = 1.4$, $\delta_0 = 2$, and $\vartheta_{r0} = 1.8$.									
IC1	10	10	10	7.44	4.74	4.72	6.00	4.00	4.00
IC2	10	10	10	7.06	4.61	4.59	6.00	4.00	4.00
IC3	10	10	10	9.74	5.36	5.39	6.00	4.00	4.00
PC1	10	10	10	7.85	5.50	5.52	6.00	4.00	4.00
PC2	10	10	10	7.50	5.31	5.32	6.00	4.00	4.00
PC3	10	10	10	9.83	6.35	6.35	6.00	4.00	4.00
$d_{r,i0} = 0.4$, $\delta_0 = 1$, and $\vartheta_{r0} = 0.7$.									
IC1	7.654	4.81	4.85	6.00	4.00	4.00	6.00	4.00	4.00
IC2	7.245	4.71	4.74	6.00	4.00	4.00	6.00	4.00	4.00
IC3	9.81	5.53	5.52	6.00	4.00	4.00	6.00	4.00	4.00
PC1	8.059	5.64	5.65	6.00	4.00	4.00	6.00	4.00	4.00
PC2	7.699	5.44	5.44	6.00	4.00	4.00	6.00	4.00	4.00
PC3	9.877	6.43	6.49	6.00	4.00	4.00	6.00	4.00	4.00
$d_{r,i0} = 0.6$, $\delta_0 = 1$, and $\vartheta_{r0} = 0.8$.									
IC1	10	10	10	6.69	4.37	4.35	6.00	4.00	4.00
IC2	10	10	10	6.52	4.30	4.29	6.00	4.00	4.00
IC3	10	10	10	7.87	4.68	4.67	6.00	4.00	4.00
PC1	10	10	10	6.90	4.75	4.74	6.00	4.00	4.00
PC2	10	10	10	6.73	4.66	4.64	6.00	4.00	4.00
PC3	10	10	10	7.92	5.18	5.15	6.00	4.00	4.00

Notes: The DGP is the same as that of Table 5. All experiments are based on 1000 replications.

Table 7
Number of common factors. Three regions. Seven blocks of data. ($N = 40, T = 500$ and $k_{max} = 10$). One global factor and one local factor in each region. Estimation is performed in first differences. $d_{r,i0} = 0.6$, $\delta_0 = 1$, and $\vartheta_{r0} = 0.8$.

	$s_{B_{R_1 \cup R_2 \cup R_3}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$	$s_{B_{R_3}}$	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_2 \cup R_3}}$
IC1	4.00	2.00	2.00	2.00	3.00	3.00	3.00
IC2	4.00	2.00	2.00	2.00	3.00	3.00	3.00
IC3	4.00	2.00	2.00	2.00	3.00	3.00	3.00
PC1	4.00	3.21	3.23	3.20	3.00	3.00	3.00
PC2	4.00	3.08	3.08	3.06	3.00	3.00	3.00
PC3	4.00	3.52	3.50	3.47	3.00	3.00	3.00

Notes: The DGP is the same as that of Table 5. All experiments are based on 1000 replications.

at daily and weekly levels, irregular cyclical movements, and spikes that are intrinsically originated in the market. See Weron (2007) for more details. Another feature that has received considerable attention is the presence of a hyperbolic decay of the autocovariances of electricity prices. In this light, Haldrup and Nielsen (2006) point out that Nord Pool prices are characterized by a high degree of long memory. Along this line, Koopman et al. (2007) consider general seasonal periodic regressions with ARFIMA-GARCH disturbances to analyze daily spot prices.

Although daily average prices are widely studied in the literature, due to the role played in the day-ahead market, as (Ramanathan et al., 1997) discuss, it is of interest to disaggregate electricity prices to strengthen the respective prediction. In this regard, Raviv et al. (2015) also point out that the daily averages of the disaggregated hourly forecasts contain useful information to study the daily average price in the Nord Pool market. In addition, it is habitually overlooked when modeling the hourly prices that the vector of 24 hourly prices is determined simultaneously in the day-ahead market. The latter means that a proper form of the data set would be a panel of prices with a natural ordering in the cross-section dimension, instead of a single time series.

Examining in detail the hourly electricity prices implies the study of a complex dependence structure in the market, which has not been extensively considered in the literature. Factor models are standard tools to analyze high-dimensional data and have been recently used in electricity markets (see e.g. Alonso et al. (2011), Dordonnat et al. (2012), and Cataño et al. (2021)). Using a long memory approach in panel models, Ergemen et al. (2016) study the dynamics of Nord Pool electricity prices in the Elspot market and suggest a fractional cointegrating relationship in the panel between electricity prices and their main unobservable common factor.

A possible limitation in the study of Ergemen et al. (2016) is the use of unconstrained equilibrium prices (called system prices) for the entire Nordic region, disregarding the available transmission capacity between market regions. The Elspot market is divided into bidding

Table 8
Number of factors using the inclusion-exclusion principle.

Individual blocks		Pairwise blocks		Triple-wise blocks		Quadruple-wise block	
s_{R_1}	7	$s_{B_{R_1}\cup B_{R_2}}$	11	$s_{B_{R_1}\cup B_{R_2}\cup B_{R_3}}$	14	$s_{B_{R_1}\cup B_{R_2}\cup B_{R_3}\cup B_{R_4}}$	16
s_{R_2}	7	$s_{B_{R_1}\cup B_{R_3}}$	12	$s_{B_{R_1}\cup B_{R_3}\cup B_{R_4}}$	14		
s_{R_3}	7	$s_{B_{R_1}\cup B_{R_4}}$	11	$s_{B_{R_1}\cup B_{R_2}\cup B_{R_4}}$	14		
s_{R_4}	7	$s_{B_{R_2}\cup B_{R_3}}$	11	$s_{B_{R_2}\cup B_{R_3}\cup B_{R_4}}$	14		
		$s_{B_{R_2}\cup B_{R_4}}$	11				
		$s_{B_{R_3}\cup B_{R_4}}$	11				

areas in which system prices do not clear the entire Nordic market, mainly because of the lack of consideration for region-specific price characteristics. In this empirical application, we delve into Nord Pool's price characteristics both by bidding areas and globally.

The data under consideration are in the form of balanced panels consisting of $N_r = 24$ hourly prices for each day for the period from January 1, 2012 to September 30, 2014, due to availability reasons, yielding a total of $T = 1004$ daily observations in each panel. We consider 12 panels, since we analyze 12 bidding areas: five Norwegian bidding areas (NO1–NO5), western Denmark (DK1), eastern Denmark (DK2), four Swedish bidding areas (SE1–SE4), and Finland (FI). All bidding areas are connected. The series are downloaded from the Nord Pool ftp server. The prices are denominated in euros per Mwh of load. We follow Ergemen et al. (2016) to prefilter the series from regular seasonalities and structural changes in the market.

We establish four regions according to the bidding areas of the market as follows: Region 1 = (DK1, DK2), Region 2 = (NO1, NO2, NO3, NO4, and NO5), Region 3 = (SE1, SE2, SE3, SE4), and Region 4 = FI. Then, the cross-sectional dimensions are $N_r = (48, 120, 96, 24)$ in regions 1, 2, 3, and 4, respectively.

We estimate the memory, $\varepsilon_{r,i}$, of hourly local prices with the ELW procedure. The number of Fourier frequencies used is $m = T^{0.7}$, with $T = 1004$ corresponding to $m = 126$. Then, prices are fractionally differenced by their respective estimated memory, $\hat{\varepsilon}_{r,i}$, so that the number of global and local factors can be estimated as described in Section 3.

To estimate the number of local and global factors, we use the procedure proposed by Alessi et al. (2010). This procedure improves the penalization in the criteria IC1 and IC2 of Bai and Ng (2002), introducing a tuning multiplicative constant in the penalty function that leads to a heteroskedasticity-robust inference. Under our assumptions, Theorem 2 in Bai and Ng (2002) is still valid. Using the inclusion-exclusion principle explained in Section 3, we find one global factor and two local factors in each of the regions.

Table 8 presents the number of factors estimated in each block of data, from which one global and two local factors in each region are defined using the inclusion-exclusion principle. The Edwards diagram (see Fig. 3) displays the decomposition of the full factor space. It is possible that remaining factors (pairwise and triple-wise factors) can be estimated following an adapted version of Rodríguez-Caballero and Caporin (2019) to long memory.

We estimate the model specified in (1) by the methodology proposed in Section 2.2 with one global factor and two local factors in each region. Fig. 4 shows the global factor and loadings. The first panel in Fig. 4 also displays the filtered system daily prices by the same method of Ergemen et al. (2016) for comparison purposes.

As seen from Fig. 4, the global factor seems to be highly persistent. Loadings of the global factor show regular behavior among bidding areas. Loadings are positive and larger during the working hours, indicating that the global factor plays a key role from 8 a.m. to 4 p.m. Levels of the loadings are similar across regions. Fig. 4 shows that the global factor fits well to the filtered system prices. Furthermore, the correlation between the global factor and the filtered daily system price is around 0.95, indicating that the global factor may be interpreted as the daily system price. The performance of the global factor loadings is in line with the nature of the operation of companies in the power market.

Figs. 5 and 6 display loadings and regional factors, respectively. Fig. 5 indicates that in general, first regional factors play a key role during non-working hours, explaining much more of the variability. In contrast, second regional factors are more relevant overnight. Furthermore, it is apparent that regional factors are also highly persistent.

We collect the global and regional factors ($\hat{G}_t, \hat{F}_{r,t}$) and estimate the fractional memory parameters with the ELW procedure proposed by Abadir et al. (2007). Fig. 7 shows that the global factor is more persistent than local factors. The regional factors of Regions 2 and 3 are similar and are more persistent than the other regions, while the Danish and Finnish regions, Regions 1 and 4, show less persistence in both regional factors.

As seen from Fig. 8, a fractional cointegration relationship is confirmed in most of the hours in each region, given that $\hat{d}_{r,i} < max(\hat{\delta}, \hat{\vartheta})$. Fig. 8 shows that the persistence levels of the residuals of the model in (1) decrease once we take into account the strong dependence of the hourly electricity prices analyzed with the global and local factors estimated.

6. Forecasting

In this section, we analyze both (univariate) system price forecast settings and (multivariate) regional price forecast settings to illustrate the empirical use of our multi-level factor model. Of special empirical interest, system price forecasts are important for obtaining information on Nord Pool reference prices in the day-ahead markets, while regional price forecasts are useful in shaping bidding strategies in those markets.

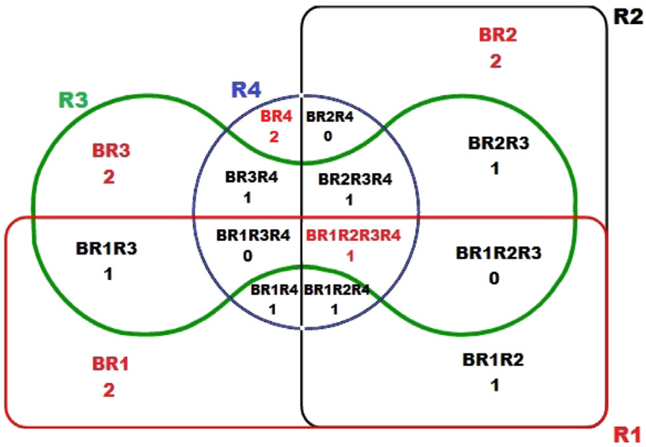

Fig. 3. Decomposition of the number of factors.

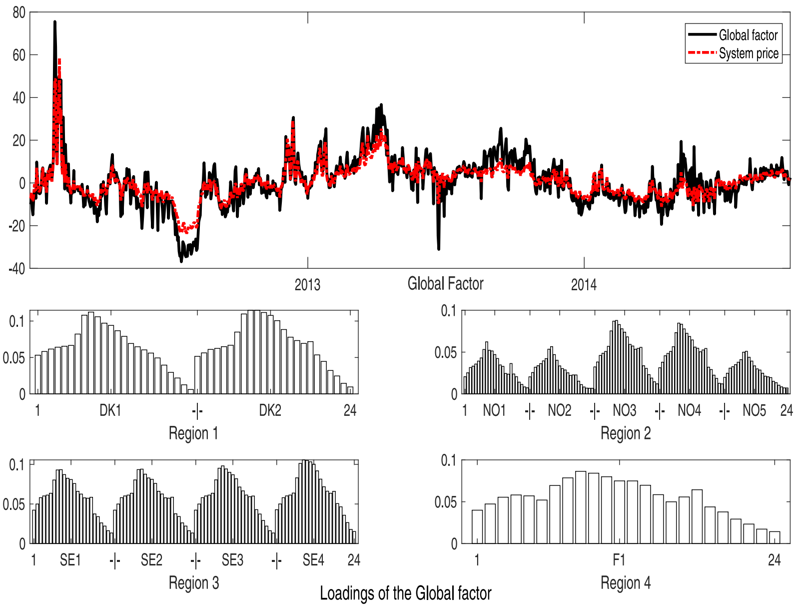

Fig. 4. Global common component of Nord Pool bidding areas.

For both forecasting exercises, we use the same estimation window from January 1, 2012 to September 30, 2014. Forecasts are generated in a window of three months, from October 1, 2014 to December 31, 2014.

For forecast evaluation and comparison purposes, we compute both the out-of-sample root mean square error (RMSE) and the out-of-sample root mean absolute deviation (RMAD) given by

$$RMSE_{i,h} = \sqrt{\frac{1}{h}\sum_{k=1}^{h}\left(y_{t+h} - \hat{y}_{i,t+k}\right)^2},$$

$$RMAD_{i,h} = \sqrt{\frac{1}{h}\sum_{k=1}^{h}|y_{t+h} - \hat{y}_{i,t+k}|},$$

where y_{t+h} is the actual realization of the variable of interest at time $t+h$, with h as the forecast horizon. The models used are indexed by $i \in \{1, \ldots, m\}$, and the out-of-sample forecast from model i is denoted by $\hat{y}_{i,t+k}$.

In both forecasting exercises, to compare the predictive accuracy of two forecasts, we employ the test developed by Diebold and Mariano (2002). In this test, the null hypothesis is that both forecasts have the same accuracy against the alternative that they do not. Particularly, in our analyses, we investigate whether the alternative model in each case is less accurate than the benchmark, which we describe below.

Furthermore, since there are several competing models, we employ the model confidence set (MCS) procedure of Hansen et al. (2011) by the R package provided

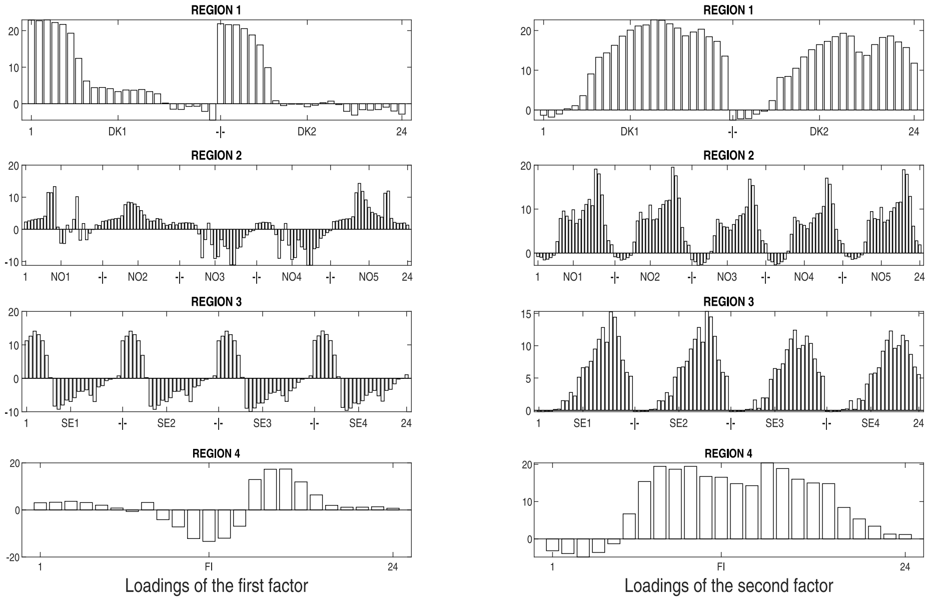

Fig. 5. Loadings of the regional factors.

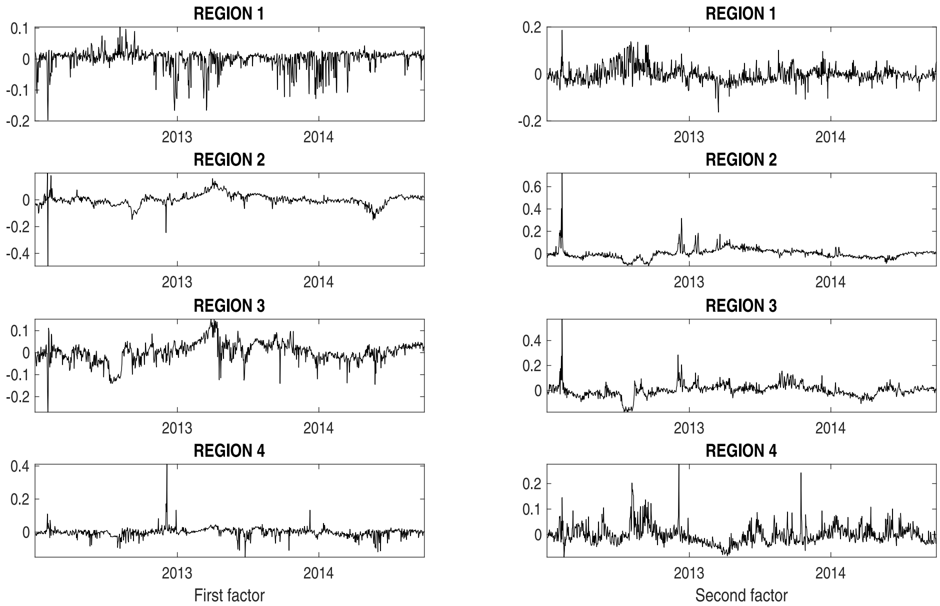

Fig. 6. Regional factors.

by Bernardi and Catania (2018) to compare all predictions jointly. The MCS is a procedure to identify the relative performance of an initial set of models and determine the superior set that can be considered statistically superior at a given confidence level. A requirement when comparing loss functions over time is to guarantee stationarity in the model's loss differentials. Hansen et al. (2011) recommend the use of a rolling window to satisfy such a condition. We refer the reader to the original paper for details.

6.1. Nord Pool reference price forecast

We set the close tie between our global factor estimate and the system price in Section 5. In this part, we assess the forecasting performance of our global factor estimate in forecasting the Nord Pool reference price in contrast to the system price under different model specifications. We study four forecast horizons at $h = (1, 7, 15, 30)$ days, based on the estimation window as above. Forecasts are produced according to a rolling window of three months,

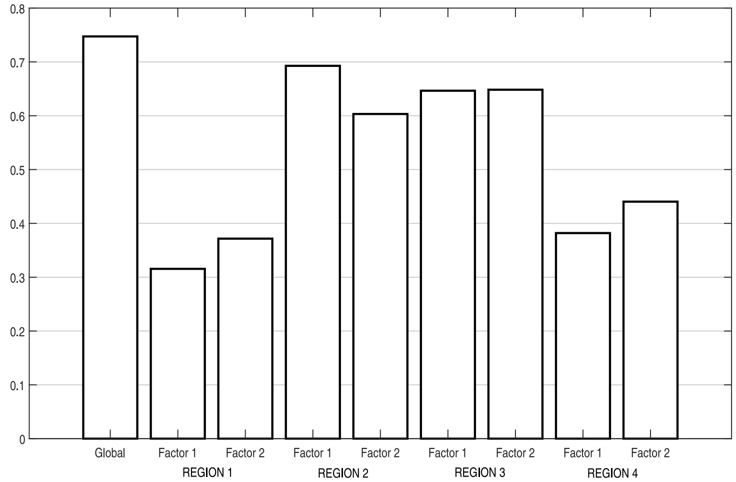

Fig. 7. Memory estimates of the global and local factors. The number of Fourier frequencies used is $m = T^{0.7}$ with $T = 1004$ corresponding to $m = 126$. The standard error of the univariate estimates is 0.044.

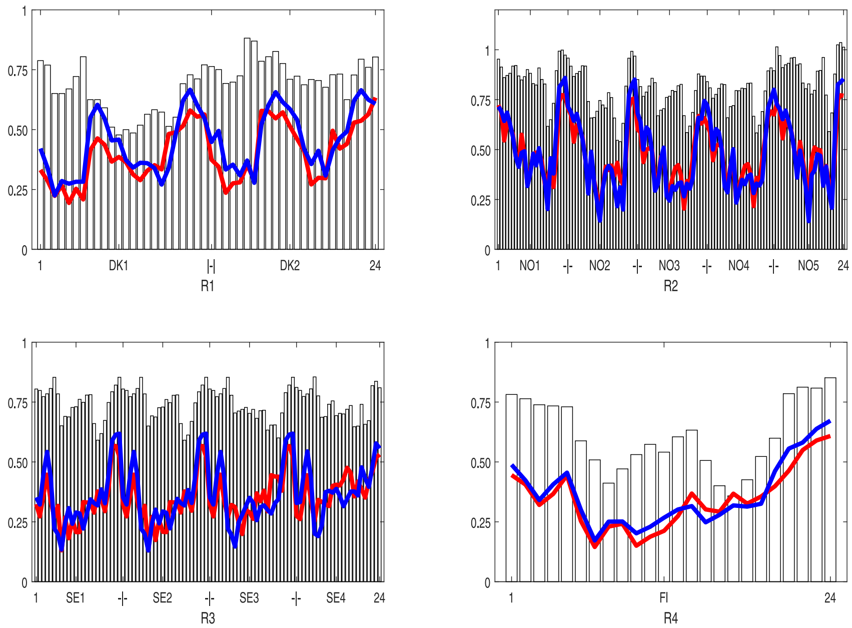

Fig. 8. Comparison of the memory estimates of the filtered local prices for each region (bars) with the residual integration order estimates by CSS (red) and ELW (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

from October 1, 2014 to December 31, 2014, for a total of 92 one-step-ahead, 86 one-week-step-ahead, 78 two-weeks-step-ahead, and 63 one-month-step-ahead forecasts for each of the competing models. Table 9 presents the starting set, $\mathcal{M}_0$, containing all models considered for the MCS approach.

The benchmark is an *ARFIMA*(p, d, q) model estimated for the global factor extracted as in Section 5 in each rolling sample. Our main interest here is to assess the predictive capabilities of the global factor even in situations with high variability in regional prices, market bottlenecks, and other market characteristics. ARFIMA models require first estimating the fractional parameter. Thereafter, an ARMA model is fitted to the filtered data by using maximum likelihood to estimate parameters, and via the use of the Schwarz information criterion (BIC) for lag selection. Note that the number of lags is determined in each rolling sample.

Table 10 reports the (mean) levels of RMSE and RMAD, and the models that belong to the superior set in the MCS.

A glance at Table 10 reveals that models $\mathcal{M}_B$ and $\mathcal{M}_3$ are always in the superior set for any forecast horizon. This result is as expected, due to the high correlation between global factor and system prices. In contrast, models $\mathcal{M}_1$ and $\mathcal{M}_2$ are the worst competing models and are eliminated from the set of superior models most of the time. These results are in line with global findings in the simulation study of Vera-Valdés (2020), given the

Table 9
The table reports all the univariate specifications considered in the starting set of models, $\mathcal{M}_0$, for the Nord Pool reference price.

Abbreviation	Full description
$\mathcal{M}_B$	*ARFIMA*(p, d, q). The best ARFIMA model according to BIC value for the estimated global factor from Section 5. Benchmark model.
$\mathcal{M}_1$	$I(1)$ model.
$\mathcal{M}_2$	*ARIMA*$(p, 1, q)$. The best ARIMA model according to BIC value for the system price.
$\mathcal{M}_3$	*ARFIMA*(p, d, q). The best ARFIMA model according to BIC value for the system price.
$\mathcal{M}_4$	*ARFIMA*$(1, d, 1)$ model for the estimated global factor.
$\mathcal{M}_5$	*ARFIMA*$(1, d, 1)$ model for the system price.

Notes: The first column is the abbreviation of the competing model, while the second provides a brief description.

Table 10
Out-of-sample predictability for system price.

	$h=1$		$h=7$		$h=15$		$h=30$	
	RMSE	RMAD	RMSE	RMAD	RMSE	RMAD	RMSE	RMAD
$\mathcal{M}_B$	0.215	0.414	**0.211**	**0.516**	0.287	0.575	**0.38**	**0.647**
$\mathcal{M}_1$	0.216	0.407	0.254***	0.541***	0.323*	0.601*	0.437***	0.689***
$\mathcal{M}_2$	0.212	0.403	0.239*	0.526**	0.307**	0.594***	0.439***	0.687***
$\mathcal{M}_3$	**0.192**	**0.378**	0.215	0.517	**0.268**	**0.569**	0.383	0.651
$\mathcal{M}_4$	0.224	0.423*	0.259**	0.533**	0.315**	0.586*	0.384	0.649
$\mathcal{M}_5$	0.209	0.403	0.271**	0.543*	0.351*	0.605*	0.447	0.679*

Notes: This table shows means of the RMSE, and RMAD computed over the forecast horizons h. The minimum value is in bold. The description of models is reported in Table 9. Gray shading indicates those models that belong to the superior set of models in the MCS using square errors (SE) or absolute errors (AE) by the range statistics defined by Hansen et al. (2011) at a 20% level of confidence. The symbols ***,**, and * correspond to statistical significance based on the Diebold–Mariano test at significance levels of 1%, 5%, and 10%, respectively.

long memory property of the electricity prices. Furthermore, as Vera-Valdés (2020) highlights, independently of the long memory generating mechanism, ARFIMA models tend to show better performance in medium and long forecast horizons, compared to short memory models.

Even when we rank the models contained in the MCS according to their RMSE or RMAD, the Diebold–Mariano test, which is valid, due to the use of a rolling window scheme, does not reject the null of equal predictive ability for $\mathcal{M}_3$ and $\mathcal{M}_B$ for any h, although the RMSE and RMAD values are lower with $\mathcal{M}_3$ for $h = 1, 15$ and with $\mathcal{M}_B$ for $h = 7, 30$. This tight forecasting performance relationship between models $\mathcal{M}_3$ and $\mathcal{M}_B$ reveals that our global factor estimate can be employed in place of the system price to forecast the Nord Pool reference price.

6.2. Regional price forecasts

Forecasting regional prices can provide useful information for designing bidding strategies in the day-ahead electricity market. Therefore, in this part, we assess the out-of-sample performance of regional price forecasts across different model specifications that include our multi-level factor model. Given all the available information from bidding areas of the Nord Pool power market, the goal is to forecast the price of region r at hour i, denoted by $P_t^{(r,i)}$.

First, we focus on the following multi-level factor augmented models:

$$P_{t+h}^{(r,i)} = \beta_0 + \beta_1 P_t^{(r,i)} + \beta_2 \hat{G}_t + \beta_3 \hat{F}_{1,t}^{(r)} + \beta_4 \hat{F}_{2,t}^{(r)} + u_{t+h}^{(r)}, \quad (8)$$

$$P_{t+h}^{(r,i)} = \beta_0 + \beta_1 \hat{G}_t + \beta_2 \hat{F}_{1,t}^{(r)} + \beta_3 \hat{F}_{2,t}^{(r)} + u_{t+h}^{(r)}, \quad (9)$$

where $h \geq 1$ is the forecast horizon, $\hat{G}_t$ is the global factor estimate, and $\hat{F}_{j,t}^{r}$ for $j = 1, 2$ are the regional factor estimates previously obtained in Section 5. Forecasting can then be done in a two-step process. First, both levels of factors are estimated from all regional prices; second, the relationship between the specific hourly price of the treated region to be forecast and the factors is estimated by linear regression. What makes models (8) and (9) useful for predicting regional prices is the fact that, on the one hand, the global factor would handle the information of the market without any congestion restriction, and, on the other hand, regional factors would correct any sudden deviation from the system prices in any congestion situation.

Then, we consider different model specifications for robustness. The starting set for the MCS approach includes particular cases of models (8)–(9) with a single local factor in each region, multi-level and standard (uni-level) factor models assuming that regional prices follow $I(1)$ processes, and univariate models. Table 11 presents the set $\mathcal{M}_0$ for Nord Pool's regional prices.

Table 11
The table reports all the specifications considered in the starting set of models, $\mathcal{M}_0$, for Nord Pool reference prices.

Abbreviation	Full description
$\mathcal{M}_1$	Model (8). Benchmark model.
$\mathcal{M}_2$	Model (9).
$\mathcal{M}_3$	Model (8) with only one local factor.
$\mathcal{M}_4$	Model (9) with only one local factor.
$\mathcal{M}_5$	Model (8) assuming that prices are $I(1)$.
$\mathcal{M}_6$	Model (9) assuming that prices are $I(1)$.
$\mathcal{M}_7$	Standard factor model with lagged price considering only one factor and assuming prices are $I(1)$.
$\mathcal{M}_8$	As $\mathcal{M}_7$ without lagged price.
$\mathcal{M}_9$	As $\mathcal{M}_7$ but considering three factors.
$\mathcal{M}_{10}$	As $\mathcal{M}_8$ but considering three factors.
$\mathcal{M}_{11}$	As $\mathcal{M}_9$ but considering prices only from the specific region.
$\mathcal{M}_{12}$	As $\mathcal{M}_{10}$ but considering prices only from the specific region.
$\mathcal{M}_{13}$	$I(1)$ model for regional price.
$\mathcal{M}_{14}$	Best $ARIMA(p, 1, q)$ model for the regional price according to BIC.
$\mathcal{M}_{15}$	Best $ARFIMA(p, d, q)$ model for the regional price according to BIC.

Notes: The first column is the abbreviation of the competing model, while the second provides a brief description.

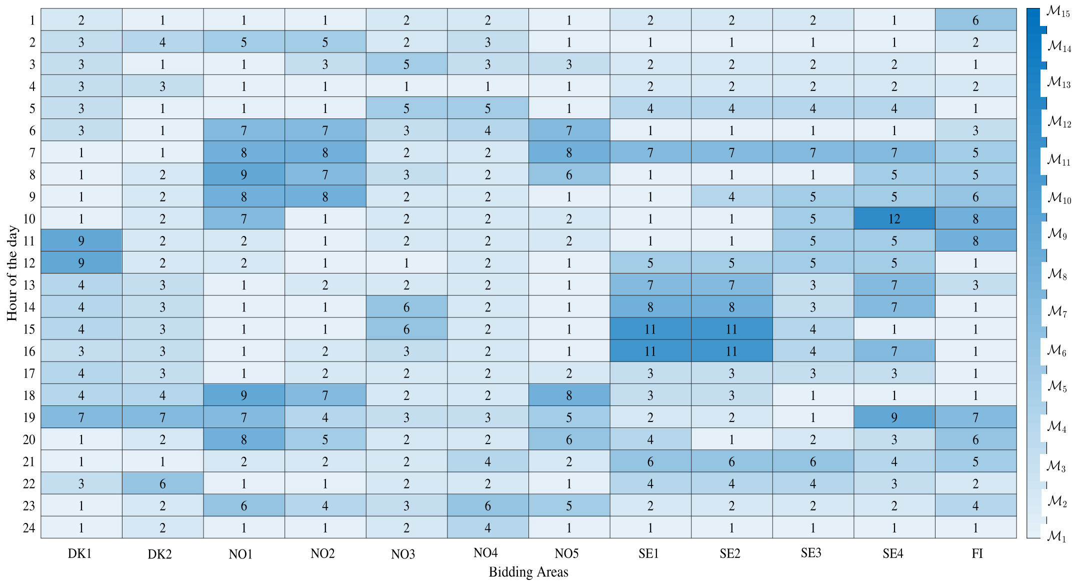

Fig. 9. Two-dimensional representation of top-ranked models in terms of the prediction for each hour of the day and each bidding area. Cells show the colors/shades associated with the winning model in each configuration. The number in the cell indicates the top model according to Table 11, where the description of the models is reported.

In models $\mathcal{M}_3$ and $\mathcal{M}_4$ we investigate the performance of the models with a single local factor, even if the previous analysis in Section 5 suggests considering two instead. With models $\mathcal{M}_5$ and $\mathcal{M}_6$, we compare whether the fractionally integrated approach shows advantages over the $I(1)$ case when predicting. Models $\mathcal{M}_7$–$\mathcal{M}_{12}$ are natural competitive models to evaluate the gain of considering a multi-level approach rather than a standard (uni-level) factor model.

Due to space restrictions, we focus only on the one-step-ahead forecast case, employing a three-month rolling window from October 1 to December 31, 2014, for each hour of the day in each bidding area for each competitive model, for a total of 393,120 forecasts.

Fig. 9 provides an immediate visual summary of the top-ranked model for each hour of the day and each bidding area of the encompassing analysis. We select the top model as in Bernardi and Catania (2018), that is, the lower value of $\bar{d}_{i,j}/(v\hat{a}r(\bar{d}_{i,j}))^{1/2}$, where $d_{i,j}$ is the loss differential between models i and j. See Hansen et al. (2011) for details.

The four lighter blue cells in Fig. 9, models $\mathcal{M}_1$–$\mathcal{M}_4$, are the dominant ones. They are included in the MCS around 75% of the time. This indicates that the fractional multi-level structure works well to predict regional prices. The remaining multi-level models assuming $I(1)$ processes are included in the MCS around 12% of the time, while standard factor models are included around 11% of the time.

Table 12
Out-of-sample predictability for hourly prices in DK1.

	1:00 a.m.	2:00 a.m.	3:00 a.m.	4:00 a.m.	5:00 a.m.	6:00 a.m.	7:00 a.m.	8:00 a.m.	9:00 a.m.	10:00 a.m.	11:00 a.m.	12:00 p.m.
$\mathcal{M}_1$	**0.935**	0.950	**0.845**	**0.858**	0.923	0.813	**0.773**	**0.696**	0.576	0.676	0.715	0.743
$\mathcal{M}_2$	0.940	0.952	0.885	0.899	0.938	0.846	0.774	0.706	0.589*	0.694**	0.725	0.746
$\mathcal{M}_3$	0.941	**0.946**	0.858	0.869	**0.919**	**0.810**	0.776	0.697	**0.574**	**0.674**	0.712	0.736
$\mathcal{M}_4$	0.947	0.961	0.897	0.907	0.941	0.856	0.778	0.701	0.584	0.690*	0.719	0.736
$\mathcal{M}_5$	0.943*	0.956	0.846	0.865	0.925	0.815	0.775	0.698	0.578	0.677	0.714	0.742
$\mathcal{M}_6$	0.948	0.961	0.894*	0.907	0.945	0.855	0.777	0.708	0.592*	0.695**	0.725	0.745
$\mathcal{M}_7$	0.961**	0.954	0.863*	0.871*	0.921	0.812	0.787	0.756**	0.639**	0.759**	0.782	0.789
$\mathcal{M}_8$	0.987**	0.992	0.930**	0.934**	0.964	0.886*	0.803	0.764**	0.641**	0.771**	0.804	0.829
$\mathcal{M}_9$	0.961*	0.961	0.870**	0.878	0.928	0.812	0.776	0.703	0.595	0.690	**0.702**	**0.728**
$\mathcal{M}_{10}$	0.953	0.965	0.903**	0.915*	0.951	0.852	0.776	0.706	0.599	0.699	0.707	0.733
$\mathcal{M}_{11}$	0.964**	0.968	0.876**	0.886*	0.939	0.820	0.767	0.699	0.585	0.677	0.719	0.74
$\mathcal{M}_{12}$	0.99*	1.017*	0.951***	0.963**	0.998*	0.912**	0.773	0.704	0.595	0.695**	0.723	0.739
$\mathcal{M}_{13}$	1.189*	1.154	1.010	1.012*	1.092*	0.916	0.951***	0.902***	0.752***	0.885***	0.896**	0.882**
$\mathcal{M}_{14}$	1.042***	1.014**	0.920**	0.921**	0.967**	0.852*	0.831*	0.794***	0.676***	0.780***	0.791***	0.8***
$\mathcal{M}_{15}$	1.03**	1.01**	0.931***	0.905**	0.967**	0.860**	0.842	0.79**	0.669***	0.749***	0.769***	0.787**

	1:00 p.m.	2:00 p.m.	3:00 p.m.	4:00 p.m.	5:00 p.m.	6:00 p.m.	7:00 p.m.	8:00 p.m.	9:00 p.m.	10:00 p.m.	11:00 p.m.	12:00 a.m.
$\mathcal{M}_1$	0.767	0.792	0.834	0.786	0.810	0.903	0.724	**0.683**	0.672	0.653	0.717	**0.953**
$\mathcal{M}_2$	0.765	0.794	0.834	0.798	0.823	0.889	0.738	0.715**	0.717**	0.709*	0.776**	1.019*
$\mathcal{M}_3$	0.754	0.774	0.819	**0.765**	**0.791**	0.901	0.721	0.684	**0.664**	**0.642**	**0.703**	**0.953**
$\mathcal{M}_4$	**0.750**	**0.771**	**0.809**	0.777	0.801	**0.883**	0.731	0.707*	0.701**	0.689*	0.756*	1.007*
$\mathcal{M}_5$	0.766	0.793	0.834	0.789**	0.812**	0.903	0.724	0.687	0.674	0.654	0.719	0.967
$\mathcal{M}_6$	0.765	0.795	0.834	0.800*	0.824	0.888	0.737	0.716**	0.721**	0.713*	0.782**	1.028*
$\mathcal{M}_7$	0.794	0.797	0.846	0.787	0.809	0.903	**0.719**	0.685*	0.674*	0.649	0.716	0.970
$\mathcal{M}_8$	0.853*	0.871*	0.908	0.869	0.910	0.964	0.778	0.743	0.734	0.734*	0.792***	1.030*
$\mathcal{M}_9$	0.753	0.772	0.816	0.767	0.794	0.896	0.726	0.691**	0.670	0.650	0.708	0.954
$\mathcal{M}_{10}$	0.755	0.775	0.814	0.777	0.806	0.884	0.741	0.715*	0.712***	0.686	0.744	0.994
$\mathcal{M}_{11}$	0.751	0.768	0.813	0.778	0.802	0.893	**0.719**	0.688*	0.679	0.650	0.710	0.981
$\mathcal{M}_{12}$	0.751	0.772	0.815	0.771	0.794	0.884	0.724	0.705**	0.716***	0.699*	0.763	1.051
$\mathcal{M}_{13}$	0.863**	0.865	0.947	0.842	0.887	1.065*	0.826	0.804**	0.777**	0.734	0.804	1.142
$\mathcal{M}_{14}$	0.807**	0.807*	0.859**	0.807**	0.835**	0.924***	0.734***	0.718***	0.711***	0.696*	0.746	1.003
$\mathcal{M}_{15}$	0.801*	0.82	0.852**	0.789	0.81*	0.913***	0.728**	0.695**	0.7***	0.682	0.718	0.976

Notes: This table shows means of the RMAD computed over a forecast horizons of one day. The minimum value is in bold. The description of the models is reported in Table 11. Gray shading indicates those models that belong to the superior set of models in the MCS, using absolute errors (AE) by the range statistics defined by Hansen et al. (2011) at a 20% level of confidence. The symbols ***,**, and * correspond to statistical significance based on the Diebold–Mariano test at significance levels of 1%, 5%, and 10%, respectively.

Standard (uni-level) factor models estimated only from the regional prices in question ($\mathcal{M}_{11}$ and $\mathcal{M}_{12}$) are covered around 2% of the time, supporting the usefulness of the full information of the entire market through a multi-level factor structure. Finally, univariate models are never the best model to forecast regional prices.

For the sake of brevity, we only show in Table 12 the mean level of RMAD and the models that belong to the superior set in the MCS for the first bidding area in Region 1, that is, DK1. The analysis from the remaining bidding areas is available upon request. According to the results, models $\mathcal{M}_1$ and $\mathcal{M}_3$ are always contained in the superior set of models in the MCS, indicating that global and local factors with a lagged value of the regional price help to obtain more accurate forecasts. This stresses that our multi-level factor model can be quite informative in forecasting regional prices.

Remarkably, it seems that the number of local factors may change depending on the hour of the day. This may incentivize two possible subsequent approaches: first, implementing models that can switch the number in the global and/or local factors, as in Cordis and Kirby (2011) and Camacho et al. (2018); and second, exploring the idea of time-varying loadings in multi-level structures, as in Mikkelsen et al. (2019) and Cataño et al. (2021). We leave these extensions for future research.

7. Concluding remarks

In this paper, we considered a dynamic multi-level factor model that allows for both global and local common factors. The multi-level factor structure and model innovations are allowed to exhibit short memory dynamics and long-range dependence without restrictions on them being either stationary $I(0)$ or nonstationary $I(1)$ processes. In order to estimate the model, we proposed a parsimonious two-step procedure and discussed how the number of global and local factors can be determined. Through an extensive simulation study, we showed that the methodology performs well in small samples, and we then applied our method to the analysis of the Nord Pool power market. While the model proposed in this paper is quite general in that it allows for a multi-level factor structure and short-term as well as long-range dynamics, it can nevertheless be extended to account for parametric spatial dependence that would be useful for analyzing economic unions and spillover effects.

Furthermore, forecasting studies using our model can be undertaken to obtain more specific information by exploiting the differences between global and local effects, as in the application to the Nord Pool market.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to Dick van Dijk, an associate editor, and two anonymous referees whose helpful suggestions and constructive comments have improved the paper's version. We are also grateful to Niels Haldrup, Carlos Velasco, Jesús Gonzalo, Esther Ruiz, Javier Hualde, Eduardo Vera-Valdés, and the participants in CREATES Seminars in 2015 and 2016, the VI Workshop in Time Series Econometrics, the IIAAE 2016 Annual Conference, and the CREATES Anniversary Meeting for constructive comments. We acknowledge support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation, and Asociación Mexicana de Cultura, A.C.

Appendix

A.1. Proof of Theorem 2.1

Taking the cross-sectional and local average of the model in (1),

$$\bar{\bar{y}}_t = \bar{\bar{\gamma}}' G_t + \bar{\bar{\lambda}}' \bar{F}_t + \bar{\bar{\epsilon}}_t$$

where $\bar{F}_t = R^{-1}\sum_{r=1}^{R} F_{r,t}$ with R fixed, and the quantities $\bar{\bar{\gamma}} \neq 0$ and $\bar{\bar{\lambda}} \neq 0$ under Assumption C do not depend on i or r, so $\bar{\bar{y}}_t$ is a pure time series that is integrated of order θ. Furthermore, note that G_t, $\bar{F}_t$, and $\bar{\bar{\epsilon}}_t$ have zero mean under Assumption A.2 and their respective variances are proportional to

$$O\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}|t-s|^{2\delta_{max}-1}\right) = O(1+T^{2\delta_{max}-1}\log T)$$

$$O\left(\frac{1}{RT}\sum_{s=1}^{T}\sum_{t=1}^{T}|t-s|^{2\vartheta_{max}-1}\right) = O(R^{-1}+R^{-1}T^{2\vartheta_{max}-1}\log T)$$

$$O\left(\frac{1}{NT}\sum_{s=1}^{T}\sum_{t=1}^{T}|t-s|^{2d_{max}-1}\right) = O(N^{-1}+N^{-1}T^{2d_{max}-1}\log T),$$

since the long-range dependence dynamics dominate the short memory dynamics as $T \to \infty$; see e.g. Robinson and Velasco (2015). Furthermore, noting that R is fixed under Assumption C and that the model components are orthogonal to each other under Assumption E,

$$\bar{\bar{y}}_t = O_p\left(1+\left(T^{\delta_{max}-1/2}+T^{\vartheta_{max}-1/2}+N^{-1/2}T^{d_{max}-1/2}\right)\log T\right).$$

To determine what θ corresponds to, we analyze the cases exhaustively as follows. If N is fixed, the contribution of each term to the variance of $\bar{\bar{y}}_t$ depends on their integration orders, so $\theta = \max\{\delta_{max}, \vartheta_{max}, d_{max}\}$. On the other hand, if $N \to \infty$, there are two cases: 1. if $\min\{\delta_{max}, \vartheta_{max}\} \geq d_{max}$ or $\max\{\delta_{max}, \vartheta_{max}\} \geq d_{max} \geq \min\{\delta_{max}, \vartheta_{max}\}$, for which $\max\{\delta_{max}, \vartheta_{max}\} \geq d_{max}$ is sufficient, $\theta = \max\{\delta_{max}, \vartheta_{max}\}$; 2. if $d_{max} > \max\{\delta_{max}, \vartheta_{max}\}$ and $N^{1/2}T^{\min\{\delta_{max},\vartheta_{max}\}-d_{max}} \to 0$, $\theta = d_{max}$ because the contribution of the averaged error term dominates those of the factors only when $N^{-1/2}T^{d_{max}-\min\{\delta_{max},\vartheta_{max}\}} \to \infty$.

For each of these cases, given that the averaged quantities are pure time series and that exact local Whittle estimation is silent about the exact data-generating process, showing the consistency and asymptotic normality of $\hat{\theta}$ follows from the proof of Theorem 4 in Shimotsu and Phillips (2005) under the conditions of Assumption A. □

A.2. Proof of Theorem 2.2

We begin by showing that the prewhitened factors and errors are asymptotically stationary, because under Assumption A.2, they have zero mean and their variances

$$Var(G_t(\theta)) \propto O\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}|t-s|^{2(\delta_{max}-\theta)-1}\right)$$
$$= O(1+T^{2(\delta_{max}-\theta)-1}\log T) = O(1)$$

$$\sup_r Var(F_{r,t}(\theta)) \propto O\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}|t-s|^{2(\vartheta_{max}-\theta)-1}\right)$$
$$= O(1+T^{2(\vartheta_{max}-\theta)-1}\log T) = O(1)$$

$$\sup_{i,r} Var(\epsilon_{r,it}(\theta)) \propto O\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}|t-s|^{2(d_{max}-\theta)-1}\right)$$
$$= O(1+T^{2(d_{max}-\theta)-1}\log T) = O(1), \tag{10}$$

since under Assumption G, $\theta = \max\{\delta_{max}, \vartheta_{max}\} \geq d_{max}$ and $\hat{\theta} = \theta + O_p(m^{-1/2})$, based on the result in Theorem 2.1.

Then, we can write

$$\hat{G}_t(\hat{\theta}) - L_G' G_t(\theta) = \hat{G}_t(\hat{\theta}) - \hat{G}_t(\theta) + \hat{G}_t(\theta) - L_G' G_t(\theta)$$
$$= I_1 + I_2 \tag{11}$$

and analyze I_1 and I_2 separately:

$$I_1 := \hat{G}_t(\hat{\theta}) - \hat{G}_t(\theta) = \dot{\hat{G}}_t(\theta^\dagger)(\hat{\theta}-\theta)$$
$$= o_p(m^{-1/2}),$$

arguing as in Robinson and Hidalgo (1997), where the first equality is obtained by applying the mean value theorem, and the second equality is because $\hat{\theta} - \theta = O_p(m^{-1/2})$.

To analyze the remainder of the terms in (11), I_2, let us reformulate the CCA considering only two regions $a, b = 1, 2$, which is enough to isolate the global factor. Arguing as in Choi et al. (2018), let $\hat{S}_{a,b} = T^{-1}\sum_{t=1}^{T}\hat{H}_{a,t}(\theta)\hat{H}_{b,t}(\theta)'$. The largest generalized eigenvalues, $\hat{\mu}$, satisfying

$$\left|\hat{S}_{1,2}\hat{S}_{2,2}^{-1}\hat{S}_{2,1} - \hat{\mu}\hat{S}_{1,1}\right| = 0$$

provide the largest squared sample correlation coefficients between arbitrary linear combinations of $\hat{H}_{1,t}(\theta)$

and $\hat{H}_{2,t}(\theta)$, and the coefficients of the linear combination of $\hat{H}_{1,t}(\theta)$ corresponding to μ_j are given by $\hat{p}_j$, such that

$$\left(\hat{S}_{1,2}\hat{S}_{2,2}^{-1}\hat{S}_{2,1} - \hat{\mu}_j\hat{S}_{1,1}\right)\hat{p}_j = 0$$
$$\text{s.t.}\quad \hat{p}_j'\hat{p}_j = 1 \text{ and } \hat{p}_j'\hat{p}_k = 0,\quad j = 1, \ldots, r_G + r_{F_1},\quad 1 \le k < j.$$

Since, for example, $\hat{G}_{tj}^{(1)}(\theta) = \hat{p}_j'\hat{H}_{1,t}(\theta)$, $j = 1, \ldots, r_G$, under Assumptions A–H as $(N, T) \to \infty$,

$$I_2 = \hat{G}_t(\theta) - L_G'G_t(\theta) = O_p\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T}\}}\right)$$

given the result in part (iii) of Proposition 2 in Choi et al. (2018) based on Theorem 1 of Bai (2003).

Therefore, noting that $1/\min\{\sqrt{N}, \sqrt{T}\} = O(1/\sqrt{N} + 1/\sqrt{T})$ as $(N, T) \to \infty$,

$$I_1 + I_2 = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}\right) + o_p(m^{-1/2}), \tag{12}$$

and

$$\sqrt{N}\,(I_1 + I_2) = O_p(1) + O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right) + o_p\left(\frac{\sqrt{N}}{\sqrt{m}}\right),$$

where the $O_p(1)$ term corresponds to the asymptotic normal distribution, as we establish below, and if $N/m \to 0$ as $(N, T) \to \infty$, the asymptotic normal distribution is guaranteed to be centered.

Write

$$\begin{aligned}\sqrt{N}\left(\hat{G}_t(\theta) - L_G'G_t(\theta)\right) &= V_G^{-1}\left(\frac{\hat{G}(\theta)'G(\theta)}{T}\right)\\ &\quad\times \frac{1}{\sqrt{N}}\sum_{r=1}^{R}\sum_{i=1}^{N_r}\gamma_{i,r}\epsilon_{r,it}(\theta) + o_p(1)\\ &\to_d N\left(0, \Xi_G\Sigma_t^*\Xi_G'\right)\end{aligned} \tag{13}$$

for fixed t as $(N, T) \to \infty$, reasoning as in the proof of Theorem 1 in Bai (2003), since under Assumption H.3, we have $(1/\sqrt{N})\sum_{r=1}^{R}\sum_{i=1}^{N_r}\gamma_{i,r}\epsilon_{r,it}(\theta) \to_d N(0, \Sigma_t^*)$ (because $\epsilon_{r,it}(\theta)$ is stationary under Assumption G, see (10), and thus all CLT conditions are still satisfied after prewhitening by θ), and by definition, $\Xi_G = plim V_G^{-1}\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(\theta)G_t(\theta)'$.

Next, we show the result for the prewhitened local factor estimate $\hat{F}_{r,t}(\hat{\theta})$. First, let us write

$$\begin{aligned}\hat{F}_{r,t}(\hat{\theta}) - L_{F_r}'F_{r,t}(\theta) &= \hat{F}_{r,t}(\hat{\theta}) - \hat{F}_{r,t}(\theta) + \hat{F}_{r,t}(\theta) - L_{F_r}'F_{r,t}(\theta)\\ &= J_1 + J_2\end{aligned} \tag{14}$$

where

$$\begin{aligned}J_1 := \hat{F}_{r,t}(\hat{\theta}) - \hat{F}_{r,t}(\theta) &= \dot{\hat{F}}_{r,t}(\theta^\dagger)(\hat{\theta} - \theta)\\ &= o_p(m^{-1/2}),\end{aligned} \tag{15}$$

$$J_2 := \hat{F}_{r,t}(\theta) - L_{F_r}'F_{r,t}(\theta) = O_p\left(\frac{1}{\min\{\sqrt{N_r}, \sqrt{T}\}}\right) \tag{16}$$

based on result (i) in Proposition 3 of Choi et al. (2018), so

$$\sqrt{N_r}\,(J_1 + J_2) = O_p(1) + O_p\left(\frac{\sqrt{N_r}}{\sqrt{T}}\right) + o_p\left(\frac{\sqrt{N_r}}{\sqrt{m}}\right)$$

where the $O_p(1)$ term corresponds to the asymptotic normal distribution, and if $N_r/m \to 0$, it is centered, arguing as before. Write

$$J_2 = K_1 + K_2 + K_3$$

where

$$K_1 = V_{RT}^{-1}\frac{1}{T}\sum_{k=1}^{T}\hat{F}_{r,k}(\theta)\frac{\epsilon_{r,k}(\theta)'\epsilon_{r,t}(\theta)}{N_r},$$

$$K_2 = V_{RT}^{-1}\left(\frac{\hat{F}_r(\theta)'F_r(\theta)}{T}\right)\frac{1}{N_r}\sum_{i=1}^{N_r}\lambda_{i,r}\epsilon_{r,it}(\theta),\quad \text{and}$$

$$K_3 = V_{RT}^{-1}\frac{1}{T}\sum_{k=1}^{T}\hat{F}_{r,k}(\theta)\frac{\epsilon_{r,k}(\theta)'\Lambda_rF_{r,t}(\theta)}{N_r}.$$

Now, arguing the same way as to obtain (13), it is easy to see that

$$\sqrt{N_r}K_2 \to_d N\left(\Xi_{F_r}\Sigma_{r,t}^*\Xi_{F_r}'\right),$$

so what remains to show is that the contribution of the remaining terms is negligible. Under Assumptions F.3 and H, and if $\frac{N_r}{T} \to 0$ as $(N_r, T) \to \infty$, we have that $\sqrt{N_r}K_1 = o_p(1)$ and $\sqrt{N_r}K_3 = o_p(1)$, as in the proof of Theorem 2 of Wang (2010). Therefore, if $N_r/m \to 0$ as $(N_r, T) \to \infty$,

$$\sqrt{N_r}\,(J_1 + J_2) \to_d N\left(\Xi_{F_r}\Sigma_{r,t}^*\Xi_{F_r}'\right). \quad \square$$

A.3. Proof of Theorem 2.3

These results can be shown using the results in Theorem 2.2, establishing the convergence rates for the loading estimates by reasoning exactly as in the proof of Theorem 2.2, and then following the proof of Theorem 3 of Bai (2003). □

References

Abadir, K. M., Distaso, W., & Giraitis, L. (2007). Nonstationarity-extended local whittle estimation. *Journal of Econometrics*, *141*(2), 1353–1384.

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

Alessi, L., Barigozzi, M., & Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters*, *80*(23), 1806–1813.

Alonso, A. M., Bastos, G., & García-Martos, C. (2016). Electricity price forecasting by averaging dynamic factor models. *Energies*, *9*(8), 600.

Alonso, A. M., García-Martos, C., Rodríguez, J., & Jesús Sánchez, M. (2011). Seasonal dynamic factor analysis and bootstrap inference: application to electricity market forecasting. *Technometrics*, *53*(2), 137–151.

Andrews, D. W., & Sun, Y. (2004). Adaptive local polynomial whittle estimation of long-range dependence. *Econometrica*, *72*(2), 569–614.

Arteche, J. (2020). Exact local whittle estimation in long memory time series with multiple poles. *Economic Theory*, *36*(6), 1064–1098.

Artis, M. J., Banerjee, A., & Marcellino, M. (2005). Factor forecasts for the UK. *Journal of Forecasting*, *24*(4), 279–298.

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

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.
Bai, J., & Ng, S. (2004). A PANIC attack on unit roots and cointegration. *Econometrica*, *74*(4), 1127–1177.
Bai, J., & Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends (R) in Econometrics*, *3*(2), 89–163.
Banerjee, A., & Marcellino, M. (2006). Are there any reliable leading indicators for US inflation and GDP growth? *International Journal of Forecasting*, *22*(1), 137–151.
Barigozzi, M., & Hallin, M. (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks. *The Econometrics Journal*, *19*(1).
Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, *120*(1), 387–422.
Bernardi, M., & Catania, L. (2018). The model confidence set package for R. *International Journal of Computational Economics and Econometrics*, *8*(2), 144–158.
Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, *132*(1), 169–194.
Bollerslev, T., Osterrieder, D., Sizova, N., & Tauchen, G. (2013). Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, *108*(2), 409–424.
Breitung, J., & Eickmeier, S. (2016). Analyzing international business and financial cycles using multi-level factor models: A comparison of alternative approaches. *Vol. 35*, In *Dynamic factor models (advances in econometrics, volume 35) emerald group publishing limited* (pp. 177–214). Emerald Group Publishing Limited.
Breitung, J., & Pigorsch, U. (2013). A canonical correlation approach for selecting the number of dynamic factors. *Oxford Bulletin of Economics and Statistics*, *75*(1), 23–36.
Camacho, M., Perez-Quiros, G., & Poncela, P. (2018). Markov-switching dynamic factor models in real time. *International Journal of Forecasting*, *34*(4), 598–611.
Cataño, D. H., Rodríguez-Caballero, C. V., Peña, D., & Chiann, C. (2021). Wavelet estimation for factor models with time-varying loadings. *International Journal of Wavelets, Multiresolution and Information Processing*, Article 2150033.
Chamberlain, G., & Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 1281–1304.
Chen, W. W., & Hurvich, C. M. (2006). Semiparametric estimation of fractional cointegrating subspaces. *The Annals of Statistics*, *34*(6), 2939–2979.
Choi, I., Kim, D., Kim, Y. J., & Kwark, N.-S. (2018). A multilevel factor model: Identification, asymptotic theory and applications. *Journal of Applied Econometrics*, *33*(3), 355–377.
Cipollini, A., & Kapetanios, G. (2008). A stochastic variance factor model for large datasets and an application to S&P data. *Economics Letters*, *100*(1), 130–134.
Cordis, A. S., & Kirby, C. (2011). Regime-switching factor models in which the number of factors defines the regime. *Economics Letters*, *112*(2), 198–201.
Corona, F., Poncela, P., & Ruiz, E. (2017). Determining the number of factors after stationary univariate transformations. *Empirical Economics*, *53*(1), 351–372.
Cristadoro, R., Forni, M., Reichlin, L., & Veronese, G. (2005). A core inflation indicator for the euro area. *Journal of Money, Credit, and Banking*, *37*(3), 539–560.
Davidson, J., & Hashimzade, N. (2009). Type I and type II fractional Brownian motions: A reconsideration. *Computational Statistics & Data Analysis*, *53*(6), 2089–2106.
Diebold, F. X., Li, C., & Yue, V. Z. (2008). Global yield curve dynamics and interactions: a dynamic nelson–Siegel approach. *Journal of Econometrics*, *146*(2), 351–363.
Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144.
Dordonnat, V., Koopman, S. J., & Ooms, M. (2012). Dynamic factors in periodic time-varying regressions with an application to hourly electricity load mo-delling. *Computational Statistics & Data Analysis*, *56*(11), 3134–3152.
Ergemen, Y. E., Haldrup, N., & Rodríguez-Caballero, C. V. (2016). Common long-range dependence in a panel of hourly nord pool electricity prices and loads. *Energy Economics*, *60*, 79–96.
Ergemen, Y. E., & Velasco, C. (2017). Estimation of fractionally integrated panels with fixed effects and cross-section dependence. *Journal of Econometrics*, *196*(2), 248–258.
Favero, C. A., Marcellino, M., & Neglia, F. (2005). Principal components at work: the empirical analysis of monetary policy with large data sets. *Journal of Applied Econometrics*, *20*(5), 603–620.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, *82*(4), 540–554.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2004). The generalized dynamic factor model consistency and rates. *Journal of Econometrics*, *119*(2), 231–255.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, *100*(471), 830–840.
Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676.
Gil-Alana, L. A., & Robinson, P. M. (1997). Testing of unit root and other nonstationary hypotheses in macroeconomic time series. *Journal of Econometrics*, *80*(2), 241–268.
Haldrup, N., & Nielsen, M. Ø. (2006). A regime switching long memory model for electricity prices. *Journal of Econometrics*, *135*(1), 349–376.
Hallin, M., & Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, *102*(478), 603–617.
Hallin, M., & Liška, R. (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics*, *163*(1), 29–41.
Hallin, M., Mathias, C., Pirotte, H., & Veredas, D. (2011). Market liquidity as dynamic factors. *Journal of Econometrics*, *163*(1), 42–50.
Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.
Hartl, T., & Weigand, R. (2018). Multivariate fractional components analysis. arXiv preprint arXiv:1812.09149.
Hendry, D. (1995). *Dynamic econometrics*. Oxford University Press.
Henry, M., & Robinson, P. (1996). Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In *Athens conference on applied probability and time series analysis* (pp. 220–232). Springer.
Hualde, J., & Nielsen, M. (2020). Truncated sum of squares estimation of fractional time series models with deterministic trends. *Economic Theory*, *36*(4), 751–772.
Iacone, F. (2010). Local whittle estimation of the memory parameter in presence of deterministic components. *Journal of Time Series Analysis*, *31*(1), 27–49.
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*(1), 63–69.
Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics*, *28*(3), 397–409.
Koopman, S. J., Ooms, M., & Carnero, M. A. (2007). Periodic seasonal reg-arfima–garch models for daily electricity spot prices. *Journal of the American Statistical Association*, *102*(477), 16–27.
Liu, K., Chen, Y., & Zhang, X. (2017). An evaluation of ARFIMA (autoregressive fractional integral moving average) programs. *Axioms*, *6*(2), 16.
Luciani, M., & Veredas, D. (2015). Estimating and forecasting large panels of volatilities with approximate dynamic factor models. *Journal of Forecasting*, *34*(3), 163–176.
Ludvigson, S. C., & Ng, S. (2007). The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics*, *83*(1), 171–222.
McCloskey, A., & Perron, P. (2013). Memory parameter estimation in the presence of level shifts and deterministic trends. *Economic Theory*, *29*(6), 1196–1237.
Michelacci, C., & Zaffaroni, P. (2000). (Fractional) beta convergence. *Journal of Monetary Economics*, *45*(1), 129–153.
Mikkelsen, J. G., Hillebrand, E., & Urga, G. (2019). Consistent estimation of time-varying loadings in high-dimensional factor models. *Journal of Econometrics*, *208*(2), 535–562.
Moench, E., Ng, S., & Potter, S. (2013). Dynamic hierarchical factor models. *The Review of Economics and Statistics*, *95*(5), 1811–1817.

Morana, C. (2007). On the macroeconomic causes of exchange rates volatility. In: ICER working papers.
Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, *92*(4), 1004–1016.
Pesaran, M. H., & Chudik, A. (2014). Aggregation in large dynamic panels. *Journal of Econometrics*, *178*, 273–285.
Ramanathan, R., Engle, R., Granger, C. W., Vahid-Araghi, F., & Brace, C. (1997). Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, *13*(2), 161–174.
Raviv, E., Bouwman, K. E., & van Dijk, D. (2015). Forecasting day-ahead electricity prices: Utilizing hourly prices. *Energy Economics*, *50*, 227–239.
Ray, B. K., & Tsay, R. S. (2000). Long-range dependence in daily stock volatilities. *Journal of Business & Economic Statistics*, *18*(2), 254–262.
Robinson, P. (2005). Efficiency improvements in inference on stationary and nonstationary fractional time series. *The Annals of Statistics*, *33*(4), 1800–1842.
Robinson, P., & Hidalgo, J. (1997). Time series regression with long-range dependence. *The Annals of Statistics*, *25*(1), 77–104.
Robinson, P. M., & Velasco, C. (2015). Efficient inference on fractionally integrated panel data models with fixed effects. *Journal of Econometrics*, *185*(2), 435–452.
Rodríguez-Caballero, C. V., & Caporin, M. (2019). A multilevel factor approach for the analysis of CDS commonality and risk contribution. *Journal of International Financial Markets, Institutions and Money*, *63*, Article 101144.
Shimotsu, K. (2010). Exact local whittle estimation of fractional integration with unknown mean and time trend. *Economic Theory*, *26*(02), 501–540.
Shimotsu, K., & Phillips, P. C. B. (2005). Exact local whittle estimation of fractional integration. *The Annals of Statistics*, *33*(4), 1890–1933.
Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.
Stock, J. H., & Watson, M. W. (2006). Forecasting with many predictors. *Handbook of Economic Forecasting*, *1*, 515–554.
Vera-Valdés, J. E. (2020). On long memory origins and forecast horizons. *Journal of Forecasting*, *39*(5), 811–826.
Wang, P. (2010). Large dimensional factor models with a multi-level factor structure: Identification, estimation, and inference. *Working Paper. Hong Kong University of Science and Technology*.
Weron, R. (2007). *vol. 403*, *Modeling and forecasting electricity loads and prices: A statistical approach*. John Wiley & Sons.