

Factor Vector Autoregressive Estimation of Heteroskedastic Persistent and Non Persistent Processes Subject to Structural Breaks

Claudio Morana^{1,2}

¹Department of Economics, Management and Statistics, University of Milan-Bicocca, Milan, Italy

²Center for Research on Pensions and Welfare Policies, Collegio Carlo Alberto, Moncalieri, Italy

Email: claudio.morana@unimib.it

Received 22 March 2014; revised 18 April 2014; accepted 6 May 2014

Copyright © 2014 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In the paper, a general framework for large scale modeling of macroeconomic and financial time series is introduced. The proposed approach is characterized by simplicity of implementation, performing well independently of persistence and heteroskedasticity properties, accounting for common deterministic and stochastic factors. Monte Carlo results strongly support the proposed methodology, validating its use also for relatively small cross-sectional and temporal samples.

Keywords

Long and Short Memory, Structural Breaks, Common Factors, Principal Components Analysis, Fractionally Integrated Heteroskedastic Factor Vector Autoregressive Model

1. Introduction

In the paper, a general strategy for large-scale modeling of macroeconomic and financial data, set within the factor vector autoregressive model (F-VAR) framework, is proposed.¹

Following the lead of dynamic factor model analysis proposed in [2], it is assumed that a small number of structural shocks are responsible for the observed comovement in economic data; it is, however, also assumed that commonalities across series are described by deterministic factors, *i.e.*, common break processes. Comovement across series is then accounted by both deterministic and stochastic factors; moreover, common factors are

¹The literature on F-VAR models is large. See [1] for a survey.

allowed in both mean and variance, covering the I(0) and I(1) persistence cases, as well as the intermediate case of long memory, *i.e.*, I(d), $0 < d < 1$. As the common factors are unobserved, accurate estimation may fail in the framework of small scale vector autoregressive (VAR) models, but succeed when cross-sectional information is employed to disentangle common and idiosyncratic features.

The proposed fractionally integrated heteroskedastic factor vector autoregressive model (FI-HF-VAR) bridges the F-VAR and (the most recent) G-VAR literature, as, similarly to [3], a weakly stationary cyclical representation is employed; yet, similarly to [4], principal components analysis (PCA) is employed for the estimation of the latent factors. Consistent and asymptotically normal estimation is performed by means of QML, also implemented through an iterative multi-step estimation procedure. Monte Carlo results strongly support the proposed methodology.

Overall, the FI-HF-VAR model can be understood as a unified framework for large-scale econometric modeling, allowing for accurate investigation of cross-sectional and time series features, independent of persistence and heteroskedasticity properties of the data, from comovement to impulse responses, forecast error variance and historical decomposition analysis.

After this introduction, the paper is organized as follows. In Section 2, the econometric model is presented; in Section 3, estimation is discussed, while Monte Carlo analysis is performed in Section 4; finally, conclusions are drawn in Section 5.

2. The FI-HF-VAR Model

Consider the following fractionally integrated heteroskedastic factor vector autoregressive (FI-HF-VAR) model

$$x_t - \Lambda_\mu \mu_t - \Lambda_f f_t = C(L)(x_{t-1} - \Lambda_\mu \mu_{t-1} - \Lambda_f f_{t-1}) + v_t \quad (1)$$

$$v_t \sim i.i.d.(0, \Sigma_v)$$

$$P(L)D(L)f_t = \eta_t = H_t^{1/2}\psi_t \quad (2)$$

$$\psi_t \sim i.i.d.(0, I_R)$$

where x_t is a $N \times 1$ vector of real valued integrated I(d) ($0 \leq d \leq 1$) and heteroskedastic processes subject to structural breaks, $t = 1, \dots, T$, in deviation from the unobserved common deterministic (μ_t) and stochastic (f_t) factors; $C(L) \equiv C_0 L^0 + C_1 L + C_2 L^2 + \dots + C_s L^s$ is a finite order matrix of polynomials in the lag operator with all the roots outside the unit circle, C_j , $j = 0, \dots, s$, is a square matrix of coefficients of order N ; v_t is a $N \times 1$ vector of zero mean idiosyncratic *i.i.d.* shocks, with contemporaneous covariance matrix Σ_v , assumed to be coherent with the condition of weak cross-sectional correlation of the idiosyncratic components (Assumption E) stated in [5] p. 143. The model in (1) actually admits the same static representation of [5], as it can be rewritten as $x_t = \Lambda_\mu \mu_t + \Lambda_f f_t + (I - C(L))^{-1} v_t$.

2.1. The Common Break Process Component

The vector of common break processes μ_t is $M \times 1$, with $M \leq N$, and $N \times M$ matrix of loadings Λ_μ ; the latter are assumed to be orthogonal to the common stochastic factors f_t , and of unknown form, measuring recurrent or non recurrent changes in mean, with smooth or abrupt transition across regimes; the generic element in μ_t is $\mu_{i,t} \equiv z_{\mu,i}(t)$, where $z_{\mu,i}(t)$, $i = 1, \dots, M$, is a function of the time index t , $t = 1, \dots, T$.

The idiosyncratic break process $z_{\mu,i}(t)$ can take different forms. For instance, [6] use a discontinuous function,

$$z_{\mu,i}(t) = \delta_{i,0} + \sum_{j=1}^J \delta_{i,j} I_{\tau_j}, \quad (3)$$

where I_{τ_j} is the indicator function, such that $I_{\tau_j} = 1$ if $t > \tau_j$ and is 0 otherwise; in [6] the break points τ_j are determined through testing; a Markov switching mechanism, as in [7], could however also be employed to

this purpose.

Differently, [8] [9] and [10] model the break process as a continuous and bounded function of time, by means of a Fourier expansion, *i.e.*,

$$z_{\mu,i}(t) = \delta_{i,0} + \sum_{j=1}^J \delta_{i,s,j} \sin(2\pi jt/T) + \delta_{i,c,j} \cos(2\pi jt/T), \quad j \leq \frac{T}{2} \quad (4)$$

Similarly [11], using a logistic specification

$$z_{\mu,i}(t) = \delta_{i,0} + \sum_{j=1}^J \delta_{i,j} g(\eta_j, c_j, t^*), \quad (5)$$

where the logistic function is $g(\eta_j, c_j, t^*) = \left(1 + \exp(-\gamma(\eta)(t^* - c_j)/\hat{\sigma}_t^*)\right)^{-1}$, $\gamma(\eta) = \exp(\eta_j)$, $c_j \in [0,1]$ and η_j are parameters, $t^* = t/T$, and $\hat{\sigma}_t^*$ is the estimated standard deviation of t^* . In particular, as $\eta_j \rightarrow \infty$, $g(\cdot)$ becomes the indicator function, yielding therefore a generalization of the specification in [6].

Also similarly [12] and [13], using a spline function

$$z_{\mu,i}(t) = S(t/T), \quad (6)$$

where $S(t/T) = \sum_{j=1}^{p+2} a_j f_j(t/T)$ is a spline function of order p , a_j are unknown regression coefficients and the functions $f_j(\cdot)$ are spline basis functions defined as $f_1 = 1$, $f_2 = (t/T)$, \dots , $f_{p+1} = (t/T)^p$, and $f_{p+2} = (t/T - \eta)^p$, with $\eta \in (1/T, 1)$.

A semiparametric approach has also been suggested by [14], using a kernel function, *i.e.*,

$$z_{\mu,i}(t) = \frac{1}{Tb} \sum_{j=1}^T K\left(\frac{t-t_j}{b}\right) x_{i,j}, \quad (7)$$

where b is the bandwidth and $K(\cdot)$ is the kernel function, specified as $K(u) = \sum_{l=0}^r \alpha_l u^{2l}$ for $|u| \leq 1$ and $K(u) = 0$ for $|u| > 1$; $r = 0,1,2,\dots$, and the coefficient α_l are such that $\int_{-1}^1 K(u) du = 1$.

Finally, a random level shift model has been proposed by [15]-[18]; for instance, [18] define the break process as

$$z_{\mu,i}(t) = \sum_{j=1}^T \delta_{T,i,j}, \quad (8)$$

where $\delta_{T,i,j} = \pi_{T,t} \eta_t$, $\eta_t \sim i.i.d.N(0, \sigma^2)$ and $\pi_{T,t} \sim i.i.d. Bernoulli(p/T, 1)$ for $p \geq 0$.

In the case $M = N$, there are no common break processes, *i.e.*, each series is characterized by its own idiosyncratic break process and the $N \times M$ factor loading matrix Λ_μ is square, diagonal and of full rank; when $M < N$, there exist M common break processes and the factor loading matrix is of reduced rank (M). Hence, in the latter case the series x_t are cotrending, according to [19], nonlinear cotrending, according to [20], or cobreaking, according to [21] and [22]. The representation in (1) emphasizes however the driving role of the common break processes, rather than the break-free linear combinations (cobreaking/cotrending relationships) relating the series x_t .

2.2. The Common Break-Free Component

The vector of (zero-mean) integrated heteroskedastic common factors f_t is $R \times 1$, with $R \leq N$, and $N \times R$ matrix of loadings Λ_f . The order of integration is d_i in mean, and b_i in variance, $0 \leq d_i \leq 1$, $0 \leq b_i \leq 1$, $i = 1, \dots, R$.

The polynomial matrix $P(L) \equiv I_R - P_1 L - P_2 L^2 - \dots - P_u L^u$ is of finite order, with all the roots outside the unit circle; P_j , $j = 1, \dots, u$, is a square matrix of coefficients of order R ; ψ_t is a $R \times 1$ vector of common zero mean i.i.d. shocks, with identity covariance matrix I_R , $E[\psi_{it} \psi_{js}] = 0$ all i, j, t, s , respectively.

The matrix $D(L)$ is a square diagonal matrix in the lag operator of order R , specified according to the integration order (in mean) of the common stochastic factors, *i.e.*, $D(L) \equiv (1-L)I_R$ for the case of $I(1)$ integration ($d_i = 1$); $D(L) \equiv I_R$ for the $I(0)$ or no integration (short memory) case ($d_i = 0$);

$$D(L) \equiv \text{diag} \left\{ (1-L)^{d_1}, (1-L)^{d_2}, \dots, (1-L)^{d_R} \right\} \text{ for the case of fractional integration } (I(d)), \text{ long memory}$$

($0 < d_i < 1$), where $(1-L)^{d_i}$ is the fractional differencing operator; the latter admits a binomial expansion, which can be compactly written in terms of the Hypergeometric function, *i.e.*,

$$(1-L)^{d_i} = F(-d_i, 1, 1; L) \equiv \sum_{k=0}^{\infty} \Gamma(k-d_i) \Gamma(k+1)^{-1} \Gamma(-d_i)^{-1} L^k \equiv \sum_{k=0}^{\infty} \pi_k L^k, \text{ where } \Gamma(\cdot) \text{ is the Gamma function.}$$

In the case $R = N$ there are no common stochastic processes, *i.e.*, each series is characterized by its own idiosyncratic persistent stochastic component, and the $N \times R$ factor loading matrix Λ_f is square, diagonal and of full rank; when $R < N$, then there exist R common stochastic processes and the factor loading matrix is of reduced rank (R). Hence, in the latter case the series x_t show common stochastic features, according to [23]. The concept of common feature is broad, encompassing the notion of cointegration ([24]), holding for the $0 < d_i \leq 1$ case. The representation in (1) emphasizes however the driving role of the common stochastic factors rather than the feature-free linear combinations (cofeature relationships) relating the series x_t .

2.3. The Conditional Variance Process

The $R \times R$ conditional variance-covariance matrix for the unconditionally and conditionally orthogonal common factors f_t is $H_t = \text{Var}(f_t | \Omega_{t-1}) \equiv \text{diag} \{ h_{1,t}, h_{2,t}, \dots, h_{R,t} \}$, where Ω_{t-1} is the information set available at time period $t-1$. Consistent with the constant conditional correlation model of [25], the i th generic element along the main diagonal of H_t is

$$m_i(L)h_{i,t} = \omega_{i,t} + n_i(L)\eta_{i,t}^2, \quad i = 1, \dots, R, \tag{9}$$

where $n_i(L) \equiv 1 - \beta_i(L) - (1 - \varphi_i(L))(1-L)^{b_i}$ for the case of fractional integration (long memory) in variance ($0 < b_i < 1$); $n_i(L) \equiv 1 - \beta_i(L) - (1 - \varphi_i(L))(1-L)$ for the case of $I(1)$ integration in variance ($b_i = 1$);

$n_i(L) \equiv 1 - \beta_i(L) - (1 - \varphi_i(L))$ for the $I(0)$ or no integration (short memory) in variance case ($b_i = 0$). In all cases $m(L) \equiv 1 - \beta_i(L)$, $\varphi_i(L) = \alpha_i(L) + \beta_i(L)$, $\alpha_i(L) \equiv \alpha_{i,1}L + \alpha_{i,2}L^2 + \dots + \alpha_{i,q}L^q$,

$\beta_i(L) \equiv \beta_{i,1}L + \beta_{i,2}L^2 + \dots + \beta_{i,p}L^p$ and all the roots of the $\alpha_i(L)$ and $\beta_i(L)$ polynomials are outside the unit circle.

The conditional variance process $h_{i,t} = \text{Var}(f_{i,t} | \Omega_{i,t-1})$, $i = 1, \dots, R$, is therefore of the *FIGARCH*(p, b_i, z) type [26], with $z = \max\{p, q\}$, or the *IGARCH*(p, q) type, for the fractionally integrated and integrated case, respectively; of the *GARCH*(p, q) type for the non integrated case. The model is however not standard as the intercept component $\omega_{i,t}$ is time-varying, allowing for structural breaks in variance; similarly to the mean part of the model, structural breaks in variance are assumed to be of unknown form, measuring recurrent or non recurrent regimes, with smooth or abrupt transition; then, $\omega_{i,t} \equiv z_{h,i}(t)$, where $z_{h,i}(t)$ is a continuous or discontinuous bounded function of the time index t , $t = 1, \dots, T$, which can be parameterized as in (3), (4), (5), (6), (7), or (8).

The following *ARCH*(∞) representation can be obtained from each of the three above models

$$h_{i,t} = \omega_{i,t}^* + \psi_i(L)\eta_{i,t}^2 \quad i = 1, \dots, R, \tag{10}$$

where $\omega_{i,t}^* = \frac{\omega_{i,t}}{m_i(1)}$ and $\psi_i(L) = \frac{n_i(L)}{m_i(L)} = \psi_{1,i}L + \psi_{2,i}L^2 + \dots$.

The term $\omega_{i,t}^*$ then bears the interpretation of break in variance process, or time-varying unconditional va-

riance process (no integration case), or long-term conditional variance level (unit root and fractional integration cases).

To guarantee the non negativity of the conditional variance process at eachpoint in time all the coefficients in the $ARCH(\infty)$ representation must be non-negative, *i.e.*, $\psi_{i,j} \geq 0$ for all $j \geq 1$ and $\omega_{i,t}^* > 0$ for any t . Sufficient conditions, for various parameterization, can be found in [26] and [27].

2.4. The Reduced Fractional VAR form

By substituting (2) into (1) and rearranging, the vector autoregressive representation for the factors f_t and the gap series $x_t - \Lambda_\mu \mu_t$ can be written as

$$\begin{bmatrix} f_t \\ x_t - \Lambda_\mu \mu_t \end{bmatrix} = \begin{bmatrix} \Pi_f^*(L) & 0 \\ \Pi^*(L) & C(L) \end{bmatrix} \begin{bmatrix} f_{t-1} \\ x_{t-1} - \Lambda_\mu \mu_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \quad (11)$$

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} I \\ \Lambda_f \end{bmatrix} \begin{bmatrix} H_t^{1/2} \psi_t \\ v_t \end{bmatrix},$$

where $\Pi^*(L) = [\Lambda_f(L)L^{-1} - C(L)\Lambda_f]$ and $\Pi_f^*(L)$ is differently defined according to persistence properties of the data. In particular, for the case of fractional integration (long memory) ($0 < d_i < 1$), by means of the binomial expansion, it follows $P(L)D(L) \equiv I - \Pi(L)$, $\Pi(L) = \Pi_1 L + \Pi_2 L^2 + \dots$, where Π_i , $i = 1, 2, \dots$, is a square matrix of coefficients of dimension R , and $\Pi_f^*(L) = \Pi(L)L^{-1}$; since the infinite order representation cannot be handled in estimation, a truncation to a suitable large lag for the polynomial matrix $\Pi(L)$ is required.² Hence, $\Pi(L) \cong \sum_{j=1}^p \Pi_j L^j$. For the case of no integration (short memory) ($d_i = 0$), recalling

that $D(L) \equiv I_R$, and therefore $P(L)D(L) = P(L)$, then $\Pi(L) = \Pi_1 L + \Pi_2 L^2 + \dots + \Pi_u L^u$; for the case of integration ($d_i = 1$), it should be firstly recalled that

$P(L)D(L) \equiv P(L)(1-L) \equiv (I_R - \rho L) - (P_1 L + P_2 L^2 + \dots + P_u L^u)(1-L)$, with $\rho = I_R$; the latter may be rewritten in the equivalent polynomial matrix form $I_R - \Gamma_1 L - \Gamma_2 L^2 - \dots - \Gamma_{u+1} L^{u+1}$, where Γ_i , $i = 1, \dots, u+1$, is a square matrix of coefficients of dimension R , and $\Gamma_1 + \Gamma_2 + \dots + \Gamma_{u+1} = \rho = I_R$, $P_i = -(\Gamma_{i+1} + \Gamma_{i+2} + \dots + \Gamma_{u+1})$, $i = 1, 2, \dots, u$; then, $\Pi(L) = P_1 L + P_2 L^2 + \dots + P_u L^u$.

Reduced Form and Structural Vector Moving Average Representation of the FI-HF-VAR Model

In the presence of unconditional heteroskedasticity, the computation of the impulse response functions and the forecast error variance decomposition (FEVD) should be made dependent on the estimated unconditional variance for each regime. In the case of (continuously) time-varying unconditional variance, policy analysis may then be computed at each point in time. For some of the conditional variance models considered in the paper, *i.e.*, the FIGARCH and IGARCH processes, the population unconditional variance does not actually exist; in the latter cases the $\omega_{i,t}$ component might bear the interpretation of long term level for the conditional variance; policy analysis is still feasible, yet subject to a different interpretation, FEVD referring, for instance, not to the proportion of forecast error (unconditional) variance accounted by each structural shock, but to the proportion of forecast error (conditional) long term variance accounted by each structural shock. With this caveat in mind, the actual computation of the above quantities is achieved in the same way as in the case of well defined population unconditional variance.

Hence, the computation of the vector moving average (VMA) representation for the FI-HF-VAR model de-
²Monte Carlo evidence reported in [28] suggests that the truncation lag should increase with the sample size and the complexity of the ARFIMA representation of the long memory process, still remaining very small relatively to the sample size. For instance, for the covariance stationary fractional white noise case and a sample of 100 observations truncation can be set as low as 6 lags, while for a sample of 10,000 observations it should be increased to 14 lags; for the case of a covariance stationary ARFIMA (1,d,1) process and a sample of 1000 observations truncation may be set to 30 lags. See [28] for further details.

depends on the persistence properties of the data. The following distinctions should then be made.

For the short memory case, *i.e.*, the zero integration order case ($d_i = 0$), the VMA representation for the factors f_t and gap series $x_t - \Lambda_\mu \mu_t$ can be written as

$$\begin{bmatrix} f_t \\ x_t - \Lambda_\mu \mu_t \end{bmatrix} = \begin{bmatrix} U(L) & 0 \\ G(L) & F(L) \end{bmatrix} \begin{bmatrix} \eta_t \\ v_t \end{bmatrix}, \quad (12)$$

where $U(L) \equiv P(L)^{-1}$, $G(L) \equiv \Lambda_f P(L)^{-1}$ and $F(L) \equiv [I - C(L)L]^{-1}$.

For the long memory case ($0 < d_i < 1$) and the case of $I(1)$ non stationarity ($d_i = 1$), the VMA representation should be computed for the differenced process, yielding

$$(1-L) \begin{bmatrix} f_t \\ x_t - \Lambda_\mu \mu_t \end{bmatrix} = \begin{bmatrix} U(L)^+ & 0 \\ G(L)^+ & F(L)^+ \end{bmatrix} \begin{bmatrix} \eta_t \\ v_t \end{bmatrix}, \quad (13)$$

where $U(L)^+ \equiv (1-L)U(L)$, $G(L)^+ \equiv (1-L)G(L)$ and $F(L)^+ \equiv (1-L)F(L)$. Impulse responses can then be computed as $I + \sum_{j=1}^k U_j^+$ for f_t and $I + \sum_{j=1}^k G_j^+$ and $I + \sum_{j=1}^k F_j^+$ for $x_t - \Lambda_\mu \mu_t$, $k = 1, 2, \dots$

The identification of the structural shocks in the FI-HF-VAR model can be implemented in two steps. Firstly, denoting by ξ_t the vector of the R structural common factor shocks, the relation between reduced and structural form common shocks can be written as $\xi_t = H\eta_t$, where H is square and invertible. Therefore, the identification of the structural common factor shocks amounts to the estimation of the elements of the H matrix. It is assumed that $E[\xi_t \xi_t'] = I_R$, and hence $H\Sigma_\eta H' = I_R$. As the number of free parameters in Σ_η is $R(R+1)/2$, at most $R(R+1)/2$ parameters in H^{-1} can be uniquely identified through the $\Sigma_\eta = H^{-1}H'^{-1}$ system of nonlinear equations in the unknown parameters of H^{-1} . Additional $R(R-1)/2$ restrictions need then to be imposed for exact identification of H^{-1} , by constraining the contemporaneous or long-run responses to structural shocks; for instance, recursive (Choleski) or non recursive structures can be imposed on the VAR model for the common factors through exclusion or linear/non-linear restrictions, as well as sign restrictions, on the contemporaneous impact matrix H^{-1} .³

Secondly, by denoting ϵ_t the vector of N structural idiosyncratic disturbances, the relation between reduced form and structural form idiosyncratic shocks can be written as $\epsilon_t = Kv_t$, where K is square and invertible. Hence, the identification of the structural idiosyncratic shocks amounts to the estimation of the elements of the K matrix. It is assumed that $E[KK'] = I_N$, and hence $K\Sigma_v K' = I_N$. Then, in addition to the $N(N+1)/2$ equations provided by $\Sigma_v = K^{-1}K'^{-1}$, $N(N-1)/2$ restrictions need to be imposed for exact identification of K^{-1} , similarly to what required for the common structural shocks.

Note that preliminary to the estimation of the Σ_v matrix, \hat{v}_t should be obtained from the residuals of an OLS regression of $\hat{\epsilon}_t$ on $\hat{\eta}_t$; the latter operation would grant orthogonality between common and idiosyncratic residuals.

The structural VMA representation can then be written as

$$\begin{bmatrix} f_t \\ x_t - \Lambda_\mu \mu_t \end{bmatrix} = \begin{bmatrix} U^*(L) & 0 \\ G^*(L) & F^*(L) \end{bmatrix} \begin{bmatrix} \xi_t \\ \epsilon_t \end{bmatrix}, \quad (14)$$

where $U^*(L) = U(L)H^{-1}$, $G^*(L) = G(L)H^{-1}$, $F^*(L) = F(L)K^{-1}$, or

$$(1-L) \begin{bmatrix} f_t \\ x_t - \Lambda_\mu \mu_t \end{bmatrix} = \begin{bmatrix} U^*(L)^+ & 0 \\ G^*(L)^+ & F^*(L)^+ \end{bmatrix} \begin{bmatrix} \xi_t \\ \epsilon_t \end{bmatrix}, \quad (15)$$

where $U^*(L)^+ = U(L)^+ H^{-1}$, $G^*(L)^+ = G(L)^+ H^{-1}$, $F^*(L)^+ = F(L)^+ K^{-1}$, according to persistence properties, and $E[\epsilon_{i,t}, \xi'_{j,t}] = 0$ any i, j .

³See [29] for a recent survey.

3. Estimation

Estimation of the model can be implemented following a multi-step procedure, consisting of persistence analysis, *QML* estimation of the common factors and VAR parameters in (1), *QML* estimation of the conditional mean model in (2) and the reduced form model in (11), *QML* estimation of the conditional variance covariance matrix in (2).

3.1. Step 1: Persistence Analysis

Each component $x_{i,t}$, $i = 1, \dots, N$, in the vector time series x_t is firstly decomposed into its purely deterministic (trend/break process; $b_{i,t}$) and purely stochastic (break-free, $l_{i,t} = x_{i,t} - b_{i,t}$) parts.

It is then assumed that the data obey the model

$$x_{i,t} = b_{i,t} + l_{i,t}, \quad t = 1, \dots, T \quad i = 1, \dots, N, \quad (16)$$

where $b_{i,t}$ and $l_{i,t}$ are orthogonal, $b_{i,t} \equiv z_{b,i}(t)$, with $z_{b,i}(t)$ a bounded function of the time index t , evolving according to discontinuous changes (step function) or showing smooth transitions across regimes.

Depending on the specification of $z_{b,i}(t)$, a joint estimate of the two components can be obtained following [7] [10] [11] [13] [14] [30], by setting up an augmented fractionally integrated ARIMA model

$$\phi(L)(1-L)^{d_i} \left((1-L)^k x_{i,t} - b_{i,t} \right) = v_{i,t}, \quad (17)$$

where $k = \{0, 1\}$ is the integer differencing parameter, d_i is the fractional differencing parameter ($-0.5 < d_i < 0.5$), $\phi(L)$ is a stationary polynomial in the lag operator and $v_{i,t}$ is a white noise disturbance.

Heteroskedastic innovations can also be considered, by specifying $v_{i,t} \equiv \sigma_{it} e_t$, with $e_t \sim i.i.d.(0, 1)$ and the conditional variance process σ_{it}^2 according to a model of the GARCH family.

Consistent and asymptotically normal estimation by means of *QML*, also implemented through iterative algorithms, is discussed in [10] [13] [14] [18] [31]. Extensions of the Markov switching [7], logistic [11] and random level shift [15]-[18] models to the long memory case have also been contributed by [32] [33] and [34], respectively.

Alternatively, following [6], a two-step procedure can be implemented: firstly, structural break tests are carried out and break points estimated; then, dummy variables are constructed according to their dating and the break process is estimated by running an OLS regression of the actual series $x_{i,t}$ on the latter dummies, as in (3); this yields $\hat{b}_{i,t}$ computed as the fitted process and the stochastic part as the estimated residual, *i.e.*, $\hat{l}_{i,t} = x_{i,t} - \hat{b}_{i,t}$; $\hat{b}_{i,t}$ and $\hat{l}_{i,t}$ are then orthogonal by construction.⁴

As neglected structural breaks may lead to processes which appear to show persistence of the long memory or unit root type, as well as spurious breaks may be detected in the data when persistence in the error component is neglected, testing procedures robust to persistence properties are clearly desirable. In this respect, the *RSS*-based testing framework in [6] yields consistent detection of multiple breaks at unknown dates for $I(0)$ processes, as well as under long range dependence [35];⁵ moreover, under long range dependence, the validity of an estimated break process (obtained, for instance, by means of [6]) may also be assessed by testing the null hypothesis of long memory in the estimated break-free series ($\hat{l}_{i,t}$), as antipersistence is expected from the removal of a spurious break process [36] [37]. Structural break tests valid for both $I(0)$ and $I(1)$ series have also recently contributed in the literature.

3.2. Step 2: Estimation of the Conditional Mean Model

QML estimation of the reduced form model in (11) is performed by first estimating the latent factors and VAR

⁴The orthogonality of $\hat{b}_{i,t}$ and $\hat{l}_{i,t}$ can however also be imposed when jointly estimating the deterministic and stochastic components by means of augmented ARFIMA models.

⁵The strong consistency of the *RSS* estimator of the break fraction, independently of the rate of decay of the autocovariance function of the error process, has been proved in [35] when the number of break points is known; a modified Bayes-Schwarz selection criterion for the number of break points is also proposed.

parameters in (1); then, by estimating the conditional mean process in (2); finally, by substituting (2) into (1) in order to obtain a restricted estimate of the polynomial matrix $\Pi^*(L)$.

3.2.1. Estimation of the Common Factors and VAR Parameters

Estimation of the common factors is performed by *QML*, writing the (misspecified) approximating model as

$$x_t - \Lambda_\mu \mu_t - \Lambda_f f_t = v_t \quad (18)$$

$$v_t \sim i.i.d.N(0, \sigma^2 I_N)$$

$$f_t \sim i.i.d.N(0, \sigma^2 I_R)$$

with log-likelihood function given by

$$l(\cdot) = -\frac{NT}{2} \ln 2\pi - \frac{T}{2} \ln |\sigma^2 I_N| - \frac{1}{2} \sum_{t=1}^T \frac{(x_t - \Lambda_\mu \mu_t - \Lambda_f f_t)' (x_t - \Lambda_\mu \mu_t - \Lambda_f f_t)}{\sigma^2}. \quad (19)$$

QML estimation of the latent factors and their loadings then requires the minimization of the objective function

$$\sum_{t=1}^T (x_t - \Lambda_\mu \mu_t - \Lambda_f f_t)' (x_t - \Lambda_\mu \mu_t - \Lambda_f f_t) \quad (20)$$

which can be rewritten as

$$\frac{1}{NT} \sum_{t=1}^T (b_t - \Lambda_\mu \mu_t)' (b_t - \Lambda_\mu \mu_t) + \frac{1}{NT} \sum_{t=1}^T (l_t - \Lambda_f f_t)' (l_t - \Lambda_f f_t), \quad (21)$$

where $x_t = l_t + b_t$, as l_t and b_t are orthogonal vectors, as well as μ_t and f_t .

The solution to the minimization problem, subject to the constraints $N^{-1} \Lambda_\mu' \Lambda_\mu = I_M$ and $N^{-1} \Lambda_f' \Lambda_f = I_R$, is given by firstly minimizing with respect to μ_t and f_t , given Λ_μ and Λ_f , yielding

$$\hat{\mu}_t \left(\Lambda_\mu (\Lambda_\mu' \Lambda_\mu)^{-1} \right) = (\Lambda_\mu' \Lambda_\mu)^{-1} \Lambda_\mu' b_t$$

$$\hat{f}_t \left(\Lambda_f (\Lambda_f' \Lambda_f)^{-1} \right) = (\Lambda_f' \Lambda_f)^{-1} \Lambda_f' l_t$$

and then concentrating the objective function to obtain

$$\frac{1}{T} \sum_{t=1}^T b_t' \left(I_N - \Lambda_\mu (\Lambda_\mu' \Lambda_\mu)^{-1} \Lambda_\mu \right) b_t + \frac{1}{T} \sum_{t=1}^T l_t' \left(I_N - \Lambda_f (\Lambda_f' \Lambda_f)^{-1} \Lambda_f \right) l_t, \quad (22)$$

which can be mimized with respect to Λ_μ and Λ_f . This is equivalent to maximizing

$$tr \left\{ (\Lambda_\mu' \Lambda_\mu)^{-1/2} \Lambda_\mu' \left(\frac{1}{T} \sum_{t=1}^T b_t b_t' \right) \Lambda_\mu (\Lambda_\mu' \Lambda_\mu)^{-1/2} \right\} + tr \left\{ (\Lambda_f' \Lambda_f)^{-1/2} \Lambda_f' \left(\frac{1}{T} \sum_{t=1}^T l_t l_t' \right) \Lambda_f (\Lambda_f' \Lambda_f)^{-1/2} \right\}, \quad (23)$$

which in turn is equivalent to maximizing

$$\Lambda_\mu \hat{\Sigma}_b \Lambda_\mu \quad (24)$$

subject to $N^{-1} \Lambda_\mu' \Lambda_\mu = I_M$, and

$$\Lambda_f \hat{\Sigma}_l \Lambda_f \quad (25)$$

subject to $N^{-1} \Lambda_f' \Lambda_f = I_R$.

The solution is then found by setting:

- $\hat{\Lambda}_\mu$ equal to the scaled eigenvectors of $\hat{\Sigma}_b$, *i.e.*, the sample variance covariance matrix of the break processes b_t , associated with its M largest eigenvalues; this yields $\hat{\mu}_t = N^{-1} \hat{\Lambda}_\mu' b_t$, *i.e.*, the scaled first M principal components of b_t ;

- $\hat{\Lambda}_f$ equal to the scaled eigenvectors of $\hat{\Sigma}_l$, *i.e.*, the sample variance covariance matrix of the break-free processes l_t , corresponding to its R largest eigenvalues; this yields $f_t' = N^{-1}\hat{\Lambda}_f' l_t$, *i.e.*, the scaled first R principal components of l_t .

Note that PCA uniquely estimates the space spanned by the unobserved factors; hence, Λ_f and f_t (Λ_μ and μ_t) are not separately identified, as the common factors $f_t(\mu_t)$ and factor loading matrix $\Lambda_f(\Lambda_\mu)$ are uniquely estimated up to a suitable invertible rotation matrix $H_f(H_\mu)$, *i.e.*, PCA delivers estimates of $f_{*,t} \equiv H_f f_t$ ($\mu_{*,t} \equiv H_\mu \mu_t$) and $\Lambda_{f*} \equiv \Lambda_f H_f^{-1}$ ($\Lambda_{\mu*} \equiv \Lambda_\mu H_\mu^{-1}$), and therefore a unique estimate of the common components $\Lambda_f f_t \equiv \Lambda_{f*} f_{*,t}$ ($\Lambda_\mu \mu_t \equiv \Lambda_{\mu*} \mu_{*,t}$) only, which is however all what is required for the computation of the gap vector.

As shown by [38], exact identification of the common factors can also be implemented, by appropriately constraining the factor loading matrix while performing PCA or after estimation. In particular, three identification structures are discussed, involving a block diagonal factor loading matrix, yield by a statistical restriction imposed in estimation, and two rotation strategies, yielding a lower triangular factor loading matrix in the former case and a two-block partitioned factor loading matrix in the latter case, with identity matrix in the upper block and an unrestricted structure in the lower block.

Moreover, the number of common factors (R, M) is unknown and needs to be determined; several criteria are available in the literature, ranging from heuristic or statistical eigenvalue-based approaches [39] [40] to the more recent information criteria [41] and “primitive” shock ([42]) based procedures.

Finally, in order to enforce orthogonality between the estimated common break processes ($\hat{\mu}_{*,t}$) and stochastic factors ($\hat{f}_{*,t}$), the above procedure may be modified by computing the stochastic component \hat{l}_t as the residuals from the OLS regression of $x_{i,t}$ on $\hat{\mu}_{*,t}$; then PCA is implemented on (the break-free residuals) \hat{l}_t to yield $\hat{f}_{*,t}$.

Estimation of the VAR parameters. Conditional on the estimated (rotated) latent factors, the polynomial matrix $C(L)$ and the $\Lambda_{f*} \equiv \Lambda_f H_f^{-1}$ and $\Lambda_{\mu*} \equiv \Lambda_\mu H_\mu^{-1}$ (rotated) factor loading matrices are obtained by means of OLS estimation of the equation system in (1). This can be obtained by first (OLS) regressing the actual series x_t on the estimated common break processes ($\hat{\mu}_{*,t}$) and stochastic factors ($\hat{f}_{*,t}$) to obtain $\hat{\Lambda}_{\mu*}$ and $\hat{\Lambda}_{f*}$; alternatively, $\hat{\Lambda}_{\mu*}$ and $\hat{\Lambda}_{f*}$ can be estimated as yield by PCA, *i.e.*, from the scaled eigenvectors of the matrices $\hat{\Sigma}_b$ and $\hat{\Sigma}_l$, respectively; then, the gap vector is computed as $x_t - \hat{\Lambda}_{\mu*} \hat{\mu}_{*,t} - \hat{\Lambda}_{f*} \hat{f}_{*,t}$, as $\hat{\Lambda}_f \hat{f}_t = \hat{\Lambda}_{f*} \hat{f}_{*,t}$ and $\hat{\Lambda}_\mu \hat{\mu}_t = \hat{\Lambda}_{\mu*} \hat{\mu}_{*,t}$, and $\hat{C}(L)$ is obtained by means of OLS estimation of the VAR model in (1).

3.2.2. Iterative Estimation of the Common Factors and VAR Parameters

The above estimation strategy may be embedded within an iterative procedure, yielding a (relatively more efficient) estimate of the latent factors and the VAR parameters in the equation system in (1).

The objective function to be minimized is then written as

$$S(\Lambda_\mu, \mu_t, \Lambda_f, f_t, C(L)) = \frac{1}{NT} \sum_{t=1}^T v_t' v_t \quad (26)$$

where $v_t = (I_N - C(L)L)(x_t - \Lambda_\mu \mu_t - \Lambda_f f_t)$.

- **Initialization.** The iterative estimation procedure requires an initial estimate of the common deterministic (μ_t) and stochastic (f_t) factors and the $C(L)$ polynomial matrix, *i.e.*, an initial estimate of the equation system in (1). The latter can be obtained as described in Section 3.2.1.
- **Updating.** An updated estimate of the equation system in (1) is obtained as follows.
 - First, a new estimate of the M (rotated) common deterministic factors, and their factor loading matrix, is obtained by the application of PCA to the (new) stochastic factor-free series

$$x_t - \hat{\Lambda}_{f*} \hat{f}_{*,t} - C(L)(x_{t-1} - \hat{\Lambda}_{f*} \hat{f}_{*,t-1} - \hat{\Lambda}_{\mu*} \hat{\mu}_{*,t-1}), \text{ yielding } \hat{\Lambda}_{\mu*}^{(new)} \text{ and } \hat{\mu}_{*,t}^{(new)}.^6$$

⁶Alternatively, $\hat{\Lambda}_{\mu*}^{(new)}$ can be obtained by regressing x_t on $\hat{\mu}_{*,t}^{(new)}$ (and the initial estimate $\hat{f}_{*,t}$), using OLS.

- Next, conditional on the new common break processes and their factor loading matrix, the new estimate of the common long memory factors is obtained from the application of PCA to (new) break-free processes $\hat{l}_t^{(new)} = x_t - \hat{\Lambda}_{\mu^*}^{(new)} \hat{\mu}_{*,t}^{(new)}$ yielding $\hat{\Lambda}_{f^*}^{(new)}$ and $\hat{f}_{*,t}^{(new)}$.⁸
- Finally, conditional on the new estimated common break processes and long memory factors, the new estimate of the gap vector $x_t - \hat{\Lambda}_{\mu^*}^{(new)} \hat{\mu}_{*,t}^{(new)} - \hat{\Lambda}_{f^*}^{(new)} \hat{f}_{*,t}^{(new)}$ is obtained, and the new estimate $\hat{C}(L)^{(new)}$ can be computed by means of OLS estimation of the VAR model in (1).
- The above procedure is iterated until convergence, yielding the final estimates $\hat{\Lambda}_{\mu^*}^{(fin)}, \hat{\mu}_{*,t}^{(fin)}, \hat{\Lambda}_{f^*}^{(fin)}, \hat{f}_{*,t}^{(fin)}$, and $\hat{C}(L)^{(fin)}$. Convergence may be assessed in various ways. For instance, the procedure may be stopped when the change in the value of the objective function is below a given threshold.⁹

3.2.3. Restricted Estimation of the Reduced Form Model

Once the final estimate of the equation system in (1) is available, the reduced VAR form in (11) is estimated as follows:

1) For the case of fractional integration (long memory) ($0 < d_i < 1$), the fractional differencing parameter is (consistently) estimated first, for each component of the (rotated) common factors vector $\hat{f}_{*,t}^{(fin)}$, yielding the estimates \hat{d}_i , $i = 1, \dots, R$, collected in $\hat{D}(L)$ matrix.

Considering then the generic element $\hat{f}_{*,t}^{(fin)}$, \sqrt{T} consistent and asymptotically normal estimation of the i th fractional differencing parameter can be obtained, for instance, by means of *QML* estimation of the fractionally integrated ARIMA model in (17); alternatively, consistent and asymptotically normal estimation can be obtained by means of the log-periodogram regression or the Whittle-likelihood function.¹⁰

Then, conditionally to the estimated fractionally differencing parameter, $\hat{P}(L)$ is obtained by means of OLS estimation of the $VAR(u)$ model for the fractionally differenced common factors $(\hat{D}(L)\hat{f}_{*,t}^{(fin)})$ in (2); hence, $I - \hat{\Pi}(L) = \hat{P}(L)\hat{D}^*(L)$, where $\hat{D}^*(L)$ is the diagonal polynomial matrix in the lag operator of order R , containing the p^* th order ($p^* > u$) truncated binomial expansion of the elements in $\hat{D}(L)$. Then,

$$\hat{\Pi}_{f^*}(L) = \hat{\Pi}(L)L^{-1} \quad \text{and} \quad \hat{\Pi}^*(L) = \hat{\Lambda}_{f^*}^{(fin)} \hat{\Pi}(L)L^{-1} - \hat{C}(L)^{(fin)} \Lambda_f^{(fin)}.$$

Alternatively, rather than by means of the two-step Box-Jenkins type of approach detailed above, *VARFIMA* estimation of the R -variate version of the model in (17) can be performed by means of Conditional-Sum-of-Squares [45], exact Maximum Likelihood [46] or Indirect [47] estimation, still yielding \sqrt{T} consistent and asymptotically normal estimates.¹¹ OLS estimation of a VAR approximation for the *VARFIMA* model has also been recently proposed in [48], which would even avoid the estimation of the fractional differencing parameter for the common stochastic factors.

For the case of no integration (short memory) ($d_i = 0$) and integration ($d_i = 1$), we also have:

2) For the case of no integration (short memory) ($d_i = 0$), $\hat{P}(L)$ is obtained by means of OLS estimation of the $VAR(u)$ model for the (rotated) common stochastic factors $(\hat{f}_{*,t}^{(fin)})$ in (2); then $\hat{\Pi}(L) = \hat{P}_1 L + \hat{P}_2 L^2 + \dots + \hat{P}_u L^u$;

3) For the $I(1)$ case ($d_i = 1$), $\hat{\Pi}(L)$ is obtained by means of OLS estimation of the $VAR(u+1)$ model in

⁷Alternatively, the new break-free process can be computed as $x_t - \hat{\Lambda}_{\mu^*} \hat{\mu}_{*,t} - C(L)(x_{t-1} - \hat{\Lambda}_{f^*} \hat{f}_{*,t-1} - \hat{\Lambda}_{\mu^*} \hat{\mu}_{*,t-1})$.

⁸Alternatively, $\hat{\Lambda}_{f^*}^{(new)}$ can be obtained by regressing x_t on $\hat{f}_{*,t}^{(new)}$ and the updated estimate $\hat{\mu}_{*,t}^{(new)}$, using OLS. This would also yield a new estimate $\hat{\Lambda}_{\mu^*}^{(new)}$ to be used in the computation of the updated gap vector.

⁹For instance, the procedure can be stopped when $c_j = \frac{S(\hat{\theta}^{(j+1)}) - S(\hat{\theta}^{(j)})}{S(\hat{\theta}^{(j+1)}) + S(\hat{\theta}^{(j)})} < -10^{-4}$, where the objective function is written as in (26).

¹⁰See [43] and [44] for a survey of alternative estimators of the fractional differencing parameter.

¹¹Depending on the parametric structure, system estimation may however become unfeasible when the number of factors is too large.

levels for the (rotated) common stochastic factors $\left(\hat{f}_{*,t}^{(fin)}\right)$ implied by (2); then,

$$\hat{\Pi}(L) = \hat{\Gamma}_1 L + \hat{\Gamma}_2 L^2 + \dots + \hat{\Gamma}_{u+1} L^{u+1}.$$

Consistent with [49] and [50], in all of the above cases VAR estimation can be performed as the estimated common factors were actually observed.

Following the thick modelling strategy in [51], median estimates of the parameters of interest, impulse responses and forecast error variance decomposition, as well as their confidence intervals, can be computed through simulation.

3.3. Step 3: Estimation of the Conditional Variance-Covariance Matrix

The estimation of the conditional variance-covariance matrix for the factors in (2) can be carried out using a procedure similar to the O-GARCH model of [52]:

1) Firstly, conditional variance estimation is carried out factor by factor, using the estimated factor residuals $\hat{\eta}_t$, yielding \hat{h}_i , $i = 1, 2, \dots, R$; *QML* estimation can be performed in a variety of settings, ranging from standard *GARCH*(p, q) and *FIGARCH*(p, b, z) models to their “adaptive” generalizations [9] [12] [53] [54], in order to allow for different sources of persistence in variance;

2) Secondly, consistent with the assumption of conditional and unconditional orthogonality of the factors, the conditional variance-covariance $(H_{x,t})$ and correlation $(R_{x,t})$ matrices for the actual series may be estimated as

$$\hat{H}_{x,t} = \hat{\Lambda}_f \hat{H}_t \hat{\Lambda}_f = \hat{\Sigma}_v \quad (27)$$

$$\hat{R}_{x,t} = \hat{H}_{x,t}^{*-1/2} \hat{H}_{x,t} \hat{H}_{x,t}^{*-1/2} \quad (28)$$

where $\hat{H}_t = \text{diag}\{\hat{h}_{1,t}, \hat{h}_{2,t}, \dots, \hat{h}_{R,t}\}$, and $\hat{H}_{x,t}^* = \text{diag}\{\hat{h}_{x_{1,t}}, \hat{h}_{x_{2,t}}, \dots, \hat{h}_{x_{N,t}}\}$.

Relaxing the assumption of conditional orthogonality of the factors is also feasible in the proposed framework, as the dynamic conditional covariances, *i.e.*, the off-diagonal elements in H_t , can be obtained, after step *i*) above, by means of the second step in the estimation of the Dynamic Conditional Correlation model [55] or the Dynamic Equicorrelation model [56].

3.4. Asymptotic Properties

The proposed iterative procedure for the system of equations in (1) bears the interpretation of *QML* estimation, using a Gaussian likelihood function, performed by means of the *EM* algorithm. In the *E*-step, the unobserved factors are estimated, given the observed data and the current estimate of model parameters, by means of *PCA*; in the *M*-step the likelihood function is maximized (OLS estimation of the $C(L)$ matrix is performed) under the assumption that the unobserved factors are known, conditioning on their *E*-step estimate. Convergence to the one-step *QML* estimate is ensured, as the value of the likelihood function is increased at each step [57] [58]. The latter implementation of the *EM* algorithm follows from considering the estimated factors by *PCA* as they were actually observed. In fact, the *E*-step would also require the computation of the conditional expectation of the estimated factors, which might be obtained, for instance, by means of Kalman smoothing [59] [60]. As shown by [49] and [50], however, when the unobserved factors are estimated by means of *PCA* in the *E*-step, the generated regressors problem is not an issue for consistent estimation in the *M*-step, due to faster vanishing of the estimation error, provided $\sqrt{T}/N \rightarrow 0$ for linear models, and $T^{5/8}/N \rightarrow 0$ for (some classes of) non linear models, *i.e.*, the factors estimated by means of *PCA* can be considered as they were actually observed, therefore not requiring a Kalman smoothing step.

Note also that the Expectation step of the *EM* algorithm relies on consistent estimation of the unobserved components. In this respect, under general conditions, $\min\{\sqrt{N}, \sqrt{T}\}$ consistency and asymptotic normality of *PCA*, at each point in time, for the unobserved common components $\Lambda_f f_t$, has been established by [5] and [61] for $N, T \rightarrow \infty$ and the case of $I(0)$ and $I(1)$ unobserved components;¹² this implies

the consistent estimation of the gap vector $x_t - \Lambda_\mu \mu_t - \Lambda_f f_t$ at the same $\min\{\sqrt{N}, \sqrt{T}\}$ rate, for $N, T \rightarrow \infty$, as well. Based on the results for $I(0)$ and $I(1)$ processes, the same properties can be conjectured also for the intermediate cases of long memory and (linear/nonlinear) trend stationarity; supporting Monte Carlo evidence is actually provided by [63] and in this study.¹³

Moreover, likewise in the Maximization step of the *EM* algorithm, \sqrt{T} consistent and asymptotically normal estimation of the polynomial matrix $C(L)$ is yielded by OLS estimation of the *VAR* model for the $I(0)$ gap vector $x_t - \Lambda_\mu \mu_t - \Lambda_f f_t$, which, according to the results in [49] and [50], can be taken as it were actually observed in the implementation of the iterative estimation procedure.

Similarly, \sqrt{T} consistent and asymptotically normal estimation of the block of equations in (2) is obtained by means of OLS estimation of the conditional mean process first, holding the estimated latent factors as they were observed, still relying on the results in [49] and [50] and on a consistent estimate of the fractional differencing parameter if needed, and then performing *QML* estimation of the conditional variance-covariance matrix.

4. Monte Carlo Analysis

Consider the following data generation process (DGP) for the $N \times 1$ vector process x_t

$$\begin{aligned} x_t - \Lambda_\mu \mu_t - \Lambda_f f_t &= C(x_{t-1} - \Lambda_\mu \mu_{t-1} - \Lambda_f f_{t-1}) + v_t \\ v_t &\sim i.i.d.(0, \sigma^2 I_N), \end{aligned} \quad (29)$$

where C is a $N \times N$ matrix of coefficients, Λ_μ and Λ_f are $N \times 1$ vectors of loadings, and μ_t and f_t are the common deterministic and long memory factors, respectively, at time period t , with

$$(1 - \phi L)(1 - L)^d f_t = \eta_t. \quad (30)$$

Then, for the conditionally heteroskedastic case it is assumed

$$\eta_t = \sqrt{h_t} \psi_t, \quad \psi_t \sim i.i.d.(0, 1)$$

$$[1 - \alpha L - \beta L](1 - L)^b (\eta_t^2 - \sigma_\eta^2) = [1 - \beta L](\eta_t^2 - h_t),$$

while

$$\eta_t \sim i.i.d.(0, 1)$$

¹²In particular, under some general conditions, given any invertible matrix Ξ , \sqrt{N} consistency and asymptotic normality of PCA for Ξf_t , at each point in time, is established for $N, T \rightarrow \infty$ and $\sqrt{T}/N \rightarrow 0$ and the case of $I(0)$ unobserved factors and idiosyncratic components, the latter also displaying limited heteroskedasticity in both their time-series and cross-sectional dimensions [5]; for $N, T \rightarrow \infty$ and $N, T^3 \rightarrow 0$ and the case of $I(1)$ (non cointegrated) unobserved factors and $I(0)$ idiosyncratic components, similarly showing limited heteroskedasticity in both the time-series and cross-sectional dimensions ([61]). The latter result is actually obtained by applying PCA to the level of the series, rather than their first differences. Moreover, for both the $I(0)$ and $I(1)$ case, \sqrt{T} consistency and asymptotic normality of PCA for $\Lambda_f \Xi^{-1}$ is established under the same conditions, as well as $\min\{\sqrt{N}, \sqrt{T}\}$ consistency and asymptotic normality of PCA for the unobserved common components $\Lambda_f f_t$, at each point in time, for $N, T \rightarrow \infty$. The conditions for consistency and asymptotic normality reported in [6] and [61] implicitly cover also the case in which PCA is implemented using the *estimated* break (\hat{b}_t) and break-free $(\hat{l}_t = x_t - \hat{b}_t)$ components, rather than the observed x_t series; in fact, by assuming $\hat{b}_t = b_t + e_{b,t}$ and $\hat{l}_t = l_t + e_{l,t}$, then $\hat{b}_t = \Lambda_\mu \mu_t + e_{b,t}$ and $\hat{l}_t = \Lambda_f f_t + e_{l,t}$, which are static factor structures as assumed in [5] and [61]. It appears that assumption E in [5], page 143, *i.e.*, weak dependence and limited cross-sectional correlation, holding for both noise (estimation error) components $e_{b,t}$ and $e_{l,t}$, augmented with the assumption of their contemporaneous orthogonality, *i.e.* $E[e_{b,t} e'_{l,t}] = 0$, is then sufficient for the validity of PCA also when implemented on noisy data. In this respect PCA acts as noise suppressor: intuitively, PCs associated with the smallest eigenvalues are noise, which should be neglected when estimating the common factors. PCA estimation of the signal component can actually be shown to be optimal in terms of minimum mean square error [62].

¹³The use of PCA for the estimation of common deterministic trends has previously been advocated by [64]. See also [65] for applications to nonstationary data.

for the conditionally homoskedastic case.

Different values for the autoregressive idiosyncratic parameter ρ , common across the N cross-sectional units ($C = \rho I_N$), have been considered, *i.e.*, $\rho = \{0, 0.2, 0.4, 0.6, 0.8\}$, as well as for the fractionally differencing parameter $d = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ and the common factor autoregressive parameter ϕ , setting $\phi = \{0.2, 0.4, 0.6, 0.8\}$ for the non integrated case and $\phi = \{0, d/2\}$ for the fractionally integrated and integrated cases; $\phi > \rho$ is always assumed in the experiment. For the conditional variance equation it is assumed $\alpha = 0.05$ and $\beta = 0.90$ for the short memory case, and $\alpha = 0.05$, $\beta = 0.30$ and $b = 0.45$ for the long memory case. The inverse signal to noise ratio $(s/n)^{-1}$ is given by σ^2/σ_η^2 , taking values

$(s/n)^{-1} = \{4, 2, 1, 0.5, 0.25\}$. Finally, Λ_μ and Λ_f are set equal to unitary vectors.

Moreover, in addition to the structural stability case, *i.e.*, $\mu_t = \mu = 0$, two designs with breaks have been considered for the component μ_t , *i.e.*,

1) Single step change in the intercept at the midpoint of the sample case, *i.e.*,

$$\mu_t = \begin{cases} 0, & t = 1, \dots, T/2 \\ 4, & t = T/2 + 1, \dots, T \end{cases}$$

2) The two step changes equally spaced throughout the sample case, with the intercept increasing at one third of the way through the sample and then decreasing at a point two thirds of the length of the sample, *i.e.*,

$$\mu_t = \begin{cases} 0, & t = 1, \dots, T/3 \\ 4, & t = T/4 + 1, \dots, 2T/3. \\ 2, & t = 2T/3 + 1, \dots, T \end{cases}$$

The sample size investigated is $T = 100,500$, and the number of cross-sectional units is $N = 30$. For the no breaks case also other cross-sectional sample sizes have been employed, *i.e.*, $N = 5, 10, 15, 50$. The number of replications has been set to 2,000 for each case.

The performance of the proposed multi-step procedure has then been assessed with reference to the estimation of the unobserved common stochastic and deterministic factors, and the ϕ and ρ autoregressive parameters. Concerning the estimation of the common factors, the Theil's inequality coefficient (IC) and the correlation coefficient ($Corr$) have been employed in the evaluation, *i.e.*,

$$IC = \sqrt{\frac{1}{T} \sum_{t=1}^T (z_t - \hat{z}_t)^2} / \left(\sqrt{\frac{1}{T} \sum_{t=1}^T z_t^2} + \sqrt{\frac{1}{T} \sum_{t=1}^T \hat{z}_t^2} \right)$$

$$Corr = \text{cov}(z_t, \hat{z}_t) / \sqrt{\text{Var}(z_t) \text{Var}(\hat{z}_t)},$$

where $z_t = \mu_t, f_t$ is the population unobserved component and \hat{z}_t its estimate. The above statistics have been computed for each Monte Carlo replication and then averaged.

In the Monte Carlo analysis, the location of the break points and the value of the fractional differencing parameter are taken as known, in order to focus on the assessment on the estimation procedure contributed by the paper; the break process is then estimated by means of the OLS regression approach in [6]. The Monte Carlo evidence provided is then comprehensive concerning the no-breaks $I(0)$ and $I(1)$ cases, as well as the no-break $I(d)$ case, concerning the estimation of the common stochastic factor. A relative assessment of the various methodologies which can be employed for the decomposition into break and break-free components is however of interest and left for further research.

4.1. Results

The results for the non integration case are reported in **Figure 1**, **Figure 2** (and **5**, columns 1 and 3), while **Figure 3**, **Figure 4** (and **5**, columns 2 and 4) refer to the fractionally integrated and integrated cases (the integrated case, independent of the type of integration, thereafter). In all cases, results refer to the estimated parameters for the first equation in the model. Since the results are virtually unaffected by the presence of condi-

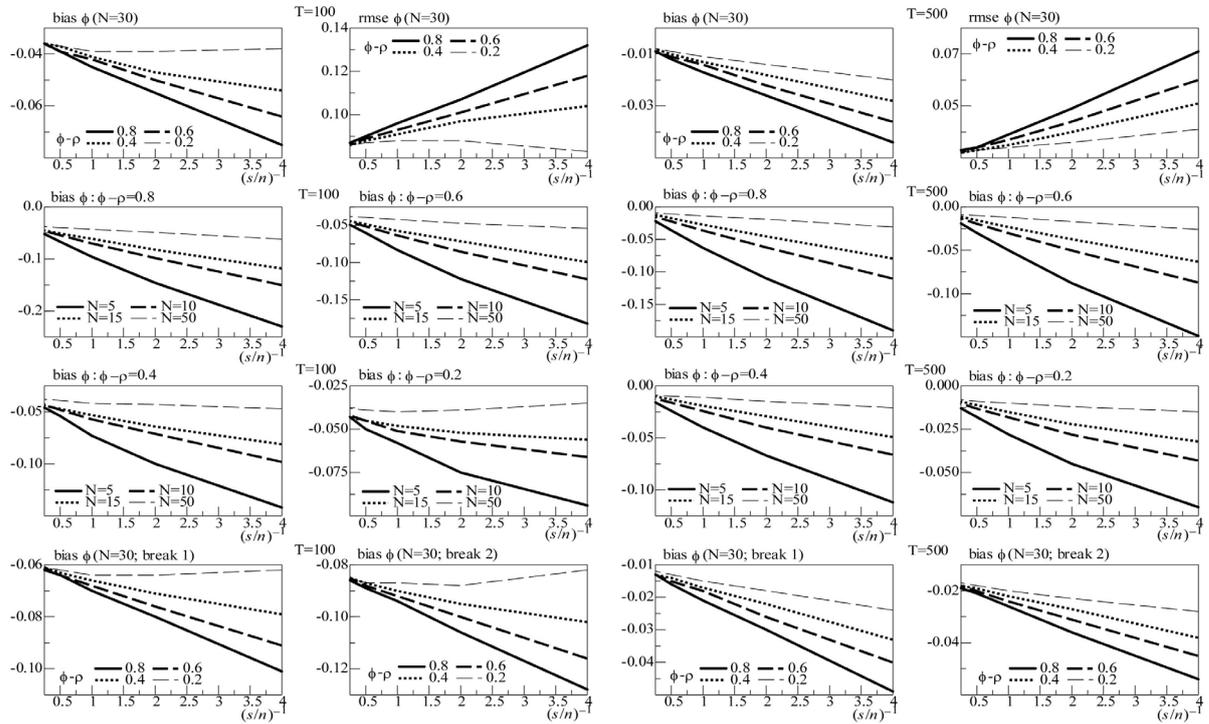


Figure 1. In the figure, Monte Carlo bias and RMSE statistics for the autoregressive parameter (ϕ) are plotted for the case of no breaks (top and center plots) and one (break 1) and two (break 2) breaks (bottom plots), and a conditionally heteroskedastic common I(0) factor. Results are reported for various values of the persistence spread $\phi - \rho$ (0.2, 0.4, 0.6, 0.8) against various values of the (inverse) signal to noise ratio $(s/n)^{-1}$ (4, 2, 1, 0.5, 0.25). The sample size T is 100 and 500 observations, the number of cross-sectional units N is 30, and the number of replications for each case is 2000. For the no breaks case, Monte Carlo bias statistics are also reported for other sample sizes N (5, 10, 15, 50) (center plots).

tional heteroskedasticity, for reasons of space, only the heteroskedastic case is discussed. Moreover, only the results for the $\phi = d/2$ case are reported for the integrated case, as similar results have been obtained for the $\phi = 0$ case.¹⁴

4.1.1. The Structural Stability Case

As shown in **Figure 5** (top plots 1-4), for a cross-sectional sample size $N = 30$ units, a negligible downward bias for the ρ parameter (on average across (inverse) signal to noise ratio values) can be noted (-0.02 and -0.03 , for the non integrated and integrated case, respectively, and $T = 100$ (top plots 1-2); -0.01 and -0.006 , respectively, and $T = 500$ (top plots 3-4)), decreasing as the serial correlation spread, $\phi - \rho$ or $d - \rho$, or the sample size T increase.

On the contrary, as shown in **Figure 1** and **Figure 3** (top plots 1 and 3), the downward bias in ϕ is increasing with the degree of persistence of the common factor d , the (inverse) signal to noise ratio s/n^{-1} , and the serial correlation spread, $\phi - \rho$ or $d - \rho$, yet decreasing with the sample size T .

For the non integrated case (**Figure 1**, plots 1 and 3), there are only few cases ($\phi - \rho = 0.4, 0.6, 0.8$) when a 10%, or larger, bias in ϕ is found, occurring when the series are particularly noisy ($s/n^{-1} = 4$); for the stationary long memory case a 10% bias, or smaller, is found for $s/n^{-1} \geq 2$, while for the non stationary long memory case for $s/n^{-1} \geq 1$ and a (relatively) large sample ($T = 500$) (**Figure 3**, plots 1 and 3). Increasing the cross-sectional dimension N yields improvements (see the next section).

Also, as shown in **Figure 2** and **Figure 4** (top plots 1-4), very satisfactory is the estimation of the unobserved common stochastic factor, as the IC statistic is always below 0.2 (0.14 (0.10)), on average, for $T = 100$ ($T = 500$) for the non integrated case (**Figure 2**, top plots 2 and 4); 0.06 (0.03), on average, for $T = 100$ ($T = 500$) for the integrated case (**Figure 4**, top plots 2 and 4). Moreover, the correlation coefficient between

¹⁴Detailed results are available in the working paper version of this paper [66] or upon request from the author.

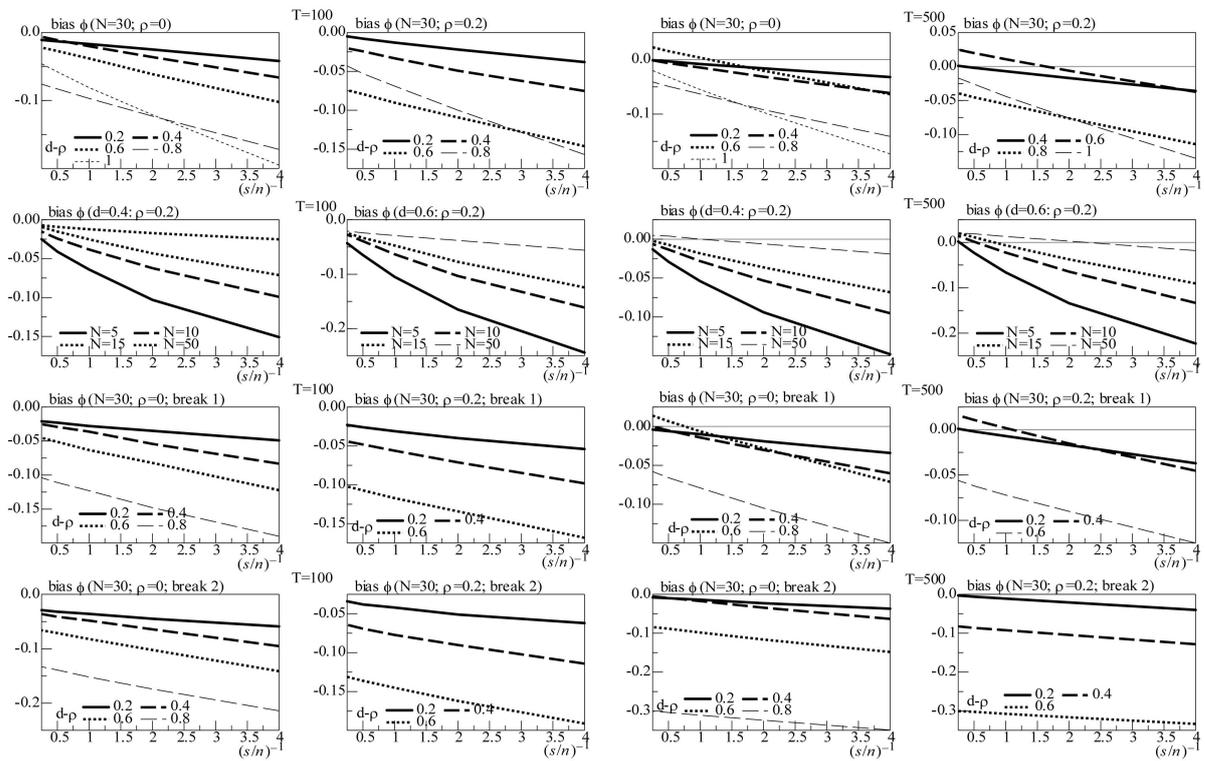


Figure 2. In the figure, Monte Carlo Theil’s index (IC) and correlation coefficient (Corr) statistics, concerning the estimation of the conditionally heteroskedastic common I(0) factor, are plotted for the case of no breaks (top and center plots) and one (break 1) and two (break 2) breaks (bottom plots). Results are reported for various values of the persistence spread $\phi-\rho$ (0.2, 0.4, 0.6, 0.8) against various values of the (inverse) signal to noise ratio $(s/n)^{-1}$ (4, 2, 1, 0.5, 0.25). The sample size T is 100 and 500 observations, the number of cross-sectional units N is 30, and the number of replications for each case is 2000. For the no breaks case, Monte Carlo correlation coefficient statistics are also reported for other sample sizes N (5, 10, 15, 50) (center plots).

the actual and estimated common factors is always very high, 0.98 and 0.99, on average, respectively, for both sample sizes (Figure 2 and Figure 4, top plots 1 and 3).

Results for smaller and larger cross-sectional samples. In Figures 1-4 (center plots, *i.e.*, rows 2 and 3), the bias for the ϕ parameter and the correlation coefficient between the actual and estimated common factors are also plotted for different cross-sectional dimensions, *i.e.*, $N = 5, 10, 15, 50$, for the non integrated and integrated cases, respectively; statistics for the ρ parameter are not reported, as the latter is always unbiasedly estimated, independently of the cross-sectional dimension.

As is shown in the plots, the performance of the estimator crucially depends on T , N , and s/n^{-1} .

For the non integrated case (Figure 1), when the (inverse) signal to noise ratio is low, *i.e.*, $s/n^{-1} \leq 0.5$, the downward bias is already mitigated by using a cross-sectional sample size as small as $N = 5$, for $T = 100$; as N increases, similar results are obtained for higher s/n^{-1} , *i.e.*, $N = 10, 15$ and $s/n^{-1} \leq 1$, or $N = 50$ and $s/n^{-1} \leq 4$ (center plots, column 1-2). For a larger sample size, *i.e.*, $T = 500$ (center plots, column 3-4), similar conclusions hold, albeit for the $N = 5$ the (inverse) signal to noise ratio can be higher, *i.e.*, $s/n^{-1} \leq 0.5$; similarly for $N = 10, 15$ with $s/n^{-1} \leq 2$.

For the integrated case (Figure 3) conditions are slightly more restrictive; in particular, for the stationary long memory case, when the (inverse) signal to noise ratio is low, *i.e.*, $s/n^{-1} \leq 0.5$, the downward bias is already mitigated by setting $N = 5$ and $T = 100$; similar results are obtained for higher s/n^{-1} and N , *i.e.*, $N = 10, 15$ and $s/n^{-1} \leq 1, 2$, or $N = 50$ and $s/n^{-1} \leq 4$ (center plots, column 1-2). Similar conclusions can be drawn for $T = 500$ (center plots, column 3-4), albeit, holding N constant, accurate estimation is

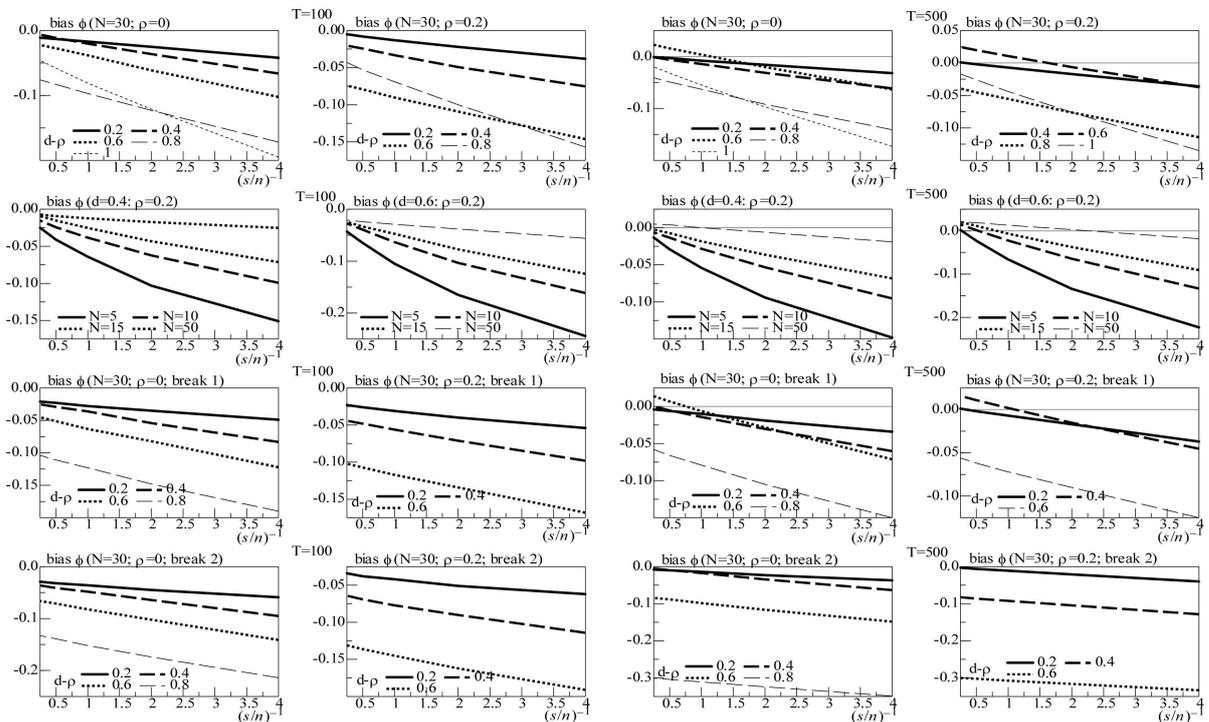


Figure 3. In the figure, Monte Carlo bias statistics for the autoregressive parameter (ϕ) are plotted for the case of no breaks (top and center plots) and one (break 1) and two (break 2) breaks (center and bottom plots), and a conditionally heteroskedastic common $I(d)$ factor ($0 < d \leq 1$). Results are reported for various values of the persistence spread $d - \rho$ (0.2, 0.4, 0.6, 0.8, 1) against various values of the (inverse) signal to noise ratio $(s/n)^{-1}$ (4, 2, 1, 0.5, 0.25). The sample size T is 100 and 500 observations, the number of cross-sectional units N is 30, and the number of replications for each case is 2000. For the no breaks case, Monte Carlo bias statistics are also reported for other sample sizes N (5, 10, 15, 50) (center plots).

obtained also for higher s/n^{-1} . Similarly also for the non stationary case (long memory or $I(1)$); yet, holding T constant, either larger N , or lower s/n^{-1} , would be required for accurate estimation.

Coherently, the correlation coefficients between the actual and estimated common factors (Figure 2 and Figure 4, center plots) point to satisfactory estimation (a correlation coefficient higher than 0.9) also in the case of a small temporal sample size, provided the (inverse) signal to noise ratio is not too high, and/or the cross-sectional dimension is not too low $s/n^{-1} \leq 1$ and $N = 5$; $s/n^{-1} \leq 2$ and $N = 10$; $s/n^{-1} \leq 4$ and $N = 15$).

4.1.2. The Structural Change Case

While concerning the estimation of the ρ parameter no sizable differences can be found for the designs with structural change, relatively to the case of structural stability¹⁵, the complexity of the break process may on the other hand affect estimation accuracy for the ϕ parameter, worsening as the number of break points increases, particularly when the temporal sample size is small ($T = 100$).

Yet, for the no integration case (Figure 1, bottom plots), already for $T = 500$ the performance is very satisfactory for both designs, independently of the (inverse) signal to noise ratio s/n^{-1} (bottom plots, columns 3 and 4); on the contrary, for $T = 100$ the performance is satisfactory (at most a 10% bias) only when the series are not too noisy ($s/n^{-1} \leq 1$) (bottom plots, columns 1 and 2). Also, similar to the structural stability case, the (downward) bias in the ϕ parameter is increasing with the degree of persistence of the common factor d , the (inverse) signal to noise ratio s/n^{-1} , and $\phi - \rho$ or $d - \rho$, yet decreasing with the sample size T .

Coherent with the above results, satisfactory estimation of the unobserved common stochastic factor (Figure 2, bottom plots) and break process can also be noted (Figure 5, bottom plots, columns 1 and 3); for

¹⁵The average bias is -0.04 and -0.01 , independent of the break process design and integration properties, when $T = 100$ and $T = 500$, respectively. Moreover, similar to the structural stability case the bias is decreasing as $\phi - \rho$, $d - \rho$, or the sample size T increase, independent of the (inverse) signal to noise ratio.

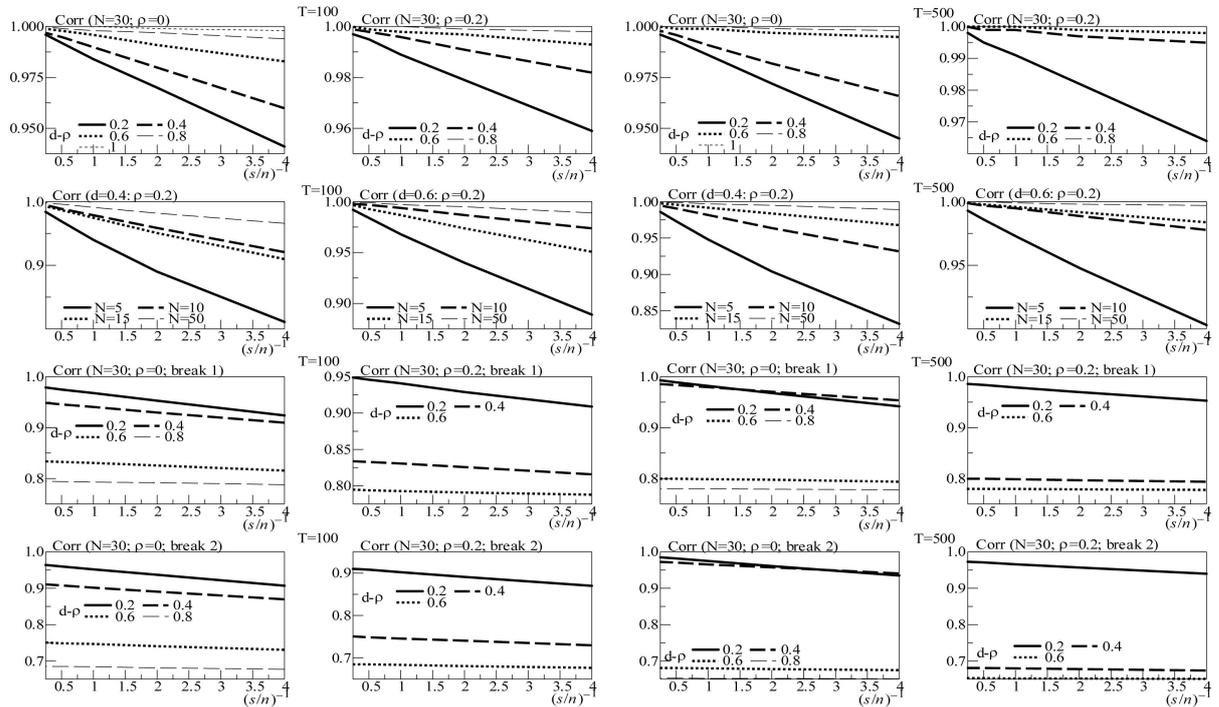


Figure 4. In the figure, Monte Carlo correlation coefficient (Corr) statistics, concerning the estimation of the conditionally heteroskedastic common $I(d)$ factor ($0 < d \leq 1$), are plotted for the case of no breaks (top and center plots) and one (break 1) and two (break 2) breaks (bottom plots). Results are reported for various values of the persistence spread $d - \rho$ (0.2, 0.4, 0.6, 0.8, 1) against various values of the (inverse) signal to noise ratio $(s/n)^{-1}$ (4, 2, 1, 0.5, 0.25). The sample size T is 100 and 500 observations, the number of cross-sectional units N is 30, and the number of replications for each case is 2000. For the no breaks case, Monte Carlo correlation coefficient statistics are also reported for other sample sizes N (5, 10, 15, 50) (center plots).

the common stochastic factor, the IC statistic (not reported) is in fact always below 0.2 for $T = 500$ (0.11 and 0.13, on average, for the single break point and two-break points case, respectively) and below 0.3 for $T = 100$ (0.17 and 0.20, on average; column 1), while the actual and estimated common stochastic factors are strongly correlated: for $T = 100$ ($T = 500$), on average, the correlation coefficient is 0.96 (0.98) for the single breakpoint case and 0.93 (0.97) for the two-break points case (column 3).

Very accurate is also the estimation of the common break process: the IC statistic is never larger than 0.15 for $T = 100$ and 0.075 for $T = 500$ (Figure 5, bottom plots, columns 1 and 3), while the correlation coefficient is virtually 1 for the single break case and never below 0.96 for $T = 100$ and 0.99 for $T = 500$ for the two-break points case (not reported). Given the assumption of known break points, the performance in terms of correlation coefficient is not surprising; yet, the very small Theil's index is indicative of success in recovering the changing level of the unobserved common break process.

Concerning the integrated case, some differences relatively to the nonintegrated case can be noted; as shown in Figure 5 (bottom plots, columns 2 and 4), albeit the overall recovery of the common break process is always very satisfactory across the various designs, independently of the sample size (the IC statistic is never larger than 0.14; bottom plots), performance slightly worsens as the complexity of the break process and persistence intensity (d) increase: the average correlation coefficient between the estimated and actual break processes (center plots) falls from 1 when $d = 0.2$ (single break point case) to 0.93 when $d = 1$ (two-break points case).

Moreover, concerning the estimation of the common stochastic factor (Figure 4, center and bottom plots, columns 1-4), for the covariance stationary case ($d < 0.5$) results are very close to the non integrated case, *i.e.*, an IC statistic (not reported) always below 0.2 for $T = 500$ (0.12 and 0.14, on average, for the single break point and two-break points case, respectively) and below 0.3 for $T = 100$ (0.21 and 0.24, on average, respectively); the correlation coefficient is also very high: 0.94 and 0.91, on average, $T = 100$ (columns 1

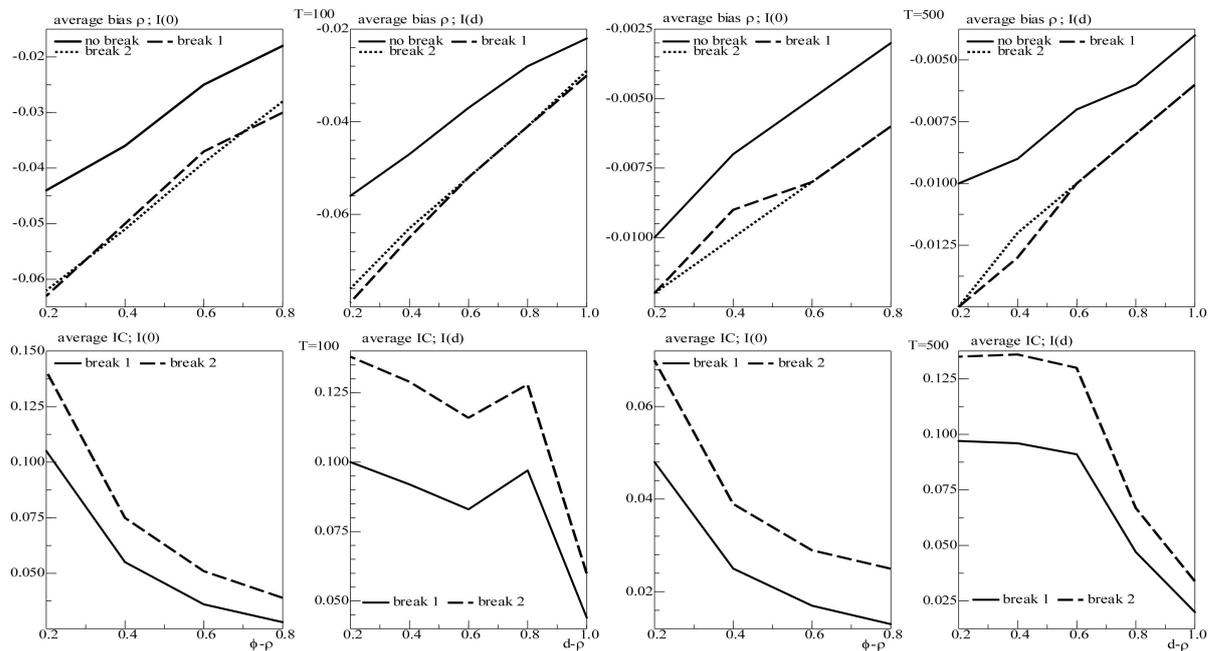


Figure 5. In the figure, average Monte Carlo statistics (across values for the inverse signal to noise ratio) for the bias in the autoregressive idiosyncratic parameter (ρ) (top plots) and Theil's index (IC) statistic for the common break process (bottom plots) are plotted for the non integrated ($I(0)$) and integrated ($I(d)$, $0 < d \leq 1$) cases. Results are reported for various values of the persistence spreads $\phi - \rho$ (0.2, 0.4, 0.6, 0.8) and $d - \rho$ (0.2, 0.4, 0.6, 0.8, 1). The sample size T is 100 and 500 observations, the number of cross-sectional units N is 30, and the number of replications for each case is 2000.

and 2); 0.97 and 0.96, on average, $T = 500$ (columns 3 and 4).

On the contrary, for the non stationary case performance is worse, showing average IC statistics (not reported) of 0.32 (0.32) and 0.42 (0.44), respectively, for the single break point (center plots) and two-break points (bottom plots) case and $T = 100$ ($T = 500$); the average correlation coefficient is 0.79 (0.78) and 0.68 (0.66), respectively. Coherently, a worsening in the estimation of the common factor autoregressive parameter ϕ , for the $d = 0.8$ and $d = 1$ case, can be noted (Figure 3, center and bottom plots), while comparable results to the short memory case can be found for $d < 0.5$. The latter findings are however not surprising, as the stronger the degree of persistence of the stochastic component (and of the series, therefore) and the less accurate the disentangling of the common break and break-free parts can be expected; overall, Monte Carlo results point to accurate decompositions also for the case of moderate nonstationary long memory, albeit deterioration in performance becomes noticeable.

5. Conclusion

In the paper, a general strategy for large-scale modeling of macroeconomic and financial data, set within the factor vector autoregressive model (F-VAR) framework is introduced. The proposed approach shows minimal pretesting requirements, performing well independently of integration properties of the data and sources of persistence, *i.e.*, deterministic or stochastic, accounting for common features of different kinds, *i.e.*, common integrated (of the fractional or integer type) or non integrated stochastic factors, also heteroskedastic, and common deterministic break processes. Consistent and asymptotically normal estimation is performed by means of QML , implemented through an iterative multi-step algorithm. Monte Carlo results strongly support the proposed approach. Empirical implementations can be found in [37] [67]-[69], showing the approach being easy to implement and effective also in the case of very large systems of dynamic equations.

Acknowledgements

A previous version of the paper was presented at the 19th and 21st Annual Symposium of the Society for Non Linear Dynamics and Econometrics, the 4th and 6th Annual Conference of the Society for Financial Econome-

trics, the 65th European Meeting of the Econometric Society (ESEM), the 2011 NBER-NSF Time Series Conference, the 5th CSDA International Conference on Computational and Financial Econometrics. The author is grateful to conference participants, N. Cassola, F. C. Bagliano, C. Conrad, R. T. Baillie, J. Bai, for constructive comments. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 3202782013-2015 and PRIN-MIUR 2009.

As Mitsuo Aida wrote in one of his poems, *somewhere in life/there is a path/that must be taken regardless of how hard we try to avoid it/at that time, all one can do is remain silent and walk the path/neither complaining nor whining/saying nothing and walking on/just saying nothing and showing no tears/it is then/as human beings, /that the roots of our souls grow deeper*. This paper is dedicated to the loving memory of A.

References

- [1] Stock, J.H. and Watson, M.W. (2011) Dynamic Factor Models. In: Clements, M.P. and Hendry, D.F., Eds., *Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford, 35-60.
- [2] Geweke, J. (1977) The Dynamic Factor Analysis of Economic Time Series. In: Aigner, D.J. and Goldberger, A.S., Eds., *Latent Variables in Socio-Economic Models 1*, North-Holland, Amsterdam.
- [3] Dees, S., Pesaran, M.H., Smith, L.V. and Smith, R.P. (2010) Supply, Demand and Monetary Policy Shocks in a Multi-Country New Keynesian Model. ECB Working Paper Series, No. 1239.
- [4] Bai, J. and Ng, S. (2004) A Panic Attack on Unit Roots and Cointegration. *Econometrica*, **72**, 1127-1177.
- [5] Bai, J.S. (2003) Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, **71**, 135-171. <http://dx.doi.org/10.1111/1468-0262.00392>
- [6] Bai, J.S. and Perron, P. (1998) Testing for and Estimation of Multiple Structural Changes. *Econometrica*, **66**, 47-78. <http://dx.doi.org/10.2307/2998540>
- [7] Hamilton, J.D. (1989) A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, **57**, 357-384. <http://dx.doi.org/10.2307/1912559>
- [8] Enders, W. and Lee, J. (2012) The Flexible Fourier Form and Dickey-Fuller Type Unit Root Tests. *Economics Letters*, **117**, 196-199. <http://dx.doi.org/10.1016/j.econlet.2012.04.081>
- [9] Baillie, R.T. and Morana, C. (2009) Modeling Long Memory and Structural Breaks in Conditional Variances: An Adaptive FIGARCH Approach. *Journal of Economic Dynamics and Control*, **33**, 1577-1592. <http://dx.doi.org/10.1016/j.jedc.2009.02.009>
- [10] Baillie, R.T. and Morana, C. (2012) Adaptive ARFIMA Models with Applications to Inflation. *Economic Modelling*, **29**, 2451-2459. <http://dx.doi.org/10.1016/j.econmod.2012.07.011>
- [11] González, A. and Teräsvirta, T. (2008) Modelling Autoregressive Processes with a Shifting Mean. *Studies in Nonlinear Dynamics and Econometrics*, **12**, 1558-3708. <http://dx.doi.org/10.2202/1558-3708.1459>
- [12] Engle, R.F. and Rangel, J.C. (2008) The Spline-GARCH Model for Low Frequency Volatility and Its Global Macroeconomic Causes. *Review of Financial Studies*, **21**, 1187-1222. <http://dx.doi.org/10.1093/rfs/hhn004>
- [13] Beran, J. and Weiershauser, A. (2011) On Spline Regression under Gaussian Subordination with Long Memory. *Journal of Multivariate Analysis*, **102**, 315-335. <http://dx.doi.org/10.1016/j.jmva.2010.09.007>
- [14] Beran, J. and Feng, Y.H. (2002) SEMIFAR Models—A Semiparametric Approach to Modelling Trends, Long-Range Dependence and Nonstationarity. *Computational Statistics and Data Analysis*, **40**, 393-419. [http://dx.doi.org/10.1016/S0167-9473\(02\)00007-5](http://dx.doi.org/10.1016/S0167-9473(02)00007-5)
- [15] Engle, R.F. and Smith, A.D. (1999) Stochastic Permanent Breaks. *The Review of Economics and Statistics*, **81**, 553-574. <http://dx.doi.org/10.1162/003465399558382>
- [16] Ray, B.K. and Tsay, R.S. (2002) Bayesian Methods for Change-Point Detection in Long-Range Dependent Processes. *Journal of Time Series Analysis*, **23**, 687-705. <http://dx.doi.org/10.1111/1467-9892.00286>
- [17] Lu, Y. and Perron, P. (2010) Modeling and Forecasting Stock Return Volatility Using a Random Level Shift Model. *Journal of Empirical Finance*, **17**, 138-156. <http://dx.doi.org/10.1016/j.jempfin.2009.10.001>
- [18] Perron, P. and Varsnekov, R.T. (2012) Combining Long Memory and Level Shifts in Modeling and Forecasting the Volatility of Asset Returns. Boston University, Boston.
- [19] Chapman, D.A. and Ogaki, M. (1998) Cotrending and the Stationarity of the Real Interest Rate. *Economics Letters*, **42**, 133-138. [http://dx.doi.org/10.1016/0165-1765\(93\)90050-M](http://dx.doi.org/10.1016/0165-1765(93)90050-M)
- [20] Bierens, H.J. (2000) Nonparametric Nonlinear Cotrending Analysis, with an Application to Interest and Inflation in the

- United States. *Journal of Business and Economic Statistics*, **18**, 323-337. <http://dx.doi.org/10.2307/1392265>
- [21] Hendry, D.F. (1996) A Theory of Co-Breaking. Nuffield College, University of Oxford, Oxford.
- [22] Hendry, D.F. and Massmann, M. (2007) Co-Breaking: Recent Advances and a Synopsis of the Literature. *Journal of Business and Economic Statistics*, **25**, 33-51. <http://dx.doi.org/10.1198/073500106000000422>
- [23] Engle, R.F. and Kozicki, S. (1993) Testing for Common Features. *Journal of Business & Economic Statistics*, **11**, 369-380. <http://dx.doi.org/10.1080/07350015.1993.10509966>
- [24] Engle, R.F. and Granger, C. (1987) Co-Integration and Error Correction: Representation, Estimation and Testing. *Econometrica*, **55**, 251-276. <http://dx.doi.org/10.2307/1913236>
- [25] Bollerslev, T. (1990) Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Approach. *Review of Economics and Statistics*, **72**, 489-505. <http://dx.doi.org/10.2307/2109358>
- [26] Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996) Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **74**, 3-30. [http://dx.doi.org/10.1016/S0304-4076\(95\)01749-6](http://dx.doi.org/10.1016/S0304-4076(95)01749-6)
- [27] Conrad, C. and Haag, B.R. (2006) Inequality Constraints in the Fractionally Integrated GARCH Model. *Journal of Financial Econometrics*, **4**, 413-449. <http://dx.doi.org/10.1093/jffinec/nbj015>
- [28] Chan, N.H. and Palma, W. (1998) State Space Modeling of Long Memory Processes. *Annals of Statistics*, **26**, 719-740. <http://projecteuclid.org/euclid.aos/1028144856>
<http://dx.doi.org/10.1214/aos/1028144856>
- [29] Kilian, L. (2011) Structural Vector Autoregressions. CEPR Discussion Paper, No. 8515.
- [30] Beran, J. and Feng, Y.H. (2001) Local Polynomial Estimation with a FARIMA-GARCH Error Process. *Bernoulli*, **7**, 733-750. <http://dx.doi.org/10.2307/3318539>
- [31] Beran, J. and Feng, Y.H. (2002) Data Driven Bandwidth Choice for SEMIFAR Models. *Journal of Computational and Graphical Statistics*, **11**, 690-713. <http://dx.doi.org/10.1198/106186002420>
- [32] Bordignon, S. and Raggi, D. (2010) Long Memory and Nonlinearities in Realized Volatility. University of Padova, Padova.
- [33] Martens, M., van Dijk, D. and de Pooter, M. (2003) Modeling and Forecasting S&P500 Volatility: Long Memory, Structural Breaks and Nonlinearity. Erasmus University of Rotterdam, Rotterdam.
- [34] Grassi, S. and de Magistris, P.S. (2011) When Long Memory Meets the Kalman Filter: A Comparative Study. Aarhus University, Aarhus.
- [35] Lavielle, M. and Moulines, E. (2000) Least-Squares Estimation of an Unknown Number of Shifts in a Time Series. *Journal of Time Series Analysis*, **21**, 33-59. <http://dx.doi.org/10.1111/1467-9892.00172>
- [36] Granger, C.W.J. and Hyung, N. (2004) Occasional Structural Breaks and Long Memory with an Application to the S&P500 Absolute Returns. *Journal of Empirical Finance*, **11**, 399-421. <http://dx.doi.org/10.1016/j.jempfin.2003.03.001>
- [37] Morana, C. (2014) New Insights on the US OIS Spreads Term Structure during the Recent Financial Turmoil. *Applied Financial Economics*, **24**, 291-317. <http://dx.doi.org/10.1080/09603107.2013.864034>
- [38] Bai, J.S. and Ng, S. (2013) Principal Components Estimation and Identification of Static Factors. *Journal of Econometrics*, **176**, 18-29. <http://dx.doi.org/10.1016/j.jeconom.2013.03.007>
- [39] Robin, J.M. and Smith, R.J. (2000) Tests of Rank. *Econometric Theory*, **16**, 151-175. <http://dx.doi.org/10.1017/S0266466600162012>
- [40] Peres-Neto, P.R., Jackson, D.A. and Somers, K.M. (2005) How Many Principal Components? Stopping Rules for Determining the Number of Non-Trivial Axes Revisited. *Computational Statistics and Data Analysis*, **49**, 974-997. <http://dx.doi.org/10.1016/j.csda.2004.06.015>
- [41] Bai, J. and Ng, S. (2002) Determining the Number of Factors in Approximate Factor Models. *Econometrica*, **70**, 191-221. <http://dx.doi.org/10.1111/1468-0262.00273>
- [42] Bai, J.S. and Ng, S. (2007) Determining the Number of Primitive Shocks in Factor Models. *Journal of Business and Economic Statistics*, **25**, 52-60. <http://dx.doi.org/10.1198/073500106000000413>
- [43] Nielsen, M.O. and Frederiksen, P.H. (2005) Finite Sample Comparison of Parametric, Semiparametric and Wavelet Estimators of Fractional Integration. *Econometric Reviews*, **24**, 405-443. <http://dx.doi.org/10.1080/07474930500405790>
- [44] Chan, N.H. and Palma, W. (2006) Estimation of Long-Memory Time Series Models: A Survey of Different Likelihood Based Approaches. In: Fomby, T.H. and Terrel, D., Eds., *Econometric Analysis of Economic and Financial Time Series, Advances in Econometrics*, Vol. 20, Emerald Group Publishing Limited, Bingley, 89-121.
- [45] Robinson, P.M. (2006) Conditional-Sum-of-Squares Estimation of Models for Stationary Time Series with Long Me-

- mory. *IMS Lecture Notes-Monograph Series, Time Series and Related Topics*, **52**, 130-137. <http://dx.doi.org/10.1214/074921706000000996>
- [46] Sowell, F. (1992) Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models. *Journal of Econometrics*, **53**, 165-188. [http://dx.doi.org/10.1016/0304-4076\(92\)90084-5](http://dx.doi.org/10.1016/0304-4076(92)90084-5)
- [47] Martin, L.V. and Wilkins, N.P. (1999) Indirect Estimation of ARFIMA and VARFIMA Models. *Journal of Econometrics*, **93**, 149-175. [http://dx.doi.org/10.1016/S0304-4076\(99\)00007-x](http://dx.doi.org/10.1016/S0304-4076(99)00007-x)
- [48] Baillie, R. and Kapetanios, G. (2013) Inference for Impulse Response Functions from Multivariate Strongly Persistent Processes. Queen Mary University of London, London.
- [49] Bai, J.S. and Ng, S. (2006) Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-Augmented Regressions. *Econometrica*, **74**, 1133-1150. <http://dx.doi.org/10.1111/j.1468-0262.2006.00696.x>
- [50] Bai, J.S. and Ng, S. (2008) Forecasting Economic Time Series Using Targeted Predictors. *Journal of Econometrics*, **146**, 304-317. <http://dx.doi.org/10.1016/j.jeconom.2008.08.010>
- [51] Granger, C.W.J. and Jeon, Y. (2004) Thick Modeling. *Economic Modelling*, **21**, 323-343.
- [52] Alexander, C.O. (2002) Principal Component Models for Generating Large GARCH Covariance Matrices. *Economic Notes*, **31**, 337-359. <http://dx.doi.org/10.1111/1468-0300.00089>
- [53] Amado, C. and Terasvirta, T. (2008) Modelling Conditional and Unconditional Heteroskedasticity with Smoothly Time-Varying Structure. CREATES Research Paper, No. 8.
- [54] Hamilton, J.D. and Susmel, R. (1994) Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics*, **64**, 307-333. [http://dx.doi.org/10.1016/0304-4076\(94\)90067-1](http://dx.doi.org/10.1016/0304-4076(94)90067-1)
- [55] Engle, R.F. (2002) Dynamic Conditional Correlation—A Simple Class of Multivariate GARCH Models. *Journal of Business and Economic Statistics*, **20**, 339-350. <http://dx.doi.org/10.1198/073500102288618487>
- [56] Engle, R.F. and Kelly, B.T. (2012) Dynamic Equicorrelation. *Journal of Business and Economics Statistics*, **30**, 212-228.
- [57] Quah, D. and Sargent, T.J. (1992) A Dynamic Index Model for Large Cross-Section. In: Stock, J. and Watson, M., Eds., *Business Cycle*, University of Chicago Press, Chicago.
- [58] Watson, M. and Engle, R.F. (1983) Alternative Algorithms for the Estimation of Dynamic Factor, Mimic and Varying Coefficient Regression Models. *Journal of Econometrics*, **23**, 385-400. <http://dx.doi.org/10.1016/0304-4076>
- [59] Doz, C., Giannone, D. and Reichlin, L. (2011) A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering. *Journal of Econometrics*, **164**, 188-205. <http://dx.doi.org/10.1016/j.jeconom.2011.02.012>
- [60] Doz, C., Giannone, D. and Reichlin, L. (2012) A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models. *Review of Economics and Statistics*, **94**, 1014-1024. http://dx.doi.org/10.1162/REST_a_00225
- [61] Bai, J.S. (2004) Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data. *Journal of Econometrics*, **122**, 137-138. <http://dx.doi.org/10.1016/j.jeconom.2003.10.022>
- [62] Castells, F., Laguna, P., Sornmo, L., Bollmann, A. and Millet-Roig, J. (2007) Principal Component Analysis in ECG Signal Processing. *EURASIP Journal on Advances in Signal Processing*, **1**, 98-119. <http://dx.doi.org/10.1155/2007/74580>
- [63] Morana, C. (2007) Multivariate Modelling of Long Memory Processes with Common Components. *Computational Statistics and Data Analysis*, **52**, 919-934. <http://dx.doi.org/10.1016/j.csda.2006.12.010>
- [64] Hatanka, M. and Yamada, H. (1994) Co-Trending: An Extended Version. University of Hiroshima, Hiroshima.
- [65] Lansang, J.R.G. and Barrios, E.B. (2009) Principal Components Analysis of Nonstationary Time Series Data. *Statistics and Computing*, **19**, 173-187. <http://dx.doi.org/10.1007/s11222-008-9082-y>
- [66] Morana, C. (2014) Factor Vector Autoregressive Estimation of Heteroskedastic Persistent and Non Persistent Processes Subject to Structural Breaks. DEMS Working Paper Series, No. 273.
- [67] Cassola, N. and Morana, C. (2012) Euro Money Market Spreads during the 2007-? Financial Crisis. *Journal of Empirical Finance*, **19**, 548-557. <http://dx.doi.org/10.1016/j.jempfin.2012.04.003>
- [68] Morana, C. (2013) Oil Price Dynamics, Macro-Finance Interactions and the Role of Financial Speculation. *Journal of Banking and Finance*, **37**, 206-226. <http://dx.doi.org/10.1016/j.jbankfin.2012.08.027>
- [69] Bagliano, F.C. and Morana, C. (2014) Determinants of US Financial Fragility Conditions. *Research in International Business and Finance*, **30**, 377-392. <http://dx.doi.org/10.1016/j.ribaf.2012.08.003>

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).

