DISCUSSION PAPER SERIES

No. 10417

A SPECTRAL EM ALGORITHM FOR DYNAMIC FACTOR MODELS

Gabriele Fiorentini, Alessandro Galesi
and Enrique Sentana

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

ISSN 0265-8003

A SPECTRAL EM ALGORITHM FOR DYNAMIC FACTOR MODELS

Gabriele Fiorentini, Alessandro Galesi and Enrique Sentana

Discussion Paper No. 10417
February 2015
Submitted 03 February 2015

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801
www.cepr.org

A SPECTRAL EM ALGORITHM FOR DYNAMIC FACTOR MODELS†

Abstract

We introduce a frequency domain version of the EM algorithm for general dynamic factor models. We consider both AR and ARMA processes, for which we develop iterative indirect inference procedures analogous to the algorithms in Hannan (1969). Although our proposed procedure allows researchers to estimate such models by maximum likelihood with many series even without good initial values, we recommend switching to a gradient method that uses the EM principle to swiftly compute frequency domain analytical scores near the optimum. We successfully employ our algorithm to construct an index that captures the common movements of US sectoral employment growth rates.

JEL Classification: C32, C38 and C51
Keywords: indirect inference, Kalman filter, sectoral employment, spectral maximum likelihood and Wiener-Kolmogorov filter

Gabriele Fiorentini fiorentini@disia.unifi.it
Università di Firenze and RCEA

Alessandro Galesi galesi@cemfi.edu.es
CEMFI

Enrique Sentana sentana@cemfi.es
CEMFI and CEPR

† We are grateful to Dante Amengual and Domenico Giannone, as well as to audiences at the Conference on Indirect Estimation Methods in Finance and Economics (Konstanz, 2014), the Econometric Methods for Banking and Finance Conference (Bank of Portugal, 2014), the NBER-NSF Time Series Conference (Federal Reserve Bank of St. Louis, 2014), the Advances in Econometrics Conference on Dynamic Factor Models (Aarhus, 2014) and the EC2 Advances in Forecasting Conference (UPF, 2014) for helpful comments, discussions and suggestions. Financial support from MIUR through the project Multivariate statistical models for risk assessment (Fiorentini), the European Research Council ERC Advanced Grant 293692 (Galesi) and the Spanish Ministry of Science and Innovation through grant ECO 2011-26342 (Sentana) is gratefully acknowledged.

1 Introduction

Dynamic factor models have been extensively used in macroeconomics and finance since their introduction by Sargent and Sims (1977) and Geweke (1977) as a way of capturing the cross-sectional and dynamic correlations between multiple series in a parsimonious way. A far from comprehensive list of early and more recent applications include not only business cycle analysis (see Litterman and Sargent (1979), Stock and Watson (1989, 1993), Diebold and Rudebusch (1996) or Gregory, Head and Raynauld (1997)) and bond yields (Singleton (1981), Jegadeesh and Pennacchi (1996), Dungey, Martin and Pagan (2000) or Diebold, Rudebusch and Aruoba (2006)), but also wages (Engle and Watson (1981)), employment (Quah and Sargent (1993)), commodity prices (Peña and Box (1987)) and financial contagion (Mody and Taylor (2007)).

In principle, Gaussian (P)MLEs of the parameters can be obtained from the usual time domain version of the log-likelihood function computed as a by-product of the Kalman filter prediction equations or from Whittle's (1962) frequency domain asymptotic approximation. Further, once the parameters have been estimated the Kalman smoother or its Wiener-Kolmogorov counterpart provide optimally filtered estimates of the latent factors. These estimation and filtering issues are well understood (see e.g. Harvey (1989)), and the same can be said of their numerical implementation (see Jungbacker and Koopman (2008)). In practice, though, researchers avoid ML except in relatively small models because of the heavy computational burden involved, which is disproportionately larger as the number of series considered increases.

To ameliorate this problem, Watson and Engle (1983) and Quah and Sargent (1993) applied the EM algorithm of Dempster, Laird and Rubin (1977) to the time domain versions of these models, thereby avoiding the computation of the likelihood function and its score. This iterative algorithm has been very popular in various areas of applied econometrics (see e.g. Hamilton (1990) in a different time series context). Its popularity can be attributed mainly to the efficiency of the procedure, as measured by its speed, and also to the generality of the approach, and its convergence properties (see Ruud (1991)). However, the time domain version of the EM algorithm has only been derived for dynamic factor models in which all the latent variables follow pure AR processes, and works best when the effects of the common factors on the observed variables are contemporaneous, which substantially limits the class of models to which they can be successfully applied. This limitation is particularly important in practice because recent macroeconomic applications of dynamic factor models have often considered moving average processes instead, sometimes treating the lagged latent variables as additional factors (see Bai and Ng (2008) and the references therein).

The purpose of this paper is to introduce a frequency domain version of the EM algorithm for general dynamic factor models with latent ARMA processes. Our algorithm reduces the computational burden so much that researchers can estimate such models by maximum likelihood with a large number of series even without good initial values. As is well known, though, this algorithm slows down considerably near the optimum. At that point, the best practical strategy would be to switch to a first derivative-based method. In that regard, we also explain in detail how to use the EM principle to swiftly compute frequency domain analytical scores. Finally, we illustrate our procedure with an empirical application to US employment data. Specifically, we follow Quah and Sargent (1993) and construct an index that captures the common movements of sectoral employment growth rates.

The rest of the paper is organised as follows. In section 2, we review the properties of dynamic factor models and their filters, as well as maximum likelihood estimation in the frequency domain. Then, we derive our estimation algorithm and present a numerical evaluation of its finite sample behaviour in section 3. This is followed by the empirical application in section 4. Finally, we discuss several interesting extensions for further research in section 5. Auxiliary results are gathered in appendices.

2 Theoretical background

2.1 Dynamic factor models

To keep the notation to a minimum, we focus on single factor models, which suffice to illustrate our procedures. A dynamic, exact, single factor model for a finite dimensional vector of N observed series, $\mathbf{y}_t$, can be defined in the time domain by the system of equations

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu} + \mathbf{c}(L)x_t + \mathbf{u}_t, \\ \alpha_x(L)x_t &= \beta_x(L)f_t, \\ \alpha_{u_i}(L)u_{i,t} &= \beta_{u_i}(L)v_{i,t}, \quad i = 1, \ldots, N, \\ (f_t, v_{1,t}, \ldots, v_{N,t})|I_{t-1}; \boldsymbol{\mu}, \boldsymbol{\theta} &\sim N[0, diag(1, \psi_1, \ldots, \psi_N)], \end{aligned}$$

where x_t is the only common factor, $\mathbf{u}_t$ the N specific factors, $\mathbf{c}(L) = \sum_{k=-m}^{n} \mathbf{c}_k L^k$ a vector of N possibly two-sided polynomials in the lag operator $c_i(L)$, $\alpha_x(L)$ and $\alpha_{u_i}(L)$ are one-sided polynomials of orders p_x and p_{u_i}, respectively, while $\beta_x(L)$ and $\beta_{u_i}(L)$ are one-sided polynomials of orders q_x and q_{u_i} coprime with $\alpha_x(L)$ and $\alpha_{u_i}(L)$, respectively, I_{t-1} is an information set that contains the values of $\mathbf{y}_t$ and f_t up to, and including time $t-1$, $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\theta}$ refers to all the remaining model parameters.

A specific example would be

$$
\begin{aligned}
\begin{pmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{pmatrix} &= \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix} + \begin{pmatrix} c_{1,0} \\ \vdots \\ c_{N,0} \end{pmatrix} x_t + \begin{pmatrix} c_{1,1} \\ \vdots \\ c_{N,1} \end{pmatrix} x_{t-1} + \begin{pmatrix} u_{1,t} \\ \vdots \\ u_{N,t} \end{pmatrix}, \qquad (1) \\
x_t &= \alpha_{x1} x_{t-1} + f_t - \beta_{x1} f_{t-1}, \\
u_{it} &= \alpha_{u_i 1} u_{it-1} + v_{it} - \beta_{u_i 1} v_{it-1}, \quad i = 1, \ldots, N.
\end{aligned}
$$

Note that the dynamic nature of the model is the result of three different characteristics:

1. The serial correlation of the common factor x_t
2. The serial correlation of the idiosyncratic factors $\mathbf{u}_t$
3. The heterogeneous dynamic impact of the common factor on each of the observed variables through the series-specific distributed lag polynomials $c_i(L)$.

Thus, we would need to shut down all three sources to go back to a traditional static factor model (see Lawley and Maxwell (1971)). Cancelling only one or two of those channels still results in a dynamic factor model. For example, Engle and Watson (1981) considered models with static factor loadings, while Peña and Box (1987) further assumed that the specific factors were white noise. To some extent, characteristics 1 and 3 overlap, as one could always write any dynamic factor model in terms of white noise common factors. In this regard, the assumption of ARMA(p_x, q_x) dynamics for the common factor can be regarded as a parsimonious way of modelling an infinite distributed lag.

2.2 Spectral density matrix

Under the assumption that $\mathbf{y}_t$ is a covariance stationary process, possibly after suitable transformations as in section 4, the spectral density matrix of the observed variables will be proportional to

$$
\begin{gathered}
\mathbf{G}_{\mathbf{yy}}(\lambda) = \mathbf{c}(e^{-i\lambda}) G_{xx}(\lambda) \mathbf{c}'(e^{i\lambda}) + \mathbf{G}_{\mathbf{uu}}(\lambda), \\
G_{xx}(\lambda) = \frac{\beta_x(e^{-i\lambda})\beta_x(e^{i\lambda})}{\alpha_x(e^{-i\lambda})\alpha_x(e^{i\lambda})}, \\
\mathbf{G}_{\mathbf{uu}}(\lambda) = diag[G_{u_1u_1}(\lambda), \ldots, G_{u_Nu_N}(\lambda)], \\
G_{u_iu_i}(\lambda) = \psi_i \frac{\beta_{u_i}(e^{-i\lambda})\beta_{u_i}(e^{i\lambda})}{\alpha_{u_i}(e^{-i\lambda})\alpha_{u_i}(e^{i\lambda})}.
\end{gathered}
$$

Thus, $\mathbf{G}_{\mathbf{yy}}(\lambda)$ is the sum of the rank 1 matrix $\mathbf{c}(e^{-i\lambda}) G_{xx}(\lambda) \mathbf{c}'(e^{i\lambda})$ and the diagonal matrix $\mathbf{G}_{\mathbf{uu}}(\lambda)$, thereby inheriting the exact single factor structure of the unconditional covariance matrix of a static factor model. We can easily ensure the separate identification of those two matrices when $\mathbf{G_{uu}}(\lambda)$ has full rank provided N is sufficiently large. The separate identification of $\mathbf{c}(e^{-i\lambda})$ and $G_{xx}(\lambda)$ is trickier, but it can be guaranteed (up to scale and time shifts) as long as the N polynomials $c_i(.)$ do not share a common root (see Geweke (1977), Geweke and Singleton (1981) and more recently Heaton and Solo (2004) for a more thorough discussion of identification in dynamic factor models). To avoid dealing with nonsensical situations, henceforth we maintain the assumption that the model that has to be estimated is identified. This will indeed be the case in our empirical application in section 4.

For the model presented in (1),

$$G_{xx}(\lambda) = \frac{\beta_x(e^{-i\lambda})\beta_x(e^{i\lambda})}{\alpha_x(e^{-i\lambda})\alpha_x(e^{i\lambda})} = \frac{1+\beta_{x1}^2 - 2\beta_{x1}\cos\lambda}{1+\alpha_{x1}^2 - 2\alpha_{x1}\cos\lambda},$$

where we have exploited the fact that the variance of f_t has been normalised to 1 for identification purposes.[1]

Similarly,

$$G_{u_iu_i}(\lambda) = \frac{\beta_{u_i}(e^{-i\lambda})\beta_{u_i}(e^{i\lambda})\psi_i}{\alpha_{u_i}(e^{-i\lambda})\alpha_{u_i}(e^{i\lambda})} = \frac{1+\beta_{u_i1}^2 - 2\beta_{u_i1}\cos\lambda}{1+\alpha_{u_i1}^2 - 2\alpha_{u_i1}\cos\lambda}\psi_i.$$

Finally,

$$\mathbf{c}(e^{-i\lambda t}) = \mathbf{c}_0 + \mathbf{c}_1 e^{-i\lambda} = \begin{pmatrix} c_{1,0} + c_{1,1}e^{-i\lambda} \\ \vdots \\ c_{N,0} + c_{N,1}e^{-i\lambda} \end{pmatrix} = \begin{pmatrix} c_1(e^{-i\lambda t}) \\ \vdots \\ c_N(e^{-i\lambda t}) \end{pmatrix}. \tag{2}$$

The fact that the heterogeneous impact of the common factor on each of the observed variables is in principle dynamic implies that the spectral density matrix of $\mathbf{y}_t$ will generally be complex but Hermitian, even though the spectral densities of x_t and u_{it} are all real because they correspond to univariate processes.

2.3 Wiener-Kolmogorov filter

By working in the frequency domain we can easily obtain smoothed estimators of the latent variables. Specifically, let

$$\begin{aligned} \mathbf{y}_t - \boldsymbol{\mu} &= \int_{-\pi}^{\pi} e^{i\lambda t} d\mathbf{Z}^{\mathbf{y}}(\lambda), \\ V[d\mathbf{Z}^{\mathbf{y}}(\lambda)] &= \mathbf{G_{yy}}(\lambda)d\lambda \end{aligned}$$

denote the spectral decomposition of the observed vector process.

Assuming that $\mathbf{G_{yy}}(\lambda_j)$ is not singular at any frequency, the Wiener-Kolmogorov two-sided filter for the common factor x_t at each frequency is given by

$$dZ^{x^K}(\lambda) = G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)d\mathbf{Z}^{\mathbf{y}}(\lambda), \tag{3}$$

[1] Other symmetric scaling assumptions would normalise the unconditional variance of x_t, or some norm of the vector of loadings $\mathbf{c}_0$. Alternatively, we could asymmetrically fix one element of $\mathbf{c}_0$ to 1.

where

$$G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)$$

is known as the transfer function of the common factor smoother. As a result, the spectral density of the smoothed values of the common factors, $x_{t|\infty}^K$, is

$$G_{x^K x^K}(\lambda) = G_{xx}^2(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})$$

thanks to the Hermitian nature of $\mathbf{G}_{\mathbf{yy}}(\lambda)$, while the spectral density of the final estimation error $x_t - x_{t|\infty}^K$ will be given by

$$G_{xx}(\lambda) - G_{xx}^2(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}) = \omega(\lambda).$$

Similarly, the Wiener-Kolmogorov smoother for the N specific factors will be

$$d\mathbf{Z}^{\mathbf{u}^K}(\lambda) = \mathbf{G}_{\mathbf{uu}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)d\mathbf{Z}^{\mathbf{y}}(\lambda)$$
$$= \left[\mathbf{I}_N - \mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\right] d\mathbf{Z}^{\mathbf{y}}(\lambda) = d\mathbf{Z}^{\mathbf{y}}(\lambda) - \mathbf{c}(e^{-i\lambda})dZ^{x^K}(\lambda).$$

Hence, the spectral density matrix of the smoothed values of the specific factors will be given by

$$\mathbf{G}_{\mathbf{u}^K\mathbf{u}^K}(\lambda) = \mathbf{G}_{\mathbf{uu}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{G}_{\mathbf{uu}}(\lambda),$$

while the spectral density of their final estimation errors $\mathbf{u}_t - \mathbf{u}_{t|\infty}^K$ is

$$\mathbf{G}_{\mathbf{uu}}(\lambda) - \mathbf{G}_{\mathbf{u}^K\mathbf{u}^K}(\lambda) = \mathbf{G}_{\mathbf{uu}}(\lambda) - \mathbf{G}_{\mathbf{uu}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{G}_{\mathbf{uu}}(\lambda) = \omega(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda}) = \mathbf{\Xi}(\lambda).$$

Finally, the co-spectrum between $x_{t|\infty}^K$ and $\mathbf{u}_{t|\infty}^K$ will be

$$\mathbf{G}_{x^K\mathbf{u}^K}(\lambda) = G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{G}_{\mathbf{uu}}(\lambda).$$

Computations can be considerably speeded up by exploiting the Woodbury formula under the assumption that neither $G_{xx}(\lambda)$ nor $\mathbf{G}_{\mathbf{uu}}(\lambda)$ are singular at any frequency (see Sentana (2000) for a generalisation):

$$\begin{aligned}
|\mathbf{G}_{\mathbf{yy}}(\lambda)| &= |\mathbf{G}_{\mathbf{uu}}(\lambda)|G_{xx}(\lambda)\omega(\lambda), \\
\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda) &= \mathbf{G}_{\mathbf{uu}}^{-1}(\lambda) - \omega(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda), \\
\omega(\lambda) &= [G_{xx}^{-1}(\lambda) + \mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}.
\end{aligned}$$

The advantage of these expressions is that $\mathbf{G}_{\mathbf{uu}}(\lambda)$ is a diagonal matrix and $\omega(\lambda)$ a scalar, which greatly simplifies the computations.

On this basis, the transfer function of the Wiener-Kolmogorov common factor smoother becomes

$$G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda) = \omega(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda),$$

so

$$G_{x^K x^K}(\lambda) = \omega(\lambda)G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}) = G_{xx}(\lambda) - \omega(\lambda),$$

where we have used the fact that

$$\omega(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}) = 1 - \omega(\lambda)G_{xx}^{-1}(\lambda), \tag{4}$$

which can be easily proved by dividing both sides by $\omega(\lambda)$.

Similarly, the transfer function of the Wiener-Kolmogorov specific factors smoother will be

$$\mathbf{G}_{\mathbf{uu}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda) = \mathbf{I}_N - \omega(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda),$$

so

$$\mathbf{G}_{\mathbf{u}^K\mathbf{u}^K}(\lambda) = \mathbf{G}_{\mathbf{uu}}(\lambda) - \omega(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda}).$$

Finally,

$$\mathbf{G}_{x^K\mathbf{u}^K}(\lambda) = \omega(\lambda)\mathbf{c}'(e^{i\lambda}).$$

2.4 The minimal sufficient statistics for $\{x_t\}$

Define $x_{t|\infty}^G$ as the spectral GLS estimator of x_t through the transformation

$$dZ^{x^G}(\lambda) = [\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}\mathbf{c}'^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)d\mathbf{Z}^{\mathbf{y}}(\lambda).$$

Similarly, define $\mathbf{u}_{t|\infty}^G$ through

$$d\mathbf{Z}^{\mathbf{u}^G}(\lambda) = \{\mathbf{I}_N - [\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\}d\mathbf{Z}^{\mathbf{y}}(\lambda).$$

It is then easy to see that the joint spectral density of $x_{t|\infty}^G$ and $\mathbf{u}_{t|\infty}^G$ will be block-diagonal, with the (1,1) element being

$$G_{xx}(\lambda) + [\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}$$

and the (2,2) block

$$\mathbf{G}_{\mathbf{yy}}(\lambda) - \mathbf{c}(e^{-i\lambda})[\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}\mathbf{c}'(e^{i\lambda}),$$

whose rank is $N-1$.

This block-diagonality allows us to factorise the spectral log-likelihood function of $\mathbf{y}_t$ as the sum of the log-likelihood function of $x_{t|\infty}^G$, which is univariate, and the log-likelihood function of

$\mathbf{u}^G_{t|\infty}$. Importantly, the parameters characterising $G_{xx}(\lambda)$ only enter through the first component. In contrast, the remaining parameters affect both components. Moreover, we can easily show that

1. $x^G_{t|T} = x_t + \zeta^G_{t|T}$, with x_t and $\zeta^G_{t|T}$ orthogonal at all leads and lags.

2. The smoothed estimator of x_t obtained by applying the Wiener- Kolmogorov filter to $x^G_{t|\infty}$ coincides with $x^K_{t|\infty}$.

This confirms that $x^G_{t|\infty}$ constitute minimal sufficient statistics for x_t, thereby generalising earlier results by Jungbacker and Koopman (2008), who considered models in which $\mathbf{c}(e^{-i\lambda}) = \mathbf{c}$ for all λ, and Fiorentini, Sentana and Shephard (2004), who looked at the related class of factor models with time-varying volatility (see also Gourieroux, Monfort and Renault (1991)). In addition, the degree of unobservability of x_t depends exclusively on the size of $[\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}$ relative to $G_{xx}(\lambda)$ (see Sentana (2004) for a closely related discussion).

2.5 Maximum likelihood estimation in the frequency domain

Let

$$\mathbf{I}_{\mathbf{yy}}(\lambda) = \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_s - \boldsymbol{\mu})' e^{-i(t-s)\lambda} \tag{5}$$

denote the periodogram matrix and $\lambda_j = 2\pi j/T$ $(j = 0, \ldots T-1)$ the usual Fourier frequencies. If we assume that $\mathbf{G}_{\mathbf{yy}}(\lambda)$ is not zero at any of those frequencies, the so-called Whittle (discrete) spectral approximation to the log-likelihood function is[2]

$$-\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1}\ln|\mathbf{G}_{\mathbf{yy}}(\lambda_j)| - \frac{1}{2}\sum_{j=0}^{T-1}\mathrm{tr}\left\{\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)[2\pi\mathbf{I}_{\mathbf{yy}}(\lambda_j)]\right\} \tag{6}$$

(see e.g. Hannan (1973) and Dunsmuir and Hannan (1976)).

Expression (5), though, is far from ideal from a computational point of view, and for that reason we make use of the Fast Fourier Transform (FFT). Specifically, given the $T \times N$ original real data matrix $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_t, \ldots, \mathbf{y}_T)'$, the FFT creates the centred and orthogonalised $T \times N$ complex data matrix $\mathbf{Z}^{\mathbf{y}} = (\mathbf{z}^{\mathbf{y}}_0, \ldots, \mathbf{z}^{\mathbf{y}}_j, \ldots, \mathbf{z}^{\mathbf{y}}_{T-1})'$ by effectively premultiplying $\mathbf{Y} - \ell_T\boldsymbol{\mu}'$ by the $T \times T$ Fourier matrix $\mathbf{W}$. On this basis, we can easily compute $\mathbf{I}_{\mathbf{yy}}(\lambda_j)$ as $2\pi\mathbf{z}^{\mathbf{y}}_j\mathbf{z}^{\mathbf{y}*}_j$, where $\mathbf{z}^{\mathbf{y}*}_j$ is the complex conjugate transpose of $\mathbf{z}^{\mathbf{y}}_j$. Hence, the spectral approximation to the log-likelihood function for a non-singular $\mathbf{G}_{\mathbf{yy}}(\lambda)$ becomes

$$-\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1}\ln|\mathbf{G}_{\mathbf{yy}}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1}\mathbf{z}^{\mathbf{y}*}_j\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{z}^{\mathbf{y}}_j,$$

[2] There is also a continuous version which replaces sums by integrals (see Dusmuir and Hannan (1976)).

which can be regarded as the log-likelihood function of T independent but heteroskedastic complex Gaussian observations.

But since $\mathbf{z}_j^{\mathbf{y}}$ does not depend on $\boldsymbol{\mu}$ for $j = 1, \ldots, T-1$ because ℓ_T is proportional to the first column of the orthogonal Fourier matrix and $\mathbf{z}_0^{\mathbf{y}} = (\bar{\mathbf{y}}_T - \boldsymbol{\mu})$, where $\bar{\mathbf{y}}_T$ is the sample mean of $\mathbf{y}_t$, it immediately follows that the ML of $\boldsymbol{\mu}$ will be $\bar{\mathbf{y}}_T$. As for the remaining parameters, the score function will be given by:

$$\mathbf{d}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=0}^{T-1} \mathbf{d}(\lambda_j; \boldsymbol{\theta}),$$

$$\begin{aligned}
\mathbf{d}(\lambda_j; \boldsymbol{\theta}) &= \frac{1}{2}\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda_j)\right]}{\partial \boldsymbol{\theta}}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j) \otimes \mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\right] vec\left[2\pi \mathbf{z}_j^{\mathbf{y}c}\mathbf{z}_j^{\mathbf{y}\prime} - \mathbf{G}_{\mathbf{yy}}'(\lambda_j)\right] \\
&= \frac{1}{2}\frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda_j)]}{\partial \boldsymbol{\theta}} \mathbf{M}(\lambda_j)\mathbf{m}(\lambda_j), \qquad (7)
\end{aligned}$$

where $\mathbf{z}_j^{\mathbf{y}c} = \mathbf{z}_j^{\mathbf{y}*\prime}$ is the complex conjugate of $\mathbf{z}_j^{\mathbf{y}}$,

$$\mathbf{m}(\lambda_j) = vec\left[2\pi \mathbf{z}_j^{\mathbf{y}c}\mathbf{z}_j^{\mathbf{y}\prime} - \mathbf{G}_{\mathbf{yy}}'(\lambda_j)\right] \qquad (8)$$

and

$$\mathbf{M}(\lambda_j) = \mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j) \otimes \mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j). \qquad (9)$$

The information matrix is block diagonal between $\boldsymbol{\mu}$ and the elements of $\boldsymbol{\theta}$, with the (1,1)-element being $\mathbf{G}_{\mathbf{yy}}(0)$ and the (2,2)-block being

$$\mathbf{Q} = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{\partial \boldsymbol{\theta}} \mathbf{M}(\lambda)\left\{\frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{\partial \boldsymbol{\theta}}\right\}^* d\lambda, \qquad (10)$$

a consistent estimator of which will be provided by either by the outer product of the score or by

$$\boldsymbol{\Phi}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=0}^{T-1} \frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda_j)]}{\partial \boldsymbol{\theta}} \mathbf{M}(\lambda_j)\left\{\frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda_j)]}{\partial \boldsymbol{\theta}}\right\}^*.$$

Formal results showing the strong consistency and asymptotic normality of the resulting ML estimators under suitable regularity conditions have been provided by Dunsmuir and Hannan (1976) and Dunsmuir (1979), who also show their asymptotic equivalence to the time domain ML estimators.[3]

Appendix A provides detailed expressions for the Jacobian of $vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]$ and the spectral score of dynamic factor models, while appendix B includes numerically reliable and efficient formulae for the information matrix. Those expressions make extensive use of the complex

[3] This equivalence is not surprising in view of the contiguity of the Whittle measure in the Gaussian case (see Choudhuri, Ghosal and Roy (2004)).

version of the Woodbury formula described in section 2.3. We can also exploit the same formula to compute the quadratic form $\mathbf{z}_j^{\mathbf{y}*}\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{z}_j^{\mathbf{y}}$ as

$$\begin{aligned}\mathbf{z}_j^{\mathbf{y}*}\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda_j)\mathbf{z}_j^{\mathbf{y}} - \mathbf{z}_j^{\mathbf{y}*}\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\omega(\lambda_j)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{z}_j^{\mathbf{y}} \\ = \mathbf{z}_j^{\mathbf{y}*}\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda_j)\mathbf{z}_j^{\mathbf{y}} - \omega(\lambda_j)z_j^{x^K*}(\boldsymbol{\theta})z_j^{x^K}(\boldsymbol{\theta}),\end{aligned}$$

where

$$z_j^{x^K}(\boldsymbol{\theta}) = E[z_j^x|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] = G_{xx}(\lambda_j)\mathbf{c}'(e^{i\lambda_j})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{z}_j^{\mathbf{y}} = \omega(\lambda)\mathbf{c}'(e^{i\lambda_j})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda_j)\mathbf{z}_j^{\mathbf{y}} \tag{11}$$

denotes the filtered value of z_j^x given the observed series and the current parameter values from (3).

Nevertheless, when N is large the number of parameters is huge, and the direct maximisation of the log-likelihood function becomes excruciatingly slow, especially without good initial values. For that reason, in the next section we describe a much faster alternative to obtain the maximum likelihood estimators of all the model parameters.

3 Spectral EM algorithm

As we mentioned in the introduction, the EM algorithm of Dempster, Laird and Rubin (1977) adapted to static factor models by Rubin and Thayer (1982) was successfully employed to handle a very large dataset of stock returns by Lehmann and Modest (1988). Watson and Engle (1983) and Quah and Sargent (1993) also applied the algorithm in the time domain to dynamic factor models and some generalisations, but they restricted common and specific factors to follow low order AR processes, which seems rather restrictive given the prevalence of the ARMA(1,1) model in univariate time series analysis.

We saw before that the spectral density matrix of a dynamic single factor model has the structure of the unconditional covariance matrix of a static factor model, but with different common and idiosyncratic variances for each frequency. Demos and Sentana (1998) applied a time domain version of the EM algorithm to conditionally heteroskedastic factor models in which the common factors followed GARCH-type processes. We could easily adapt their algorithm to models with white noise idiosyncratic factors and contemporaneous effects of the common factors on the observed variables if we replaced the subscript t for time with the subscript j for frequency. However, since we want to consider more complex models, we need to do some additional algebra.

3.1 Complete log-likelihood function

Consider a situation in which the common factor x_t was also observed. The joint spectral density of $\mathbf{y}_t$ and x_t, which is given by

$$\begin{bmatrix} \mathbf{G}_{\mathbf{yy}}(\lambda) & \mathbf{G}_{\mathbf{y}x}(\lambda) \\ \mathbf{G}^*_{\mathbf{y}x}(\lambda) & G_{xx}(\lambda) \end{bmatrix} = \begin{bmatrix} \mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda}) + \mathbf{G}_{\mathbf{uu}}(\lambda) & \mathbf{c}(e^{-i\lambda})G_{xx}(\lambda) \\ G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda}) & G_{xx}(\lambda) \end{bmatrix},$$

could be diagonalised as

$$\begin{bmatrix} \mathbf{I}_N & \mathbf{c}(e^{-i\lambda}) \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{G}_{\mathbf{uu}}(\lambda) & \mathbf{0} \\ \mathbf{0} & G_{xx}(\lambda) \end{bmatrix} \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{c}'(e^{i\lambda}) & 1 \end{bmatrix},$$

with

$$\left| \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{c}'(e^{i\lambda}) & 1 \end{bmatrix} \right| = 1$$

and

$$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{c}'(e^{i\lambda}) & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ -\mathbf{c}'(e^{i\lambda}) & 1 \end{bmatrix}.$$

Let us define as $[\mathbf{Z}^{\mathbf{y}}|\mathbf{z}^x]$ as the Fourier transform of the $T \times (N+1)$ matrix $[\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{x}] = [\mathbf{Y}|\mathbf{x}]$ so that the joint periodogram of $\mathbf{y}_t$ and x_t at frequency λ_j could be quickly computed as

$$2\pi \begin{pmatrix} \mathbf{z}^{\mathbf{y}}_j \\ z^x_j \end{pmatrix} \begin{pmatrix} \mathbf{z}^{\mathbf{y}*}_j & z^{x*}_j \end{pmatrix},$$

where we have implicitly assumed that either the elements of $\mathbf{y}$ have zero mean, or else that they have been previously demeaned by subtracting their sample averages.

In this notation, the spectral approximation to the joint log-likelihood function would become

$$l(\mathbf{y}, x) = -\frac{(N+1)T}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1} \ln \left| \begin{bmatrix} \mathbf{G}_{\mathbf{yy}}(\lambda) & \mathbf{G}_{\mathbf{y}x}(\lambda_j) \\ \mathbf{G}^*_{\mathbf{y}x}(\lambda_j) & G_{xx}(\lambda_j) \end{bmatrix} \right|$$

$$-\frac{2\pi}{2}\sum_{j=0}^{T-1} \begin{pmatrix} \mathbf{z}^{\mathbf{y}*}_j & z^{x*}_j \end{pmatrix} \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ -\mathbf{c}'(e^{i\lambda_j}) & 1 \end{bmatrix} \begin{bmatrix} \mathbf{G}^{-1}_{\mathbf{uu}}(\lambda_j) & \mathbf{0} \\ \mathbf{0} & G^{-1}_{xx}(\lambda_j) \end{bmatrix} \begin{bmatrix} \mathbf{I}_N & \mathbf{c}(e^{-i\lambda_j}) \\ \mathbf{0} & 1 \end{bmatrix} \begin{pmatrix} \mathbf{z}^{\mathbf{y}}_j \\ z^x_j \end{pmatrix}$$

$$= -\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1} \ln|\mathbf{G}_{\mathbf{uu}}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1} \mathbf{z}^{\mathbf{u}*}_j \mathbf{G}^{-1}_{\mathbf{uu}}(\lambda_j)\mathbf{z}^{\mathbf{u}}_j$$

$$-\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1} \ln|G_{xx}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{xx}(\lambda_j) z^x_j z^{x*}_j$$

$$= \sum_{i=1}^{N}\left[-\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1} \ln|G_{u_iu_i}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{u_iu_i}(\lambda_j) z^{u_i}_j z^{u_i*}_j \right] \tag{12}$$

$$-\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1} \ln|G_{xx}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{xx}(\lambda_j) z^x_j z^{x*}_j \tag{13}$$

$$= \sum_{i=1}^{N} l(\mathbf{y}_i|\mathbf{x}) + l(\mathbf{x}) = l(\mathbf{Y}|\mathbf{x}) + l(\mathbf{x}),$$

where[4]

$$z_j^{u_i} = z_j^{y_i} - c_i(e^{-i\lambda_j})z_j^x = z_j^{y_i} - \sum_{k=-m}^{n} c_{i,k}e^{-ik\lambda}z_j^x, \tag{14}$$

so that

$$\begin{aligned} z_j^{u_i}z_j^{u_i*} &= z_j^{y_i}z_j^{y_i*} - c_i(e^{-i\lambda_j})z_j^x z_j^{y_i*} - c_i(e^{i\lambda_j})z_j^{y_i}z_j^{x*} + c_i(e^{-i\lambda_j})c_i(e^{i\lambda_j})z_j^x z_j^{x*} \\ &= I_{y_iy_i}(\lambda_j) - c_i(e^{-i\lambda_j})I_{xy_i}(\lambda_j) - c_i(e^{i\lambda_j})I_{y_ix}(\lambda_j) + c_i(e^{-i\lambda_j})c_i(e^{i\lambda_j})I_{xx}(\lambda_j) = I_{u_iu_i}(\lambda_j). \end{aligned}$$

In this way, we have decomposed the joint log-likelihood function of $\mathbf{y}_1, \ldots, \mathbf{y}_N$ and $\mathbf{x}$ as the sum of the marginal log-likelihood of $\mathbf{x}$ in (13) and the log-likelihood function of $\mathbf{y}_1, \ldots, \mathbf{y}_N$ given $\mathbf{x}$, $l(\mathbf{Y}|\mathbf{x})$, which in turn can be decomposed as the sum of N univariate components in (12) by exploiting the diagonality of $\mathbf{G_{uu}}(\lambda_j)$.

Importantly, these expressions can be computed using real arithmetic only since

$$c_i(e^{-i\lambda_j})I_{xy_i}(\lambda_j) + c_i(e^{i\lambda_j})I_{y_ix}(\lambda_j) = 2\text{real}\left[c_i(e^{-i\lambda_j})I_{xy_i}(\lambda_j)\right]$$

and

$$c_i(e^{-i\lambda_j})c_i(e^{i\lambda_j})I_{xx}(\lambda_j) = \left\|c_i(e^{-i\lambda_j})\right\|^2 I_{xx}(\lambda_j).$$

Let us classify the parameters into three blocks:

1. the parameters that characterise the spectral density of $x_t : \boldsymbol{\theta}_x$
2. the parameters that characterise the spectral density of $\mathbf{u}_t : \boldsymbol{\psi} = (\psi_1, \ldots, \psi_N)'$ and $\boldsymbol{\theta}_{\mathbf{u}} = (\boldsymbol{\theta}'_{u_i}, , \ldots, \boldsymbol{\theta}'_{u_N})'$
3. the parameters that characterise the dynamic idiosyncratic impact of the common factors on each observed variable:

   $\mathbf{c} = (\mathbf{c}'_{1\cdot}, \ldots, \mathbf{c}'_{i\cdot}, \ldots, \mathbf{c}'_{N\cdot})'$, where $\mathbf{c}'_{i\cdot} = (c_{i,-m}, \ldots, c_{i,0}, \ldots, c_{i,n})$.

Importantly, $\boldsymbol{\theta}_x$ only appear in (13) while $\boldsymbol{\theta}_{\mathbf{u}}$ and $\mathbf{c}$ appear in (12). This sequential cut on the joint spectral density confirms that z^x and therefore x_t would be weakly exogenous for $\boldsymbol{\psi}$, $\boldsymbol{\theta}_{\mathbf{u}}$ and $\mathbf{c}$ (see Engle, Hendry and Richard (1983)). Moreover, the fact that f_t is uncorrelated at all leads and lags with $\mathbf{v}_t$ implies that x_t would be strongly exogenous too.

[4] Note that we could have expressed those log-likelihood in terms of $I_{xx}(\lambda_j) = z_j^x z_j^{x*}$, $\mathbf{I_{uu}}(\lambda) = \mathbf{z}_j^{\mathbf{u}}\mathbf{z}_j^{\mathbf{u}*}$and $\mathbf{I}_{\mathbf{u}x}(\lambda) = \mathbf{z}_j^{\mathbf{u}}z_j^{x*}$, but for the EM algorithm it is more convenient to work with the underlying complex random variables.

We can also exploit the aforementioned log-likelihood decomposition to obtain the score of the complete log-likelihood function. In this way, we can write

$$\frac{\partial l(\mathbf{Y},\mathbf{x})}{\partial \boldsymbol{\theta}_x} = \frac{\partial l(\mathbf{x})}{\partial \boldsymbol{\theta}_x} = \frac{1}{2}\sum_{j=0}^{T-1} \frac{\partial G_{xx}(\lambda_j)}{\partial \boldsymbol{\theta}_x} G_{xx}^{-2}(\lambda_j) \left[2\pi z_j^x z_j^{x*} - G_{xx}(\lambda_j)\right] \tag{15a}$$

$$\frac{\partial l(\mathbf{Y},\mathbf{x})}{\partial \boldsymbol{\theta}_{u_i}} = \frac{\partial l(\mathbf{y}_i|\mathbf{x})}{\partial \boldsymbol{\theta}_{u_i}} = \frac{1}{2}\sum_{j=0}^{T-1} \frac{\partial G_{u_iu_i}(\lambda_j)}{\partial \boldsymbol{\theta}_{u_i}} G_{u_iu_i}^{-2}(\lambda_j) \left[2\pi z_j^{u_i} z_j^{u_i*} - G_{u_iu_i}(\lambda_j)\right] \tag{15b}$$

$$\frac{\partial l(\mathbf{Y},\mathbf{x})}{\partial c_{i,k}} = \frac{\partial l(\mathbf{y}_i|\mathbf{x})}{\partial c_{i,k}} = \frac{2\pi}{2}\sum_{j=0}^{T-1} G_{u_iu_i}^{-1}(\lambda_j) \left[z_j^{u_i} e^{ik\lambda_j} z_j^{x*} + e^{-ik\lambda_j} z_j^x z_j^{u_i*}\right]$$

$$= \frac{2\pi}{2}\sum_{j=0}^{T-1} G_{u_iu_i}^{-1}(\lambda_j) \left[\left(z_j^{y_i} - \sum_{l=-m}^{n} c_{i,l} e^{-il\lambda} z_j^x\right) e^{ik\lambda_j} z_j^{x*} + e^{-ik\lambda_j} z_j^x \left(z_j^{y_i*} - \sum_{l=-m}^{n} c_{i,l} e^{il\lambda} z_j^{x*}\right)\right], \tag{15c}$$

where we have used the fact that

$$\frac{\partial z_j^{u_i}}{\partial c_{i,k}} = -e^{-ik\lambda} z_j^x$$

in view of (14).

Expression (15a) confirms that the MLE of $\boldsymbol{\theta}_x$ would be obtained from a univariate time series model for x_t. However, since $G_{xx}(\lambda_j)$ also depends on $\boldsymbol{\theta}_x$, there are no closed form solutions for models with MA components, and we would have to resort to the numerical optimisation of (13). We revisit this issue in section 3.2.

In an AR(1) example, in contrast, the derivative of $G_{xx}(\lambda)$ with respect to α_{x1} would be

$$\frac{\partial G_{xx}(\lambda)}{\partial \alpha_{x1}} = \frac{2(\cos\lambda - \alpha_{x1})}{(1+\alpha_{x1}^2 - 2\alpha_{x1}\cos\lambda)^2}.$$

Hence, the log-likelihood score would become

$$\begin{aligned}\frac{\partial l(\mathbf{x})}{\partial \alpha_{x1}} &= \frac{1}{2}\sum_{j=0}^{T-1} \frac{2(\cos\lambda_j - \alpha_{x1})}{(1+\alpha_{x1}^2 - 2\alpha_{x1}\cos\lambda_j)^2}(1+\alpha_{x1}^2 - 2\alpha_{x1}\cos\lambda_j)^2 \times \\ &\times \left[2\pi z_j^x z_j^{x*} - \frac{1}{(1+\alpha_{x1}^2 - 2\alpha_{x1}\cos\lambda_j)}\right] = 2\pi\sum_{j=0}^{T-1}(\cos\lambda_j - \alpha_{x1}) z_j^x z_j^{x*},\end{aligned}$$

where we have exploited the fact that

$$\sum_{j=0}^{T-1} \frac{(\cos\lambda_j - \alpha_{x1})}{(1+\alpha_{x1}^2 - 2\alpha_{x1}\cos\lambda_j)} = \gamma_{xx}(1) - \alpha_{x1}\gamma_{xx}(0) = 0.$$

As a result, when we set the score to 0 and solve for α_{x1} we obtain

$$\hat{\alpha}_{x1} = \frac{\sum_{j=0}^{T-1}\cos\lambda_j z_j^x z_j^{x*}}{\sum_{j=0}^{T-1} z_j^x z_j^{x*}} = \frac{\sum_{j=0}^{T-1}\cos\lambda_j I_{xx}(\lambda_j)}{\sum_{j=0}^{T-1} I_{xx}(\lambda_j)}.$$

But since

$$I_{xx}(\lambda_j) = \hat{\gamma}_{xx}(0) + 2\sum_{k=1}^{T-1}\hat{\gamma}_{xx}(k)\cos(k\lambda_j),$$

we would have that

$$\sum_{j=0}^{T-1} 2\pi I_{xx}(\lambda_j) = T\hat{\gamma}_{xx}(0)$$

and

$$\sum_{j=0}^{T-1} \cos\lambda_j[2\pi I_{xx}(\lambda_j)] = T[\hat{\gamma}_{xx}(1) + \hat{\gamma}_{xx}(T-1)],$$

which is the first sample (circulant) autocovariance of x_t. Therefore, the expression for $\hat{\alpha}_{x1}$ is (almost) identical to the one we would obtain in the time domain, which will be given by $\hat{\gamma}_{xx}(1)/\hat{\gamma}_{xx}(0)$, because $\hat{\gamma}_{xx}(T-1) = T^{-1}x_T x_1 = o_p(1)$.

Similar expressions would apply to the dynamic parameters that appear in $\boldsymbol{\theta}_{u_i}$ for a given value of $\mathbf{c}_{i\cdot}$ in view of (15b), since in this case it would be possible to estimate the variances of the innovations ψ_i in closed form.

Specifically, for an AR(1) example the partial derivatives of $G_{u_iu_i}(\lambda)$ with respect to ψ_i and α_{u_i1} would be

$$\begin{aligned}
\frac{\partial G_{u_iu_i}(\lambda)}{\partial \psi_i} &= \frac{1}{1+\alpha_{u_i1}^2 - 2\alpha_{u_i1}\cos\lambda}, \\
\frac{\partial G_{u_iu_i}(\lambda)}{\partial \alpha_{u_i1}} &= \frac{2(\cos\lambda - \alpha_{u_i1})\psi_i}{(1+\alpha_{u_i1}^2 - 2\alpha_{u_i1}\cos\lambda)^2}.
\end{aligned}$$

Hence, the corresponding log-likelihood scores would be

$$\begin{aligned}
\frac{\partial l(\mathbf{y}_i|\mathbf{x})}{\partial \psi_i} &= \frac{1}{2}\sum_{j=0}^{T-1} \frac{(1+\alpha_{u_i1}^2 - 2\alpha_{u_i1}\cos\lambda_j)^2}{(1+\alpha_{u_i1}^2 - 2\alpha_{u_i1}\cos\lambda_j)\,\psi_i^2}\left[2\pi z_j^{u_i} z_j^{u_i*} - \frac{\psi_i}{1+\alpha_{u_i1}^2 - 2\alpha_{u_i1}\cos\lambda_j}\right] \\
&= \frac{1}{2\psi_i^2}\sum_{j=0}^{T-1}\left[(1+\alpha_{u_i1}^2 - 2\alpha_{u_i1}\cos\lambda_j)2\pi z_j^{u_i} z_j^{u_i*} - \psi_i\right], \\
\frac{\partial l(\mathbf{y}_i|\mathbf{x})}{\partial \alpha_{u_i1}} &= \frac{1}{2}\sum_{j=0}^{T-1} \frac{2(\cos\lambda_j - \alpha_{u_i1})\psi_i(1+\alpha_{u_i1}^2 - 2\alpha_{u_i1}\cos\lambda_j)^2}{(1+\alpha_{u_i1}^2 - 2\alpha_{u_i1}\cos\lambda_j)^2\psi_i^2} \\
&\times \left[2\pi z_j^{u_i} z_j^{u_i*} - \frac{\psi_i}{(1+\alpha_{x1}^2 - 2\alpha_{x1}\cos\lambda_j)}\right] = \frac{2\pi}{\psi_i}\sum_{j=0}^{T-1}(\cos\lambda_j - \alpha_{u_i1}) z_j^{u_i} z_j^{u_i*}.
\end{aligned}$$

As a result, the spectral ML estimators of ψ_i and α_{u_i1} for fixed values of $\mathbf{c}_{i\cdot}$ would satisfy

$$\begin{aligned}
\tilde{\psi}_i &= \frac{2\pi}{T}\sum\nolimits_{j=0}^{T-1}(1+\tilde{\alpha}_{u_i1}^2 - 2\tilde{\alpha}_{u_i1}\cos\lambda_j) z_j^{u_i} z_j^{u_i*}, \\
\tilde{\alpha}_{u_i1} &= \frac{\sum_{j=0}^{T-1}\cos\lambda_j z_j^{u_i} z_j^{u_i*}}{\sum_{j=0}^{T-1} z_j^{u_i} z_j^{u_i*}}.
\end{aligned}$$

Intuitively, these parameter estimates are, respectively, the sample analogues to the variance of v_{it}, which is the residual variance in the regression of u_{it} on u_{it-1}, and the slope coefficient in the same regression.

Finally, (15c) would allow us to obtain the ML estimators of $\mathbf{c}_{i\cdot}$ for given values of $\boldsymbol{\theta}_{u_i}$. In particular, if we write together the derivatives for $c_{i,k}$ for $k = -m, \ldots, 0, \ldots, n$ we end up with the "weighted" normal equations:

$$\sum_{j=0}^{T-1}\left[G_{u_iu_i}^{-1}(\lambda_j)\begin{pmatrix} e^{im\lambda_j}z_j^x z_j^{x*}e^{-im\lambda_j}+e^{im\lambda_j}z_j^x z_j^{x*}e^{-im\lambda_j} & \cdots & e^{-in\lambda_j}z_j^x z_j^{x*}e^{-im\lambda_j}+e^{im\lambda_j}z_j^x z_j^{x*}e^{in\lambda_j} \\ \vdots & \ddots & \vdots \\ e^{im\lambda_j}z_j^x z_j^{x*}e^{in\lambda_j}+e^{-in\lambda_j}z_j^x z_j^{x*}e^{-im\lambda_j} & \cdots & e^{-in\lambda_j}z_j^x z_j^{x*}e^{in\lambda_j}+e^{-in\lambda_j}z_j^x z_j^{x*}e^{in\lambda_j} \end{pmatrix}\right]\begin{pmatrix}\tilde{c}_{i,-m}\\ \vdots \\ \tilde{c}_{i,n}\end{pmatrix}$$
$$=\sum_{j=0}^{T-1}G_{u_iu_i}^{-1}(\lambda_j)\begin{pmatrix} z_j^{y_i}z_j^{x*}e^{-im\lambda_j}+z_j^{y_i*}z_j^x e^{im\lambda_j}\\ \vdots \\ z_j^{y_i}z_j^{x*}e^{in\lambda_j}+z_j^{y_i*}z_j^x e^{-in\lambda_j}\end{pmatrix}.$$

Thus, unrestricted MLE's of $\mathbf{c}$ could be obtained from N univariate distributed lag weighted least squares regressions of each y_{it} on x_t that take into account the residual serial correlation in u_{it}. Interestingly, given that $G_{u_iu_i}(\lambda_j)$ is real, the above system of equations would not involve complex arithmetic. In addition, the terms in ψ_i would cancel, so the WLS procedure would only depend on the dynamic elements in $\boldsymbol{\theta}_{u_i}$.

Let us derive these expressions for the model in (1). In that case, the matrix on the left hand of the normal equations becomes

$$\sum_{j=0}^{T-1}G_{u_iu_i}^{-1}(\lambda_j)\begin{pmatrix} 2z_j^x z_j^{x*} & (e^{-i\lambda_j}+e^{i\lambda_j})z_j^x z_j^{x*}\\ (e^{i\lambda_j}+e^{-i\lambda_j})z_j^x z_j^{x*} & 2z_j^x z_j^{x*}\end{pmatrix}$$
$$=\sum_{j=0}^{T-1}G_{u_iu_i}^{-1}(\lambda_j)2z_j^x z_j^{x*}\begin{pmatrix}1 & \cos\lambda_j\\ \cos\lambda_j & 1\end{pmatrix},$$

while the vector on the right hand side will be

$$\sum_{j=0}^{T-1}G_{u_iu_i}^{-1}(\lambda_j)\begin{pmatrix} z_j^{y_i}z_j^{x*}+z_j^{y_i*}z_j^x\\ e^{i\lambda_j}z_j^{y_i}z_j^{x*}+e^{-i\lambda_j}z_j^{y_i*}z_j^x\end{pmatrix}.$$

In principle, we could carry out a zig-zag procedure that would estimate $\mathbf{c}_{i\cdot}$ and ψ_i for given $\boldsymbol{\theta}_{u_i}$, and then $\boldsymbol{\theta}_{u_i}$ for a given $\mathbf{c}_{i\cdot}$ and ψ_i. This would correspond to the spectral analogue to the Cochrane-Orcutt (1949) procedure. Obviously, iterations would be unnecessary when $\mathbf{G}_{\mathbf{uu}}(\lambda_j)$ is in fact constant, so that the idiosyncratic terms are static.

Unfortunately, we would have to resort once again to numerical optimisation in models with MA components. While this would be relatively costless if those components only appear in the common factor, it would be far more taxing if they also appeared in the idiosyncratic ones because there would be N such optimisations at each Cochrane-Orcutt iteration. For that reason, it would be useful to have a very fast way of estimating the parameters of processes with MA components, which would nevertheless remain asymptotically efficient.

3.2 Dealing with ARMA models by indirect inference

3.2.1 Pure MA terms

Consider the following MA(1) model

$$x_t = f_t - \beta f_{t-1}, \quad |\beta| < 1, \quad f_t|x_{t-1}, x_{t-2}, \ldots \sim N(0,1)$$

The simplest consistent estimator of β is an indirect inference (II) one based on the misspecified AR(1) auxiliary model

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad \varepsilon_t|x_{t-1}, x_{t-2}, \ldots \sim N(0,1)$$

(see e.g. Gouriéroux, Monfort and Renault (1993)), Chumacero (2001) or Ghysels, Khalaf and Vodounou (2003)). This estimator is equivalent to the GMM estimator of β based on

$$E[m_t(\beta)|\beta] = 0, \; m_t(\beta) = \left(x_t + \frac{\beta}{1+\beta^2} x_{t-1}\right) x_{t-1},$$

which coincides with the score of the AR(1) parameter ρ evaluated at the binding function

$$\rho(\beta) = -\frac{\beta}{1+\beta^2}.$$

We could increase the efficiency with which we estimate β by II if we considered higher order AR(k) models for $k \geq 2$. Unfortunately, for any finite order k those II estimators of β are generally inefficient relative to the ML estimator, which is effectively based on the moment condition

$$E[s_t(\beta)|\beta] = 0, \quad s_t(\beta) = [x_t - \nu_t(\beta)] \frac{\partial \nu_t(\beta)}{\partial \beta},$$

where

$$\nu_t(\beta) = E(x_t|x_{t-1}, x_{t-2}, \ldots; \beta) = -\sum_{j=1}^{\infty} \beta^j x_{t-j} = -\frac{\beta L}{1-\beta L} x_t$$

is the conditional mean of x_t given its past under the maintained assumption that the MA(1) process is invertible.

At first sight, it would appear that this highly non-linear estimator cannot be obtained by applying OLS to some auxiliary linear autoregressive model, but appearances can sometimes be misleading. Define

$$f_t(\beta) = x_t - \nu_t(\beta) = \sum_{j=0}^{\infty} \beta^j x_{t-j} = \frac{1}{1-\beta L} x_t$$

as the "innovations" in x_t. Similarly, let us use the shorthand notation

$$w_t(\beta) = -\frac{\partial \nu_{t+1}(\beta)}{\partial \beta} = \sum_{j=0}^{\infty} (j+1)\beta^j x_{t-j} = \frac{1}{(1-\beta L)^2} x_t.$$

We know that at the true value of β, say β_0, $f_t(\beta_0)$ will be white noise while $w_t(\beta_0)$ will be an AR(1). In addition, it is easy to see that

$$w_t(\beta) = \frac{1}{1-\beta L} f_t(\beta)$$

so that

$$f_t(\beta) = w_t(\beta) - \beta w_{t-1}(\beta).$$

Therefore, we can re-write the score of the MA(1) model as

$$s_t(\beta) = -[w_t(\beta) - \beta w_{t-1}(\beta)]w_{t-1}(\beta),$$

which coincides with the (minus) score of an AR(1) model for $w(\beta)$.

This regression is infeasible, but we can compute $\bar{\delta}_T$ as the OLS estimator in the regression of $w_t(\bar{\beta}_T)$ on $w_{t-1}(\bar{\beta}_T)$, where $\bar{\beta}_T$ is the II estimator of β based on the misspecified AR(1) auxiliary model for x_t.

Unfortunately, $\bar{\delta}_T$ is even less efficient than $\bar{\beta}_T$. Nevertheless, we can optimally combine those two different consistent but inefficient II estimators. Specifically, we can easily prove that $\tilde{\beta}_T = 2\bar{\beta}_T - \bar{\delta}_T$ is the outcome of a Gauss-Newton iteration, and therefore asymptotically equivalent to the ML estimator. In fact, it is possible to iterate the above procedure and obtain a new estimator $\bar{\delta}_T^1$ by regressing $w_t(\tilde{\beta}_T)$ on $w_{t-1}(\tilde{\beta}_T)$, which preserves asymptotic efficiency. The fixed point of these iterations is the ML estimator.

It turns out that Hannan (1969) proposed a simple iterative frequency domain procedure, which is effectively identical to the iterated indirect inference procedure we have just discussed.

3.2.2 Mixed models

Let us now consider the extension of our indirect inference procedure to the ARMA(1,1) model

$$x_t = \alpha x_{t-1} + f_t - \beta f_{t-1}, \quad |\alpha|, |\beta| < 1, \quad f_t|x_{t-1}, x_{t-2}, \ldots \sim N(0,1)$$

The simplest consistent estimator of α and β is an indirect inference one based on the misspecified AR(2) auxiliary model

$$x_t = \delta_1 x_{t-1} + \delta_2 x_{t-2} + w_t, \quad w_t|x_{t-1}, x_{t-2}, \ldots \sim N(0,1)$$

(see again Chumacero (2001)). This estimator is asymptotically equivalent to the GMM estimator of α and β based on the moment conditions

$$\begin{aligned} E[\mathbf{m}_t(\alpha,\beta)|\alpha,\beta] &= 0, \\ m_{1t}(\alpha,\beta) &= \left(x_t - \frac{(\alpha-\beta)(1-\alpha\beta)}{1-\alpha^2}x_{t-1}\right)x_{t-1}, \\ m_{2t}(\alpha,\beta) &= \left(x_t - \frac{\alpha(\alpha-\beta)(1-\alpha\beta)}{1-\alpha^2}x_{t-2}\right)x_{t-2}. \end{aligned}$$

The exactly identified nature of these moment conditions implies that the indirect inference estimator of α will coincide with the ratio of the second to the first autocorrelation of x_t, which is always between -1 and 1. As for the indirect inference estimator of β, we can obtain it from the first moment condition if we keep α fixed at its indirect inference value. In large samples, this procedure is effectively identical to the indirect inference estimator of β described in the previous section obtained by fitting an AR(1) model to the filtered series $\eta_t(\alpha) = x_t - \alpha x_{t-1}$.

Once again, we could increase the efficiency with which we estimate α and β if we considered higher order AR(k) models. Unfortunately, for any finite order k those II estimators are generally inefficient relative to the ML estimator, which is effectively based on the moment conditions

$$\begin{aligned} E[\mathbf{s}_t(\beta)|\beta] &= 0, \\ s_{\alpha t}(\alpha,\beta) &= [x_t - \nu_t(\beta)]\frac{\partial \nu_t(\alpha,\beta)}{\partial \alpha}, \\ s_{\beta t}(\alpha,\beta) &= [x_t - \nu_t(\beta)]\frac{\partial \nu_t(\alpha,\beta)}{\partial \beta} \end{aligned}$$

where

$$\nu_t(\alpha,\beta) = E(x_t|x_{t-1}, x_{t-2}, \ldots; \alpha, \beta) = (\alpha - \beta)\sum_{j=1}^{\infty} \beta^{j-1} x_{t-j} = \frac{(\alpha-\beta)L}{1-\beta L} x_t$$

This highly non-linear estimator can also be related to a couple of auxiliary linear autoregressive models. Specifically, define

$$f_t(\alpha,\beta) = x_t - \nu_t(\alpha,\beta) = \frac{1-\alpha L}{1-\beta L} x_t$$

as the "innovations" in x_t. Similarly, let us use the shorthand notation

$$\begin{aligned} r_t(\beta) &= \frac{\partial \nu_{t+1}(\alpha,\beta)}{\partial \alpha} = \frac{1}{(1-\beta L)} x_t \\ w_t(\alpha,\beta) &= -\frac{\partial \nu_{t+1}(\alpha,\beta)}{\partial \beta} = \frac{(1-\alpha L)}{(1-\beta L)^2} x_t. \end{aligned}$$

Then it is easy to see that

$$\begin{aligned} s_{\alpha t}(\alpha,\beta) &= [r_t(\beta) - \alpha r_{t-1}(\beta)] r_{t-1}(\beta) \\ s_{\beta t}(\alpha,\beta) &= -[w_t(\alpha,\beta) - \beta w_{t-1}(\alpha,\beta)] w_{t-1}(\alpha,\beta) \end{aligned}$$

so that we can estimate α for given β from the autoregression of $r_t(\beta)$ and β for given α from the autoregression of $w_t(\alpha,\beta)$. Again, these alternative indirect inference estimators will be inefficient when the unknown ARMA parameters are replaced by the indirect inference estimators $\bar{\alpha}_T$ and $\bar{\beta}_T$ based on the misspecified AR(2) auxiliary model for x_t, but we can combine them by means of a Gauss-Newton iteration of the form

$$\begin{pmatrix} \tilde{\alpha}_T \\ \tilde{\beta}_T \end{pmatrix} = \begin{pmatrix} \bar{\alpha}_T \\ \bar{\beta}_T \end{pmatrix} + \left\{ \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} r_{t-1}^2(\bar{\beta}_T) & -r_{t-1}(\bar{\beta}_T) w_{t-1}(\bar{\alpha}_T, \bar{\beta}_T) \\ -r_{t-1}(\bar{\beta}_T) w_{t-1}(\bar{\alpha}_T, \bar{\beta}_T) & w_{t-1}^2(\bar{\alpha}_T, \bar{\beta}_T) \end{bmatrix} \right\}^{-1}$$
$$\times \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} r_{t-1}(\bar{\beta}_T) f_t(\bar{\alpha}_T, \bar{\beta}_T) \\ -w_{t-1}(\bar{\alpha}_T, \bar{\beta}_T) f_t(\bar{\alpha}_T, \bar{\beta}_T) \end{bmatrix}.$$

Once again, it is possible to iterate the above procedure while preserving asymptotic efficiency, the fixed point of these iterations being the ML estimator.

Analogous procedures apply to general Arma(p,q) models[5] if we define

$$f_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\alpha_x(L)}{\beta_x(L)} x_t, \quad r_t(\boldsymbol{\beta}) = \frac{1}{\beta_x(L)} x_t, \quad w_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\alpha_x(L)}{\beta_x^2(L)} x_t.$$

Importantly, the variances, autocovariances and cross-covariances of the different filtered series can be computed much faster in the frequency domain than in the time domain, which makes these indirect inference estimators an ideal match to our spectral estimation techniques (see again Hannan (1969)).

3.3 Expected log-likelihood function

In practice, of course, we do not observe x_t. Nevertheless, the EM algorithm can be used to obtain values for $\boldsymbol{\theta}$ as close to the optimum as desired. At each iteration, the EM algorithm maximises the expected value of $l(\mathbf{y}_1, \ldots, \mathbf{y}_N|\mathbf{x}) + l(\mathbf{x})$ conditional on $\mathbf{Y}$ and the current parameter estimates, $\boldsymbol{\theta}^{(n)}$. The rationale stems from the fact that $l(\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{x})$ can also be factorized as $l(\mathbf{y}_1, \ldots, \mathbf{y}_N) + l(\mathbf{x}|\mathbf{y}_1, \ldots, \mathbf{y}_N)$. Since the expected value of the latter, conditional on $\mathbf{Y}$ and $\boldsymbol{\theta}^{(n)}$, reaches a maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(n)}$, *any increase* in the expected value of $l(\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{x})$ must represent an increase in $l(\mathbf{y}_1, \ldots, \mathbf{y}_N)$. This is the generalised EM principle.

In the E step we must compute

$$E[l(\mathbf{x})|\mathbf{Z}^{\mathbf{y}}, \boldsymbol{\theta}^{(n)}] = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1} \ln|G_{xx}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G_{xx}^{-1}(\lambda_j) E[z_j^x z_j^{x*}|\mathbf{Z}^{\mathbf{y}}, \boldsymbol{\theta}^{(n)}],$$
$$E[l(\mathbf{y}_i|\mathbf{x})|\mathbf{Z}^{\mathbf{y}}, \boldsymbol{\theta}^{(n)}] = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1} \ln|G_{u_i u_i}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G_{u_i u_i}^{-1}(\lambda_j) E[z_j^{u_i} z_j^{u_i*}|\mathbf{Z}^{\mathbf{y}}, \boldsymbol{\theta}^{(n)}].$$

But

$$E[z_j^x z_j^{x*}|\mathbf{Z}_y, \boldsymbol{\theta}^{(n)}] = z_j^{xK}(\boldsymbol{\theta}^{(n)}) z_j^{xK*}(\boldsymbol{\theta}^{(n)}) + E\left\{[z_j^x - z_j^{xK}(\boldsymbol{\theta}^{(n)})][z_j^{x*} - z_j^{x*}(\boldsymbol{\theta}^{(n)})]|\mathbf{z}_j^{\mathbf{y}}, \boldsymbol{\theta}^{(n)}\right\}$$
$$= I_{x^K x^K}^{(n)}(\lambda_j) + \omega^{(n)}(\lambda_j),$$

[5] The stationarity and strict invertibility of the estimated Ar and Ma polynomials in high order models could be achieved by reparametrising them in terms of partial autocorrelations, as in Barndorff-Nielsen and Schou (1973).

where

$$\begin{aligned} z_j^{x_K}(\boldsymbol{\theta}) = E[z_j^x|\mathbf{Z^y},\boldsymbol{\theta}] &= G_{xx}(\lambda_j)\mathbf{c}'(e^{i\lambda_j})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{z}_j^{\mathbf{y}}, \\ E\left\{[z_j^x - z_j^{x_K}(\boldsymbol{\theta})][z_j^{x_K*} - z_j^{x*}(\boldsymbol{\theta})]|\mathbf{Z^y},\boldsymbol{\theta}\right\} &= \omega(\lambda_j), \end{aligned}$$

and

$$\begin{aligned} I_{x^K x^K}(\lambda) &= 2\pi G_{xx}^2(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}) \\ &= 2\pi\omega^2(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}). \end{aligned} \tag{16}$$

is the periodogram of the smoothed values of the common factor.

In turn, if we define

$$\mathbf{I}_{\mathbf{y}x^K}(\lambda) = \mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})G_{xx}(\lambda) = \mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})\omega(\lambda)$$

as the cross-periodogram between the observed series $\mathbf{y}$ and the smoothed values of the common factor, we will have that

$$\begin{aligned} \mathbf{I}_{\mathbf{uu}}^{(\text{N})}(\lambda_j) &= E[\mathbf{z}_j^{\mathbf{u}}\mathbf{z}_j^{\mathbf{u}*}|\mathbf{Z^y},\boldsymbol{\theta}^{(n)}] = E\left\{\left[\mathbf{z}_j^{\mathbf{y}} - \mathbf{c}(e^{-i\lambda_j})z_j^x\right]\left[\mathbf{z}_j^{\mathbf{y}*} - z_j^{x*}\mathbf{c}'(e^{i\lambda_j})\right]|\mathbf{Z^y},\boldsymbol{\theta}^{(n)}\right\} \\ &= [\mathbf{z}_j^{\mathbf{y}} - \mathbf{c}(e^{-i\lambda_j})z_j^{x_K}(\boldsymbol{\theta}^{(n)})][\mathbf{z}_j^{\mathbf{y}*} - \mathbf{c}'(e^{i\lambda_j})z_j^{x_K*}(\boldsymbol{\theta}^{(n)})] + \mathbf{c}(e^{-i\lambda_j})\omega^{(n)}(\lambda_j)\mathbf{c}'(e^{i\lambda_j}) \\ &= \mathbf{I}_{\mathbf{yy}}(\lambda_j) - \mathbf{I}_{\mathbf{y}x^K}^{(n)}(\lambda)\mathbf{c}'(e^{i\lambda_j}) - \mathbf{c}(e^{-i\lambda_j})\mathbf{I}_{x^K\mathbf{y}}^{(n)}(\lambda) + \mathbf{c}(e^{-i\lambda_j})[I_{x^K x^K}^{(n)}(\lambda_j) + \omega^{(n)}(\lambda_j)]\mathbf{c}'(e^{i\lambda_j}), \end{aligned}$$

which resembles the expected value of $\mathbf{I}_{\mathbf{uu}}(\lambda_j)$ but the values at which the expectations are evaluated are generally different from the values at which the distributed lags are computed.

For the i^{th} series, this expression reduces to

$$\begin{aligned} I_{u_i u_i}^{(\text{N})}(\lambda_j) &= E[z_j^{u_i}z_j^{u_i*}|\mathbf{Z}_y,\boldsymbol{\theta}^{(n)}] = [z_j^{y_i} - c_i(e^{-i\lambda_j})z_j^{x_K}(\boldsymbol{\theta}^{(n)})][z_j^{y_i*} - c_i(e^{i\lambda_j})z_j^{x_K*}(\boldsymbol{\theta}^{(n)})] \\ &\quad + \omega^{(n)}(\lambda_j)c_i(e^{-i\lambda_j})c_i(e^{i\lambda_j}) \\ &= I_{y_i y_i}(\lambda_j) - c_i(e^{-i\lambda_j})I_{x^K y_i}^{(n)}(\lambda_j) - I_{y_i x^K}^{(n)}(\lambda_j)c_i(e^{i\lambda_j}) + [I_{x^K x^K}^{(n)}(\lambda_j) + \omega^{(n)}(\lambda_j)]c_i(e^{-i\lambda_j})c_i(e^{i\lambda_j}). \end{aligned}$$

Therefore, if we put all these expressions together we end up with

$$E[l(\mathbf{x})|\mathbf{Y},\boldsymbol{\theta}^{(n)}] = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1}\ln|G_{xx}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1}G_{xx}^{-1}(\lambda_j)\left[I_{x^K x^K}^{(n)}(\lambda_j) + \omega^{(n)}(\lambda_j)\right], \tag{17}$$

$$E[l(\mathbf{y}_i|\mathbf{x})|\mathbf{Y},\boldsymbol{\theta}^{(n)}] = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1}\ln|G_{u_i u_i}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1}G_{u_i u_i}^{-1}(\lambda_j)I_{u_i u_i}^{(\text{N})}(\lambda_j). \tag{18}$$

We can then maximise $E[l(\mathbf{x})|\mathbf{Y},\boldsymbol{\theta}^{(n)}]$ in (17) with respect to $\boldsymbol{\theta}_x$ to update those parameters. Similarly, we can maximise $E[l(\mathbf{y}_i|\mathbf{x})|\mathbf{Y},\boldsymbol{\theta}^{(n)}]$ in (18) with respect to $\mathbf{c}_{i\cdot}$, ψ_i and $\boldsymbol{\theta}_{u_i}$ to update those parameters.

In order to conduct those maximisations, we need the scores of the expected log-likelihood functions.

Given the similarity between (17) and (13), it is easy to see that

$$\frac{\partial E[l(\mathbf{x})|\mathbf{Y},\boldsymbol{\theta}^{(n)}]}{\partial \boldsymbol{\theta}_x} = \frac{1}{2}\sum_{j=0}^{T-1}\frac{\partial G_{xx}(\lambda_j)}{\partial \boldsymbol{\theta}_x}G_{xx}^{-2}(\lambda_j)\left\{2\pi\left[I_{x^K x^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)\right]-G_{xx}(\lambda_j)\right\},$$

which, not surprisingly, coincides with the the expected value of (15a) given $\mathbf{Y}$ and the current parameter estimates, $\boldsymbol{\theta}^{(n)}$.

In the case of an AR(1) common factor the expected log-likelihood score becomes

$$\frac{\partial E[l(\mathbf{x})|\mathbf{Y},\boldsymbol{\theta}^{(n)}]}{\partial \alpha_{x1}} = 2\pi\sum_{j=0}^{T-1}(\cos\lambda_j-\alpha_{x1})\left[I_{x^K x^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)\right].$$

As a result,

$$\hat{\alpha}_{x1}^{(n+1)} = \frac{\sum_{j=0}^{T-1}\cos\lambda_j\left[I_{x^K x^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)\right]}{\sum_{j=0}^{T-1}\left[I_{x^K x^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)\right]}.$$

It is also straightforward to modify the indirect inference procedures discussed in section 3.2 to handle models with ARMA terms if we replace the periodogram of the common factor by its expected value given observables, which coincides with sum of the periodogram of the smoothed values of the factor and its estimation error. Those periodograms can be obtained in no time in the E step of the algorithm from the minimal "sufficient statistics" discussed in section 2.4.

Similar expressions would apply to the dynamic parameters that appear in $\boldsymbol{\theta}_{u_i}$ and ψ_i for a given value of $\mathbf{c}_{i\cdot}$. Specifically, when the idiosyncratic terms follow AR(1) processes

$$\begin{aligned}\frac{\partial E[l(\mathbf{y}_i|\mathbf{x})|\mathbf{Y},\boldsymbol{\theta}^{(n)}]}{\partial \psi_i} &= \frac{1}{2\psi_i^2}\sum_{j=0}^{T-1}(1+\alpha_{u_i1}^2-2\alpha_{u_i1}\cos\lambda)\left[2\pi I_{u_iu_i}^{(\mathrm{N})}(\lambda_j)-\psi_i\right],\\ \frac{E[l(\mathbf{y}_i|\mathbf{x})|\mathbf{Y},\boldsymbol{\theta}^{(n)}]}{\partial \alpha_{u_i1}} &= \frac{2\pi}{\psi_i}\sum_{j=0}^{T-1}(\cos\lambda_j-\alpha_{u_i1})I_{u_iu_i}^{(\mathrm{N})}(\lambda_j).\end{aligned}$$

As a result, the spectral ML estimators of ψ_i and α_{u_i1} given $\mathbf{c}_{i\cdot}$ will satisfy

$$\begin{aligned}\hat{\psi}_i^{(n+1)} &= \frac{2\pi}{T}\sum\nolimits_{j=0}^{T-1}\left[1+\left(\hat{\alpha}_{u_i1}^{(n+1)}\right)^2-2\hat{\alpha}_{u_i1}^{(n+1)}\cos\lambda_j\right]I_{u_iu_i}^{(\mathrm{N})}(\lambda_j),\\ \hat{\alpha}_{u_i1}^{(n+1)} &= \frac{\sum_{j=0}^{T-1}\cos\lambda_j I_{u_iu_i}^{(\mathrm{N})}(\lambda_j)}{\sum_{j=0}^{T-1}I_{u_iu_i}^{(\mathrm{N})}(\lambda_j)}.\end{aligned}$$

Finally, the derivatives of (18) with respect to $c_{i,k}$ for $k=-m,\ldots,0,\ldots,n$ for fixed values of $\boldsymbol{\theta}_{u_i}$ will give rise to the modified "weighted" normal equations:

$$\sum_{j=0}^{T-1}\left[G_{u_iu_i}^{-1}(\lambda_j)\begin{pmatrix} e^{im\lambda_j}[I_{x^Kx^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)]e^{-im\lambda_j}+e^{im\lambda_j}[I_{x^Kx^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)]e^{-im\lambda_j} & \cdots & e^{-in\lambda_j}[I_{x^Kx^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)]e^{-im\lambda_j}+e^{im\lambda_j}[I_{x^Kx^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)]e^{in\lambda_j} \\ \vdots & \ddots & \vdots \\ e^{im\lambda_j}[I_{x^Kx^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)]e^{in\lambda_j}+e^{-in\lambda_j}[I_{x^Kx^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)]e^{-im\lambda_j} & \cdots & e^{-in\lambda_j}[I_{x^Kx^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)]e^{in\lambda_j}+e^{-in\lambda_j}[I_{x^Kx^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)]e^{in\lambda_j} \end{pmatrix}\right]\begin{pmatrix}\hat{c}_{i,-m}^{(n+1)} \\ \vdots \\ \hat{c}_{i,n}^{(n+1)}\end{pmatrix}$$

$$=\sum_{j=0}^{T-1}G_{u_iu_i}^{-1}(\lambda_j)\begin{pmatrix} I_{y_ix^K}^{(n)}(\lambda_j)e^{-im\lambda_j}+I_{x^Ky_i}^{(n)}(\lambda_j)e^{im\lambda_j} \\ \vdots \\ I_{y_ix^K}^{(n)}(\lambda_j)e^{in\lambda_j}+I_{x^Ky_i}^{(n)}(\lambda_j)e^{-in\lambda_j}\end{pmatrix}.$$

For the example in (2), the matrix on the left hand of the normal equations becomes

$$\sum_{j=0}^{T-1}G_{u_iu_i}^{-1}(\lambda_j)2[I_{x^Kx^K}^{(n)}(\lambda_j)+\omega^{(n)}(\lambda_j)]\begin{pmatrix}1 & \cos\lambda_j \\ \cos\lambda_j & 1\end{pmatrix},$$

while the vector on the right hand side will be

$$\sum_{j=0}^{T-1}G_{u_iu_i}^{-1}(\lambda_j)\begin{pmatrix} I_{y_ix^K}^{(n)}(\lambda_j)+I_{x^Ky_i}^{(n)}(\lambda_j) \\ e^{i\lambda_j}I_{y_ix^K}^{(n)}(\lambda_j)+e^{-i\lambda_j}I_{x^Ky_i}^{(n)}(\lambda_j)\end{pmatrix}.$$

In principle, we could carry out a zig-zag procedure that would estimate $\mathbf{c}_{i\cdot}$ and ψ_i for given $\boldsymbol{\theta}_{u_i}$ and $\boldsymbol{\theta}_{u_i}$ for a given $\mathbf{c}_{i\cdot}$ and ψ_i, although it is not clear that we really need to fully maximise the expected log-likelihood function at each EM iteration since the generalised EM principle simply requires us to increase it. Obviously, such iterations would be unnecessary when the idiosyncratic terms are static.

3.4 Alternative marginal scores

The EM principle can also be exploited to simplify the computation of the score. Since the Kullback inequality implies that $E\left[l(\mathbf{x}|\mathbf{Y};\boldsymbol{\theta})|\mathbf{Y};\boldsymbol{\theta}\right]=0$, it is clear that $\partial l(\mathbf{Y};\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ can be obtained as the expected value (given $\mathbf{Y}$ and $\boldsymbol{\theta}$) of the sum of the unobservable scores corresponding to $l(\mathbf{y}_1,\ldots,\mathbf{y}_N|\mathbf{x})$ and $l(\mathbf{x})$. This yields

$$\begin{aligned}
\frac{\partial l(\mathbf{Y})}{\partial\boldsymbol{\theta}_x} &= \frac{1}{2}\sum_{j=0}^{T-1}\frac{\partial G_{xx}(\lambda_j)}{\partial\boldsymbol{\theta}_x}G_{xx}^{-2}(\lambda_j)\left[2\pi E[z_j^x z_j^{x*}|\mathbf{Z}_y,\boldsymbol{\theta}]-G_{xx}(\lambda_j)\right] \\
\frac{\partial l(\mathbf{Y})}{\partial\boldsymbol{\theta}_{u_i}} &= \frac{1}{2}\sum_{j=0}^{T-1}\frac{\partial G_{u_iu_i}(\lambda_j)}{\partial\boldsymbol{\theta}_{u_i}}G_{u_iu_i}^{-2}(\lambda_j)\left[2\pi E[z_j^{u_i}z_j^{u_i*}|\mathbf{Z}_y,\boldsymbol{\theta}]-G_{u_iu_i}(\lambda_j)\right] \\
\frac{\partial l(\mathbf{Y})}{\partial c_{i,k}} &= \frac{2\pi}{2}\sum_{j=0}^{T-1}G_{u_iu_i}^{-1}(\lambda_j)\left[e^{ik\lambda_j}E[z_j^{u_i}z_j^{x*}|\mathbf{Z}_y,\boldsymbol{\theta}]+e^{-ik\lambda_j}E[z_j^x z_j^{u_i*}|\mathbf{Z}_y,\boldsymbol{\theta}]\right].
\end{aligned}$$

But since the scores are now evaluated at the values of the parameters at which the expectations are computed, we will have that

$$
\begin{aligned}
E[z_j^x z_j^{x*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] &= I_{x^K x^K}(\lambda_j)+\omega(\lambda_j),\\
E[\mathbf{z}_j^{\mathbf{u}}\mathbf{z}_j^{\mathbf{u}*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] &= E[\mathbf{z}_j^{\mathbf{u}}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]E[\mathbf{z}_j^{\mathbf{u}*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]+E\left[\left\{\mathbf{z}_j^{\mathbf{u}}-E[\mathbf{z}_j^{\mathbf{u}}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]\right\}\left\{\mathbf{z}_j^{\mathbf{u}*}-E[\mathbf{z}_j^{\mathbf{u}*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]\right\}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}\right]\\
&= I_{\mathbf{u}^K\mathbf{u}^K}(\lambda_j)+\mathbf{c}(e^{-i\lambda_j})\omega(\lambda_j)\mathbf{c}'(e^{i\lambda_j}).\\
E[\mathbf{z}_j^{\mathbf{u}} z_j^{x*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] &= E[\mathbf{z}_j^{\mathbf{u}}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]E[z_j^{x*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]+E\left[\left\{\mathbf{z}_j^{\mathbf{u}}-E[\mathbf{z}_j^{\mathbf{u}}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]\right\}\left\{z_j^{x*}-E[z_j^{x*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]\right\}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}\right]\\
&= I_{\mathbf{u}^K x^K}(\lambda_j)-\mathbf{c}(e^{-i\lambda_j})\omega(\lambda_j)
\end{aligned}
$$

where

$$
\begin{aligned}
z_j^{\mathbf{u}^K}=E[z_j^{\mathbf{u}}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] &= \mathbf{G}_{\mathbf{uu}}(\lambda_j)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{z}_j^{\mathbf{y}}=\mathbf{z}_j^{\mathbf{y}}-\mathbf{c}(e^{-i\lambda_j})z_j^{x^K},\\
E[(\mathbf{z}_j^{\mathbf{u}}-\mathbf{z}_j^{\mathbf{u}^K})(\mathbf{z}_j^{\mathbf{u}*}-\mathbf{z}_j^{\mathbf{u}^K*})|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] &= \mathbf{c}(e^{-i\lambda_j})\omega(\lambda_j)\mathbf{c}'(e^{i\lambda_j}),\\
E[(\mathbf{z}_j^{\mathbf{u}}-\mathbf{z}_j^{\mathbf{u}^K})(z_j^{x*}-z_j^{x^K*})|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] &= \mathbf{c}(e^{-i\lambda_j})\omega(\lambda_j),
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{I}_{\mathbf{u}^K\mathbf{u}^K}(\lambda)=2\pi\mathbf{G}_{\mathbf{uu}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{G}_{\mathbf{uu}}(\lambda)\\
&=2\pi\left[\mathbf{I}_N-\omega(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\right]\mathbf{I}_{\mathbf{yy}}(\lambda)\left[\mathbf{I}_N-\omega(\lambda)\mathbf{c}(e^{i\lambda})\mathbf{c}'(e^{-i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\right] \quad (19)
\end{aligned}
$$

is the periodogram of the smoothed values of the specific factors, and

$$
\begin{aligned}
\mathbf{I}_{x^K\mathbf{u}^K}(\lambda) &= 2\pi G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{G}_{\mathbf{uu}}(\lambda)\\
&= 2\pi\omega(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\left[\mathbf{I}_N-\omega(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\right] \quad (20)
\end{aligned}
$$

is the co-periodogram between $x_{t|\infty}^K$ and $\mathbf{u}_{t|\infty}^K$.

Tedious algebra shows that these scores coincide with the expressions in appendix A. They also closely related to the scores of the expected log-likelihoods in the previous subsection, but the difference is that the expectations were taken there with respect to the conditional distribution of $\mathbf{x}$ given $\mathbf{Y}$ evaluated at $\boldsymbol{\theta}^{(n)}$, not $\boldsymbol{\theta}$.

3.5 Some illustrations

We have generated samples of size $T = 100$ from the dynamic factor model in (1) in which common and idiosyncratic factors follow ARMA(1,1) processes. We carry out 5 Cochrane-Orcutt iterations only within each EM iteration. As starting values for the EM algorithm, we assume unit loadings on the contemporaneous and lagged values of the common factor, unit specific variances, and all autoregressive and moving average coefficients set to 0.5 and 0.1, respectively. These initial values are far away from the true parameters.

Figure 1 illustrates a typical example with ten series, while Figure 2 corresponds to a model with one hundred series. Remarkably, the first iteration of the EM yields a massive increase in the log-likelihood function in both cases. In addition, successive iterations also provide noticeable gains. As expected, though, the algorithm slows down considerably as we approach the optimum. Nevertheless, if we conduct a sufficiently large number of iterations, the value of the estimated parameters coincides with the estimates obtained by maximising the marginal log-likelihood function directly using the method of scoring with the analytical expressions for the score and information matrix in appendices A and B.

4 Common dynamics in sectoral employment

There is a long tradition of analysing comovements of sectoral activity indicators (see for example Abraham and Katz (1986), Lilien (1982) or Rogerson (1987)). In this context, dynamic factor models have proved useful in assessing the extent to which observed fluctuations in sectoral aggregates are accounted for by common sources of variation. In their seminal paper, Quah and Sargent (1993) studied the behavior of annual employment series across sixty US industries over the period 1948-1989. They found that the bulk of the time variation of the different sectors was explained by a common factor, and that their estimated measure of "business activity" captured aggregate dynamics in sectoral employment very well.

Motivated by their results, we downloaded employment series from the Bureau of Labor Statistics corresponding to the 81 NAICS 3-digit sectors, measured at monthly frequency and seasonally adjusted, for the period 1990M1-2014M4, which was the longest available (see Table 1 for the list of sectors). We decided to work with (annualised) growth rates for $T = 291$ months in view of the overwhelming evidence that the (log) levels of those employment series are nonstationary.[6]

Since our latent factor is meant to capture the common source of variation across sectoral employment growth rates, we followed Quah and Sargent (1993) and considered a dynamic single factor model. In order to determine the dynamic specification of common and specific factors, as well as the dynamic impact of the former on each sectoral series, we carried out some preliminary empirical analysis. Given that we expected the common factor to mimic the dynamics of total nonfarm employment, we fitted univariate ARMA models of various orders to the (geometric) rate of growth of this observed series. We found that an ARMA(1,1) yields the best fit according to both the Schwartz and Akaike criteria. Next we regressed the demeaned

[6]A preliminary check on data quality indicated that a handful of series display abnormal values. We treated them as additive outliers in the (log) levels and replaced them by the average of the adjacent observations, which is a simple fillter that is nevertheless optimal under the random walk hypothesis.

changes of employment on the demeaned contemporaneous and one-month lagged changes of total nonfarm employment. We found that the coefficients associated to the lagged changes were significantly different from zero for a sizeable fraction of the series, which strongly suggests that the sector-specific employment growth rates may be driven not only by the contemporaneous value of the latent factor but also by its one-month lagged values. In addition, we conducted LM tests for first- and second-order residual autocorrelation to assess whether the idiosyncratic disturbances are likely to be serially correlated. We found that roughly 2/3 of the series require serially correlated idiosyncratic terms.

In view of those findings, we began by estimating a special case of model (1) in which both x_t and x_{t-1} heterogeneously affect each of the sectoral growth rates, x_t follows an ARMA(1,1) process while the idiosyncratic terms u_{it} follow simple AR(1)'s. Individual tests for $H_0 : \alpha_i = 0$ indicated that there are 35 series for which the white noise hypothesis is not rejected,[7] which we decided to impose thenceforth. For the remaining 46 series we jointly tested the null of AR(1) against ARMA(1,1) specific factors, the likelihood ratio statistic taking the extremely significant value of 1369.9.[8]

Estimation of the final model with 46 ARMA(1,1) and 35 white noise idiosyncratic processes was conducted by means of the EM algorithm using the iterated indirect procedure discussed in previous sections. As starting values, we assumed again unit loadings on the contemporaneous and lagged values of the common factor, unit specific variances, and autoregressive and moving average coefficients set 0.5 and 0.1, respectively, for both common and idiosyncratic factors. In order to speed up the EM iterations, we conducted five Cochrane-Orcutt iterations only instead of continuing until convergence. Despite the hundreds of parameters involved, this procedure worked very well to begin with. Eventually, though, the norm of the gradients corresponding to the idiosyncratic parameters of three series reached a positive lower bound. A careful inspection suggested that the corresponding AR and MA coefficients were probably too close to each other, and the resulting quasi cancellations made the likelihood function rather flat. For that reason, we switched to an alternative, slower version of the EM algorithm that replaced our iterative indirect inference procedure by the direct maximisation of the expected log-likelihood function in (18) using a scoring algorithm with analytical derivatives and information matrix. Although the estimated parameters did not change much, the log-likelihood function improved slightly and the norm of the gradients went down all the way to 0.

Finally, we computed standard errors of the parameter estimators using the analytical ex-

[7] The series are: 5 8 18 19 22 23 26 28 32 34 35 36 37 38 39 40 41 43 45 47 48 51 54 58 62 63 65 66 70 71 73 75 77 79 81, which by and large coincide with the LM tests computed for the total nonfarm regressions

[8] See Fiorentini and Sentana (2013) for computationally simple and intuitive individual and joint LM tests for neglected serial correlation in common and specific factors.

pressions for the information matrix in appendix B. The estimation results are reported in Tables 2 and 3.

As is well known, the usual Wiener-Kolmogorov filter can lead to filtering distortions at both ends of the sample. For that reason, we wrote the model in a state-space form and applied the Kalman filter in the time domain with exact initial conditions derived from the stationary distribution of the 165 state variables (3 for the common factor and 2 for each of the idiosyncratic ones; see appendix C for details).[9] Given that the standard fixed interval smoother was numerically unstable with such a big state vector, we used the modified Bryson-Frazier smoother instead (see Bierman (1977)). Figure 3 plots the yearly growth rate of total nonfarm employment (red dashed line) and our estimated employment factor (solid blue line). Importantly, our smoothed factor tracks remarkably well the actual growth rate of aggregate employment, specially during recession phases, such as in 1991, 2001, and 2009, although it is unsurprisingly smoother than the observed series.

5 Directions for further work

The spectral EM algorithm developed in the previous sections can be extended in several interesting directions. One obvious possibility would be to models with multiple common factors. Although this would be intensive in notation, such an extension would be otherwise straightforward after dealing with identification issues before estimating the model. In fact, in a follow up paper (Fiorentini, Galesi and Sentana (2015)) we consider models with two levels of factors:

1. Pervasive common factors that affect all N series
2. Factors that only affect a subset of the series, such as the ones belonging to the same country or region.

The main complication is keeping track of what factors affect which series.

Another interesting extension would deal with models in which the heterogeneous dynamic impact of the common factor on each observed variable, which is characterised by the $c_i(L)$ polynomials, can be represented by the ratio of two low order polynomials (see Hannan (1965) and Hannan and Nichols (1972) for frequency domain estimators of some rational distributed lag models when x_t is observable).

[9] The main difference between the Wiener-Kolmogorov filtered values, $x^K_{t|\infty}$, and the Kalman filter smoothed values, $x^K_{t|T}$, results from the implicit dependence of the former on a doubly infinite sequence of past and future observations. As shown by Fiorentini (1995) and Gómez (1999), though, they can be made numerically identical by replacing both pre- and post- sample observations by their least squares projections onto the linear span of the sample observations.

Given their ubiquitousness in the recent empirical literature (see e.g. Bai and Ng (2008) and the references therein), the extension of our methods to approximate factor models in which (i) the cross-sectional dimension is non-negligible relative to the time series dimension; and (ii) there is some mild contemporaneous and dynamic correlation between the idiosyncratic terms would constitute a very valuable addition. In fact, a very large number of series constitutes a computational blessing in this framework, because for large N the unobservable factors will be consistently estimated by the Kalman-Wiener-Kolmogorov filter, and the model effectively becomes a multivariate regression model. In this regard, Doz, Giannone and Reichlin (2012) have recently proved the consistency of the Gaussian pseudo ML estimators that we have used in such contexts. In principle, we could easily extend our numerical procedures to models with non-diagonal idiosyncratic spectral density matrices because in those models the factorisation of the complete log-likelihood function of observed series and common factors will still be true. However, we would have to use frequency domain versions of multivariate regressions in those contexts, whose efficient estimation deserve further consideration.

Finally, it is worth mentioning that although we have exploited some specificities of dynamic factor models, our procedures can be easily extended to most unobserved components time series processes, and in particular to Ucarima models (see Fiorentini and Sentana (2014) for a closely related analysis) and the state-space models underlying the recent nowcasting literature (see Banbura, Giannone and Reichlin (2011) and the references therein). We are currently pursuing some of these research avenues.

References

Abraham, K. G., and Katz, L. F. (1986): "Cyclical unemployment: sectoral shifts or aggregate disturbances", *Journal of Political Economy* 94, 507-522.

Bai, J. and Ng, S. (2008): "Large dimensional factor analysis", *Foundations and Trends in Econometrics* 3, 89–163.

Banbura, M., Giannone, D., and Reichlin, L. (2011): "Nowcasting", in M.P. Clemens and D.F. Hendry (eds.), *Oxford handbook on economic forecasting*, 193-224, Oxford University Press.

Barndorff-Nielsen, O. and Schou, G. (1973): "On the parametrization of autoregressive models by partial autocorrelations", *Journal of Multivariate Analysis* 3, 408–419.

Bierman, G.J. (1977) *Factorization methods for discrete sequential estimation*, Academic Press, New York.

Choudhuri , N., Ghosal S. and Roy, A. (2004): "Contiguity of the Whittle measure for a Gaussian time series", *Biometrika* 91, 211-218.

Chumacero, R. (2001): "Estimating ARMA models efficiently", Central Bank of Chile Working Paper 92.

Cochrane, and Orcutt, (1949): "Application of least squares regression to relationships containing auto-correlated error terms", *Journal of the American Statistical Association* 44, 32-61.

Demos, A., and Sentana, E. (1998): "An EM algorithm for conditionally heteroskedastic factor models", *Journal of Business and Economic Statistics* 16, 357-361.

Dempster, A., Laird, N., and Rubin, D. (1977): "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society B* 39, 1-38.

Diebold, F.X. and Rudebusch, G.D. (1996), "Measuring business cycles: a modern perspective", *Review of Economics and Statistics*, 78, 67-77.

Diebold, F.X., Rudebusch, G.D. and Aruoba, B. (2006), "The macroeconomy and the yield curve: a dynamic latent factor approach", *Journal of Econometrics* 131, 309-338.

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.

Dungey, M., Martin, V.L. and Pagan, A.R. (2000): "A multivariate latent factor decomposition of international bond yield spreads", *Journal of Applied Econometrics* 15, 697-715.

Dunsmuir, W. (1979): "A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise", *Annals of Statistics* 7, 490-506.

Dunsmuir, W. and Hannan, E.J. (1976): "Vector linear time series models", *Advances in Applied Probability* 8, 339-364.

Engle, R.F., Hendry, D.F. and Richard, J.-F. (1983): “Exogeneity”, *Econometrica*, 277-304.

Engle, R.F. and Watson, M.W. (1981): “A one-factor multivariate time series model of metropolitan wage rates”, *Journal of the American Statistical Association* 76, 774-781.

Fiorentini, G. (1995): *Conditional heteroskedasticity: some results on estimation, inference and signal extraction, with an application to seasonal adjustment*, unpublished Doctoral Dissertation, European University Institute.

Fiorentini, G. and Sentana, E. (2013): “Dynamic specification tests for dynamic factor models”, CEMFI Working Paper 1306.

Fiorentini, G. and Sentana, E. (2014): “Neglected serial correlation tests in UCARIMA models”, CEMFI Working Paper 1406.

Fiorentini, G., Sentana, E. and Shephard, N. (2004): “Likelihood estimation of latent generalised ARCH structures”, *Econometrica* 72, 1481-1517.

Fiorentini, G., Galesi, A. and Sentana (2015): “A flexible dynamic factor model for international business cycles”, mimeo, CEMFI.

Geweke, J.F. (1977): “The dynamic factor analysis of economic time series models”, in D. Aigner and A. Goldberger (eds.), *Latent variables in socioeconomic models*, 365-383, North-Holland.

Geweke, J.F. and Singleton, K.J (1981): “Maximum likelihood "confirmatory" factor analysis of economic time series”, *International Economic Review* 22, 37-54.

Ghysels, E., Khalaf, L. and Vodounou, C. (2003): “Simulation based inference in moving average models”, *Annales d’Economie et de Statistique* 69, 85-99.

Gómez, V. (1999): “Three equivalent methods for filtering finite nonstationary time series”, *Journal of Business and Economic Statistics* 17, 109-116.

Gouriéroux, C., Monfort, A. and Renault, E. (1991): “A general framework for factor models”, mimeo, INSEE.

Gouriéroux, C., Monfort, A. and Renault, E. (1993): “Indirect inference”, *Journal of Applied Econometrics* 8, S85-S118.

Gregory, A.W., Head, A.C. and Raynauld, J. (1997): “Measuring world business cycles”, *International Economic Review* 38, 677-701.

Hamilton, J. (1990): “Analysis of time series subject to changes in regime”, *Journal of Econometrics* 45, 39-70

Hannan, E.J. (1965): “The estimation of relationships involving distributed lags”, *Econometrica* 33, 206-224.

Hannan, E.J. (1969): "The estimation of mixed moving average autoregressive systems", *Biometrika* 56, 579-593.

Hannan, E.J. (1973): "The asymptotic theory of linear time series models", *Journal of Applied Probability* 10, 130-145 (Corrigendum 913).

Hannan, E.J. and Nicholls, D.F. (1972): "The estimation of mixed regression, autoregression, moving average, and distributed lag models", *Econometrica* 40, 529-547.

Harvey, A.C. (1989): *Forecasting, structural models and the Kalman filter*, Cambridge University Press.

Heaton, C. and Solo, V. (2004): "Identification of causal factor models of stationary time series", *Econometrics Journal*, 7, 618-627.

Jegadeesh, N. and G.G. Pennacchi (1996): "The behavior of interest rates implied by the term structure of Eurodollar futures", *Journal of Money, Credit and Banking* 28, 426-446.

Jungbacker, B. and S.J. Koopman (2008): "Likelihood-based analysis for dynamic factor models", Tinbergen Institute Discussion Paper 2008-0007.

Lawley, D.N. and Maxwell, A.E. (1971): *Factor analysis as a statistical method*, 2nd ed., Butterworths.

Lehmann, B.N., and Modest, D.M. (1988): "The empirical foundations of the arbitrage pricing theory ", *Journal of Financial Economics* 21, 213-254.

Lilien, D. M. (1982): "Sectoral shifts and cyclical unemployment", *Journal of Political Economy* 90, 777-793.

Litterman, R. and Sargent, T.J. (1979): "Detecting neutral price level changes and the effects of aggregate demand with index models", Research Department Working Paper 125, Federal Reserve Bank of Minneapolis.

Magnus, J.R. (1988): *Linear structures*, Oxford University Press.

Magnus, J.R. and Neudecker, H. (1988): *Matrix differential calculus with applications in Statistics and Econometrics*, Wiley.

Mody, A. and Taylor, M.P. (2007): "Regional vulnerability: the case of East Asia", *Journal of International Money and Finance* 26, 1292-1310.

Peña, D. and Box, G.E.P. (1987): "Identifying a simplifying structure in time series", *Journal of the American Statistical Association* 82, 836-843.

Quah, D. and Sargent, T. (1993): "A dynamic index model for large cross sections", in Stock, J.H. and Watson, M.W. (eds) *Business cycles, indicators and forecasting*, 285-310, University of Chicago Press.

Rogerson, R. (1987): “An equilibrium model of sectoral reallocation”, *Journal of Political Economy* 95, 824-834.

Rubin, D. and Thayer, D. (1982): “EM algorithms for ML factor analysis”, *Psychometrika* 47, 69-76.

Ruud, P. (1991): “Extensions of estimation methods using the EM algorithm”, *Journal of Econometrics* 49, 305-341.

Sargent, T.J., and Sims, C.A. (1977): “Business cycle modeling without pretending to have too much a priori economic theory”, *New methods in business cycle research* 1, 145-168.

Sentana, E. (2000): “The likelihood function of conditionally heteroskedastic factor models”, *Annales d’Economie et de Statistique* 58, 1-19.

Sentana, E. (2004): “Factor representing portfolios in large asset markets”, *Journal of Econometrics* 119, 257-289.

Singleton, K.J. (1981): “A latent time series model of the cyclical behavior of interest rates”, *International Economic Review* 21, 559-575.

Stock, J.H. and Watson, M.W. (1989): “New indexes of coincident and leading economic indicators”, *NBER Macroeconomics Annual* 4, 351-394.

Stock, J., and Watson, M. (1991), “A probability model of the coincident economic indicators”, in K. Lahiri and G. Moore (eds.) *Leading economic indicators: new approaches and forecasting records*, Cambridge University Press.

Stock, J.H. and Watson, M.W. (1993): “A procedure for predicting recessions with leading indicators: econometric issues and recent experience”, in J.H. Stock and M.W. Watson (eds.) *Business cycles, indicators, and forecasting*, 95-153, University of Chicago Press.

Watson, M.W. and Engle, R.F. (1983): “Alternative algorithms for estimation of dynamic MIMIC, factor, and time varying coefficient regression models”, *Journal of Econometrics* 23, 385-400.

Watson, M.W. and Kraft, D.F. (1984): “Testing the interpretation of indices in a macroeconomic index model”, *Journal of Monetary Economics* 13, 165-182.

Whittle, P. (1962): “Gaussian estimation in stationary time series”, *Bulletin of the International Statistical Institute* 39, 105-129.

Appendices

A Spectral scores

The score function for all the parameters other than the mean is given by (7). Since

$$d\mathbf{G}_{\mathbf{yy}}(\lambda) = d\mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda}) + \mathbf{c}(e^{-i\lambda})dG_{xx}(\lambda)\mathbf{c}'(e^{i\lambda}) + \mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)d\mathbf{c}'(e^{i\lambda}) + d\mathbf{G}_{\mathbf{uu}}(\lambda)$$

(see Magnus and Neudecker (1988)), it immediately follows that

$$\begin{aligned} dvec\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right] &= \left[\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)\otimes\mathbf{I}_N\right]d\mathbf{c}(e^{-i\lambda}) + \left[\mathbf{I}_N\otimes\mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\right]d\mathbf{c}(e^{i\lambda}) \\ &+ \left[\mathbf{c}(e^{i\lambda})\otimes\mathbf{c}(e^{-i\lambda})\right]dG_{xx}(\lambda) + \mathbf{E}_N dvecd\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right], \end{aligned}$$

where $\mathbf{E}'_N = (\mathbf{e}_1\mathbf{e}'_1|\ldots|\mathbf{e}_N\mathbf{e}'_N)$, with $(\mathbf{e}_1|\ldots|\mathbf{e}_N) = \mathbf{I}_N$, is the unique $N^2\times N$ "diagonalisation" matrix that transforms $vec(\mathbf{A})$ into $vecd(\mathbf{A})$ as $vecd(\mathbf{A}) = \mathbf{E}'_N vec(\mathbf{A})$, and $\mathbf{K}_{mn}$ is the commutation matrix of orders m and n (see Magnus (1988)). But

$$\mathbf{c}(e^{-i\lambda}) = \sum_{m=0}^{M}\mathbf{c}_m(\boldsymbol{\theta})e^{-im\lambda}, \tag{A1}$$

so

$$d\mathbf{c}(e^{-i\lambda}) = \sum_{m=0}^{M}d\mathbf{c}_m(\boldsymbol{\theta})e^{-im\lambda}.$$

Consequently, we can write

$$\begin{aligned} dvec\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right] &= \sum_{m=0}^{M}\left\{\left[e^{-im\lambda}\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)\otimes\mathbf{I}_N\right] + \left[\mathbf{I}_N\otimes e^{im\lambda}\mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\right]\right\}d\mathbf{c}_m(\boldsymbol{\theta}) \\ &+ \left[\mathbf{c}(e^{i\lambda})\otimes\mathbf{c}(e^{-i\lambda})\right]dG_{xx}(\lambda) + \mathbf{E}_N dvecd\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right]. \end{aligned}$$

Hence, the Jacobian of $vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]$ will be

$$\begin{aligned} \frac{\partial vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial\boldsymbol{\theta}'} &= \sum_{m=0}^{M}\left\{\begin{array}{c}\left[e^{-im\lambda}\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)\otimes\mathbf{I}_N\right] \\ +\left[\mathbf{I}_N\otimes e^{im\lambda}\mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\right]\end{array}\right\}\frac{\partial\mathbf{c}_m}{\partial\boldsymbol{\theta}'} \\ &+ \left[\mathbf{c}(e^{i\lambda})\otimes\mathbf{c}(e^{-i\lambda})\right]\frac{\partial G_{xx}(\lambda)}{\partial\boldsymbol{\theta}'} + \mathbf{E}_N\frac{\partial vecd\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right]}{\partial\boldsymbol{\theta}'}. \end{aligned} \tag{A2}$$

If we combine this expression with the fact that

$$\begin{aligned} &\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\right]vec\left[\mathbf{z}^{\mathbf{y}c}_j\mathbf{z}^{\mathbf{y}\prime}_j - \mathbf{G}'_{\mathbf{yy}}(\lambda_j)\right] \\ = \;& vec\left[2\pi\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{z}^{\mathbf{y}c}_j\mathbf{z}^{\mathbf{y}\prime}_j\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda) - \mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right] \end{aligned}$$

and $\mathbf{I}'_{\mathbf{yy}}(\lambda)=\mathbf{z}_j^{\mathbf{y}c}\mathbf{z}_j^{\mathbf{y}\prime}$ we obtain:

$$
\begin{aligned}
2\mathbf{d}(\lambda;\boldsymbol{\theta}) &= \sum_{m=0}^{M}\frac{\partial \mathbf{c}'_m}{\partial\boldsymbol{\theta}}\left\{\begin{array}{c}\left[e^{-im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\otimes\mathbf{I}_N\right]\\ +\left[\mathbf{I}_N\otimes e^{im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{-i\lambda})\right]\end{array}\right\}vec\left[2\pi\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)-\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right]\\
&\quad+\frac{\partial G_{xx}(\lambda)}{\partial\boldsymbol{\theta}}\left[\mathbf{c}'(e^{i\lambda})\otimes\mathbf{c}'(e^{-i\lambda})\right]vec\left[2\pi\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)-\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right]\\
&\quad+\frac{\partial vecd'\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right]}{\partial\boldsymbol{\theta}}\mathbf{E}_N vec\left[2\pi\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)-\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right]\\
&= \sum_{m=0}^{M}\frac{\partial \mathbf{c}'_m}{\partial\boldsymbol{\theta}}vec\left[\begin{array}{c}2\pi\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)e^{-im\lambda}-\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)e^{-im\lambda}\\ +2\pi e^{im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)-e^{im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\end{array}\right]\\
&\quad+\frac{\partial G_{xx}(\lambda)}{\partial\boldsymbol{\theta}}vec\left[2\pi\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{c}(e^{i\lambda})-\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{c}(e^{i\lambda})\right]\\
&\quad+\frac{\partial vecd'\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right]}{\partial\boldsymbol{\theta}}\mathbf{E}_N vec\left[2\pi\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)-\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right].
\end{aligned}
$$

Let us now try to interpret the different components of this expression. The first thing to note is that

$$
e^{-im\lambda}vec\left[2\pi\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)-\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)\right]
$$

and

$$
e^{im\lambda}vec\left[2\pi G_{xx}(\lambda)\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)-G_{xx}(\lambda)\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right]
$$

are complex conjugates because the conjugate of a product is the product of the conjugates, so it suffices to analyse one of them.

If we further assume that $G_{xx}(\lambda)>0$ and $\mathbf{G}_{\mathbf{uu}}(\lambda)>\mathbf{0}$ we can write

$$
\begin{aligned}
2\pi\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)&\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)e^{-im\lambda}-\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)e^{-im\lambda}\\
&=\mathbf{G}'^{-1}_{\mathbf{uu}}(\lambda)\left[2\pi e^{-im\lambda}\mathbf{I}'_{x^K\mathbf{u}^K}(\lambda)-e^{-im\lambda}\mathbf{G}'_{x^K\mathbf{u}^K}(\lambda)\right],\\
2\pi\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)&\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{c}(e^{i\lambda})-\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{c}(e^{i\lambda})\\
&=G^{-2}_{xx}(\lambda)\left[2\pi I_{x^Kx^K}(\lambda)-G_{x^Kx^K}(\lambda)\right]
\end{aligned}
$$

and

$$
2\pi\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)-\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)=\mathbf{G}'^{-1}_{\mathbf{uu}}(\lambda)\left[2\pi\mathbf{I}'_{\mathbf{u}^K\mathbf{u}^K}(\lambda)-\mathbf{G}'_{\mathbf{u}^K\mathbf{u}^K}(\lambda)\right]\mathbf{G}'^{-1}_{\mathbf{uu}}(\lambda).
$$

Therefore, the component of the score associated to $\mathbf{c}_m$ will be the sum across frequencies of terms of the form

$$
\mathbf{G}'^{-1}_{\mathbf{uu}}(\lambda)\left[2\pi e^{-im\lambda}\mathbf{I}'_{x^K\mathbf{u}^K}(\lambda)-e^{-im\lambda}\mathbf{G}'_{x^K\mathbf{u}^K}(\lambda)\right]
$$

(and their conjugate transposes), which capture the difference between the cross-periodogram and cross-spectrum of x^K_{t-m} and u^K_{it} inversely weighted by the spectral density of u_{it}. As a result, we can understand this term as arising from the normal equation in the spectral regression of y_{it} onto x_{t-m} but taking into account the unobservability of the regressor.

Similarly, the component of the score associated to the parameters that determine $G_{xx}(\lambda)$ will be the cross-product across frequencies of the product of the derivatives of the spectral density of x_t with the difference between the periodogram and spectrum of x_t^K inversely weighted by the squared spectral density of x_t. In this case, we can interpret this term as arising from a marginal log-likelihood function for x_t that takes into account the unobservability of x_t.

Finally, the component of the score associated to the parameters that determine $G_{u_iu_i}(\lambda)$ will be the cross-product across frequencies of the product of the derivatives of the spectral density of u_{it} with the difference between the periodogram and spectrum of u_{it}^K inversely weighted by the squared spectral density of u_{it}. Once again, we can interpret this term as arising from the conditional log-likelihood function of u_{it} given x_t that takes into account the unobservability of u_{t_i}.

As usual, we can then exploit the Woodbury formula, as in expressions (16), (19) and (20), to greatly speed up the computations. In particular, we will get

$$
\begin{aligned}
& G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda) - G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda) \\
&= G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\left[\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda) - \omega(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\right] - \omega(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda) \\
&= G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\left[\mathbf{I}_N - \omega(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\right]\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda) - \omega(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda) \\
&= \left[\mathbf{I}_{x^K\mathbf{u}^K}(\lambda) - \omega(\lambda)\mathbf{c}'(e^{i\lambda})\right]\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda),
\end{aligned}
$$

$$
\begin{aligned}
& \mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda) - \mathbf{G}_{\mathbf{yy}}^{-1}(\lambda) \\
&= \left[\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda) - \omega(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\right]\mathbf{I}_{\mathbf{yy}}(\lambda)\left[\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda) - \omega(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\right] \\
& \quad - \left[\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda) - \omega(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\right] \\
& \mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\left[\mathbf{I}_N - \omega(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\right]\mathbf{I}_{\mathbf{yy}}(\lambda)\left[\mathbf{I}_N - \omega(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\right]\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda) \\
& -\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\left[\mathbf{G}_{\mathbf{uu}}(\lambda) - \omega(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\right]\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda) \\
&= \mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\left[\mathbf{I}_{\mathbf{u}^K\mathbf{u}^K}(\lambda) - \mathbf{G}_{\mathbf{u}^K\mathbf{u}^K}(\lambda)\right]\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda),
\end{aligned}
$$

and

$$
\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}) - \mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}) = G_{xx}^{-1}(\lambda)[I_{x^Kx^K}(\lambda) - G_{x^Kx^K}(\lambda)]G_{xx}^{-1}(\lambda).
$$

B Spectral information matrix

Given the expression for the Jacobian matrix (A2), we will have that

$$\begin{aligned}\frac{\partial vec'\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial\boldsymbol{\theta}} &= \sum_{m=0}^{M}\frac{\partial\mathbf{c}'_m}{\partial\boldsymbol{\theta}}\left\{\begin{array}{c}\left[e^{-im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\otimes\mathbf{I}_N\right]\\ +\left[\mathbf{I}_N\otimes e^{im\lambda}\mathbf{c}'(e^{-i\lambda})G_{xx}(\lambda)\right]\end{array}\right\}\\ &\quad+\frac{\partial G_{xx}(\lambda)}{\partial\boldsymbol{\theta}}\left[\mathbf{c}'(e^{i\lambda})\otimes\mathbf{c}'(e^{-i\lambda})\right]+\frac{\partial vecd\left[\mathbf{G_{uu}}(\lambda)\right]}{\partial\boldsymbol{\theta}}\mathbf{E}'_N\end{aligned}$$

and

$$\begin{aligned}\left\{\frac{\partial vec'\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial\boldsymbol{\theta}}\right\}^* &= \sum_{m=0}^{M}\left\{\begin{array}{c}\left[e^{im\lambda}\mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\otimes\mathbf{I}_N\right]\\ +\left[\mathbf{I}_N\otimes e^{-im\lambda}\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)\right]\end{array}\right\}\frac{\partial\mathbf{c}_m}{\partial\boldsymbol{\theta}'}\\ &\quad+\left[\mathbf{c}(e^{-i\lambda})\otimes\mathbf{c}(e^{i\lambda})\right]\frac{\partial G_{xx}(\lambda)}{\partial\boldsymbol{\theta}'}+\mathbf{E}_N\frac{\partial vecd\left[\mathbf{G_{uu}}(\lambda)\right]}{\partial\boldsymbol{\theta}'}.\end{aligned}$$

Hence, it is straightforward to see that the elements of the block of the information matrix (10) corresponding to the dynamic factor loadings will be

$$\begin{aligned}&\frac{\partial vec'[\mathbf{G_{yy}}(\lambda)]}{\partial\boldsymbol{c}_m}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\right]\left\{\frac{\partial vec'\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial\mathbf{c}_n}\right\}^*\\ =&\left\{\begin{array}{c}\left[e^{-im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\otimes\mathbf{I}_N\right]\\ +\left[\mathbf{I}_N\otimes e^{im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{-i\lambda})\right]\end{array}\right\}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\right]\left\{\begin{array}{c}\left[e^{in\lambda}\mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\otimes\mathbf{I}_N\right]\\ +\left[\mathbf{I}_N\otimes e^{-in\lambda}\mathbf{c}(e^{i\lambda})G_{xx}(\lambda)\right]\end{array}\right\}\\ =&\ G^2_{xx}(\lambda)\left\{\begin{array}{c}e^{-i(m+n)\lambda}\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{c}(e^{i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\\ +e^{i(m+n)\lambda}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\right]\\ e^{-i(m-n)\lambda}\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{c}(e^{-i\lambda})]\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\\ e^{i(m-n)\lambda}\left[\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{c}(e^{i\lambda})\right]\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\end{array}\right\}\end{aligned}$$

Notice that since the information matrix is real, there will be cancellation between the complex parts of the above matrices.

Similarly,

$$\begin{aligned}&\frac{\partial vec'[\mathbf{G_{yy}}(\lambda)]}{\partial\boldsymbol{c}_m}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\right]\frac{\partial vec\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial\boldsymbol{\theta}'_x}\\ =&\left\{\begin{array}{c}\left[e^{-im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\otimes\mathbf{I}_N\right]\\ +\left[\mathbf{I}_N\otimes e^{im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{-i\lambda})\right]\end{array}\right\}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\right]\left[\mathbf{c}(e^{-i\lambda})\otimes\mathbf{c}(e^{i\lambda})\right]\frac{\partial G_{xx}(\lambda)}{\partial\boldsymbol{\theta}'_x}\\ =&\ G_{xx}(\lambda)\left\{\begin{array}{c}\left[e^{-im\lambda}\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{c}(e^{-i\lambda})\right]\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{c}(e^{i\lambda})\\ +\left[e^{im\lambda}\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{c}(e^{i\lambda})\right]\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{c}(e^{-i\lambda})\end{array}\right\}\frac{\partial G_{xx}(\lambda)}{\partial\boldsymbol{\theta}'_x},\end{aligned}$$

which again will be real.

In addition

$$\begin{aligned}&\frac{\partial vec'[\mathbf{G_{yy}}(\lambda)]}{\partial\boldsymbol{c}_m}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\right]\frac{\partial vec\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial\boldsymbol{\theta}'_{u_j}}\\ =&\left\{\begin{array}{c}\left[e^{-im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\otimes\mathbf{I}_N\right]\\ +\left[\mathbf{I}_N\otimes e^{im\lambda}G_{xx}(\lambda)\mathbf{c}'(e^{-i\lambda})\right]\end{array}\right\}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\right]\mathbf{E}_N\left(\mathbf{e}_j\frac{\partial G_{u_ju_j}(\lambda)}{\partial\boldsymbol{\theta}'_{u_j}}\right)\\ =&\ G_{xx}(\lambda)\left\{\begin{array}{c}\left[e^{-im\lambda}\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{e}_j\right]\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{e}_j\\ +\left[e^{im\lambda}\mathbf{c}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{e}_j\right]\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{e}_j\end{array}\right\}\frac{\partial G_{u_ju_j}(\lambda)}{\partial\boldsymbol{\theta}'_{u_j}}\end{aligned}$$

since

$$\mathbf{e}_j'\mathbf{E}_N' = \mathbf{e}_j'(\mathbf{e}_1\mathbf{e}_1'|\ldots|\mathbf{e}_N\mathbf{e}_N') = \mathbf{e}_j' \otimes \mathbf{e}_j'.$$

In turn,

$$\begin{aligned} & \frac{\partial vec'\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_x}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\otimes \mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\right]\frac{\partial vec\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_x'} \\ = & \frac{\partial G_{xx}(\lambda)}{\partial \boldsymbol{\theta}_x}\left[\mathbf{c}'(e^{i\lambda})\otimes \mathbf{c}'(e^{-i\lambda})\right]\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\otimes \mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\right]\left[\mathbf{c}(e^{-i\lambda})\otimes \mathbf{c}(e^{i\lambda})\right]\frac{\partial G_{xx}(\lambda)}{\partial \boldsymbol{\theta}_x'} \\ = & \left[\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{c}(e^{-i\lambda})\right]\left[\mathbf{c}'(e^{-i\lambda})\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\mathbf{c}(e^{i\lambda})\right]\frac{\partial G_{xx}(\lambda)}{\partial \boldsymbol{\theta}_x}\frac{\partial G_{xx}(\lambda)}{\partial \boldsymbol{\theta}_x'}. \end{aligned}$$

Further

$$\begin{aligned} & \frac{\partial vec'\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_x}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\otimes \mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\right]\frac{\partial vec\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_{u_i}'} \\ = & \frac{\partial G_{xx}(\lambda)}{\partial \boldsymbol{\theta}_x}\left[\mathbf{c}'(e^{i\lambda})\otimes \mathbf{c}'(e^{-i\lambda})\right]\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\otimes \mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\right]\mathbf{E}_N\mathbf{e}_j\frac{\partial G_{u_ju_j}(\lambda)}{\partial \boldsymbol{\theta}_{u_j}'} \\ = & \left[\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{e}_j\right]\left[\mathbf{c}'(e^{-i\lambda})\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\mathbf{e}_j\right]\frac{\partial G_{xx}(\lambda)}{\partial \boldsymbol{\theta}_x}\frac{\partial G_{u_ju_j}(\lambda)}{\partial \boldsymbol{\theta}_{u_j}'}. \end{aligned}$$

Finally,

$$\begin{aligned} & \frac{\partial vec'\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_{u_i}}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\otimes \mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\right]\frac{\partial vec\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_{u_j}'} \\ = & \frac{\partial G_{u_iu_i}(\lambda)}{\partial \boldsymbol{\theta}_{u_i}}\mathbf{e}_i'\mathbf{E}_N'\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\otimes \mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\right]\mathbf{E}_N\mathbf{e}_j\frac{\partial G_{u_ju_j}(\lambda)}{\partial \boldsymbol{\theta}_{u_j}'} \\ = & \mathbf{e}_i'\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\odot \mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\right]\mathbf{e}_j\frac{\partial G_{u_iu_i}(\lambda)}{\partial \boldsymbol{\theta}_{u_i}}\frac{\partial G_{u_ju_j}(\lambda)}{\partial \boldsymbol{\theta}_{u_j}'}, \end{aligned}$$

where $\odot$ denotes the Hadamard (or element by element) product of two matrices of equal size.

If we assume that both $G_{xx}(\lambda)$ and $\mathbf{G_{uu}}(\lambda)$ are strictly positive, we can use again the Woodbury formula to considerably simplify the previous expressions. In particular,

$$\begin{aligned} \mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{c}(e^{-i\lambda}) &= \mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}) - \omega(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}), \\ \mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\mathbf{c}(e^{i\lambda}) &= \mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{i\lambda}) - \omega(\lambda)\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{i\lambda})\mathbf{c}'(e^{-i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{i\lambda}), \end{aligned}$$

so that

$$\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{c}(e^{-i\lambda}) = \left[\mathbf{c}'(e^{-i\lambda})\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j)\mathbf{c}(e^{i\lambda})\right] = \mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda})G_{xx}^{-1}(\lambda)\omega(\lambda)$$

in view of (4). Finally, further speed gains can be achieved by noticing that

$$\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}) = \sum_{j=1}^{N}\frac{\left\|c_j(e^{i\lambda})\right\|^2}{G_{u_ju_j}(\lambda)}.$$

C State space representation in the time domain

There are several ways of casting the dynamic factor model in (1) into state-space format, but the most straightforward one is to consider a huge state vector of dimension $2N+3$ in which the ARMA(1,1) process for the common factor is written as a trivariate VAR(1) in (x_t, x_{t-1}, f_t) and the N ARMA(1,1) processes for the specific factors are written as first order bivariate VARs in (u_{it}, v_{it}). As a result, we can write the measurement equation without an error term as

$$\mathbf{y}_t = \mathbf{\Gamma}\mathbf{x}_t,$$

$$\mathbf{x}_t = (x_t, x_{t-1}, f_t; u_{1t}, v_{1t}; \ldots; u_{it}, v_{it}; \ldots; u_{Nt}, v_{Nt})'$$

and $\mathbf{\Gamma}$ is an $N \times (2N+3)$ matrix with typical row equal to

$$[c_{i0}, c_{i1}, 0; 0, 0; \ldots; 1, 0; \ldots; 0, 0].$$

In turn, the transition equation will be

$$\begin{bmatrix} x_t \\ x_{t-1} \\ f_t \end{bmatrix} = \begin{bmatrix} \alpha & 0 & -\beta \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ f_{t-1} \end{bmatrix} + \begin{bmatrix} f_t \\ 0 \\ f_t \end{bmatrix},$$

$$\begin{bmatrix} u_{it} \\ v_{it} \end{bmatrix} = \begin{bmatrix} \alpha & -\beta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{it-1} \\ v_{it-1} \end{bmatrix} + \begin{bmatrix} v_{it} \\ v_{it} \end{bmatrix} \quad (i = 1, \ldots, N),$$

with a block diagonal covariance matrix for its innovations.

Given our stationary assumption, the initial conditions for the state will trivially be $\mathbf{x}_{1|0} = \mathbf{0}_{(2N+3)x1}$ and

$$\mathbf{P}_{1|0} = \begin{bmatrix} \mathbf{Q}_x & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_1 & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{Q}_i & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{Q}_N \end{bmatrix},$$

in which the first 3×3 block is

$$\mathbf{Q}_x = \begin{bmatrix} \gamma_{x0} & \gamma_{x1} & 1 \\ \gamma_{x1} & \gamma_{x0} & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

$$\gamma_{x0} = \frac{1+\beta^2-2\alpha\beta}{1-\alpha^2}, \qquad \gamma_{x1} = \frac{(1-\alpha\beta)(\alpha-\beta)}{1-\alpha^2},$$

and the other N 2×2 blocks are

$$\mathbf{Q}_i = \begin{bmatrix} \gamma_{i0} & \psi_i \\ \psi_i & \psi_i \end{bmatrix},$$

$$\gamma_{i0} = \frac{1+\beta_i^2-2\alpha_i\beta_i}{1-\alpha_i^2}\psi_i.$$

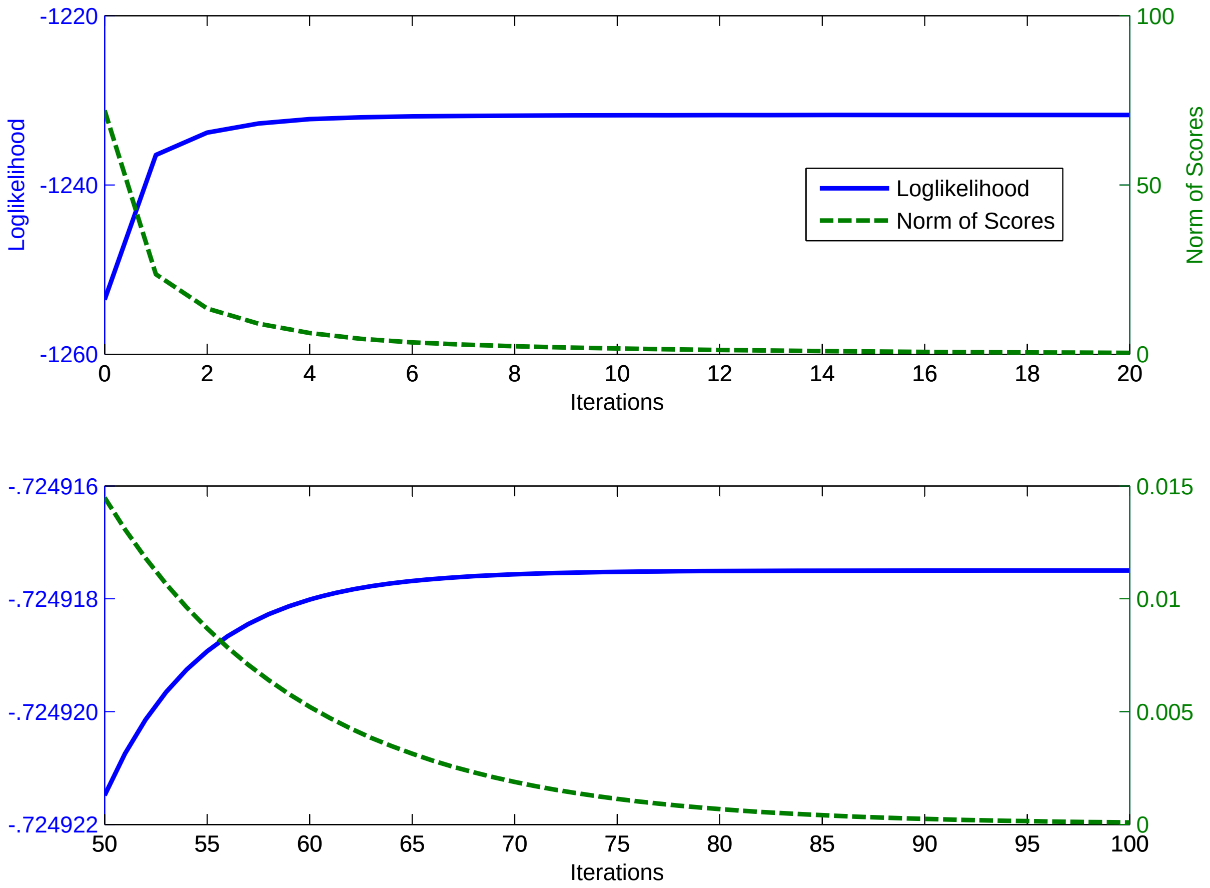

Figure 1: A model with $N = 10$ series

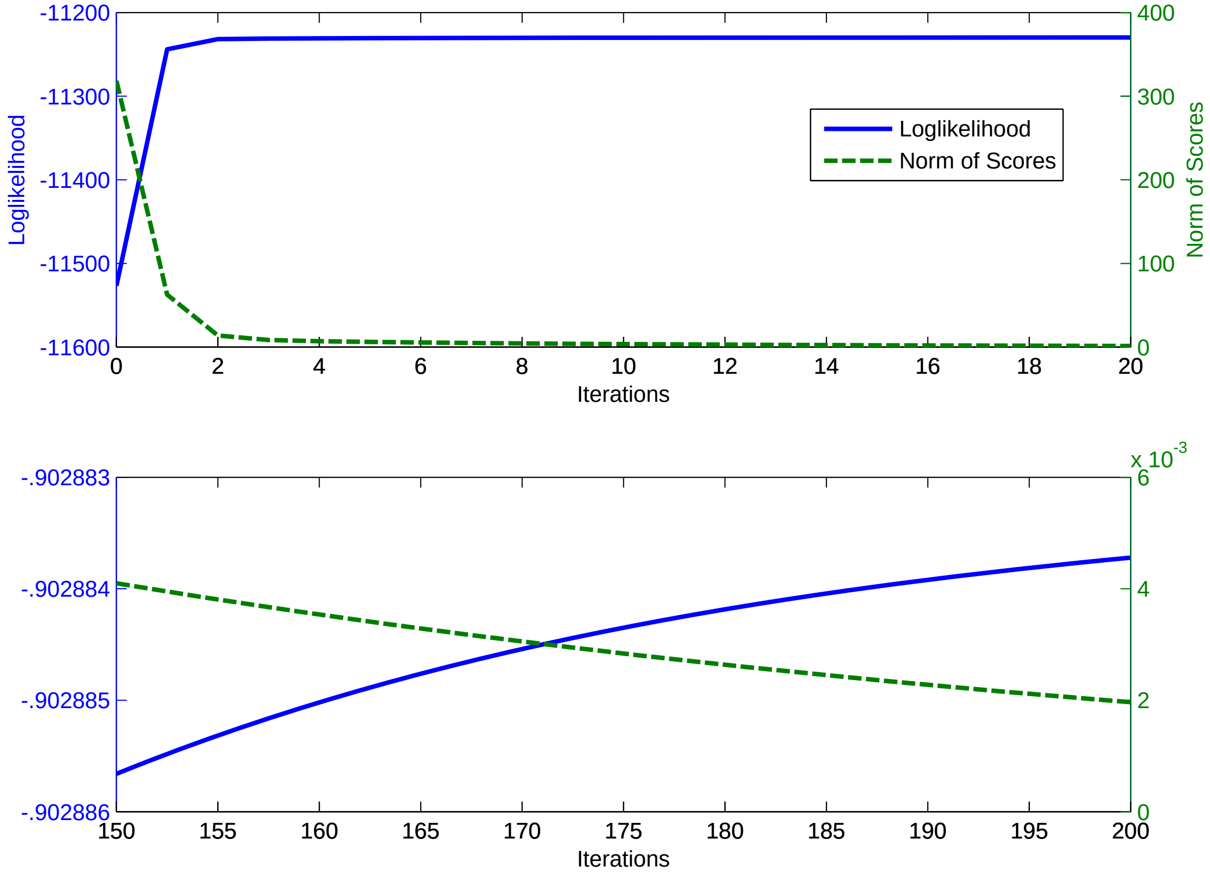

Figure 2: A model with $N = 100$ series

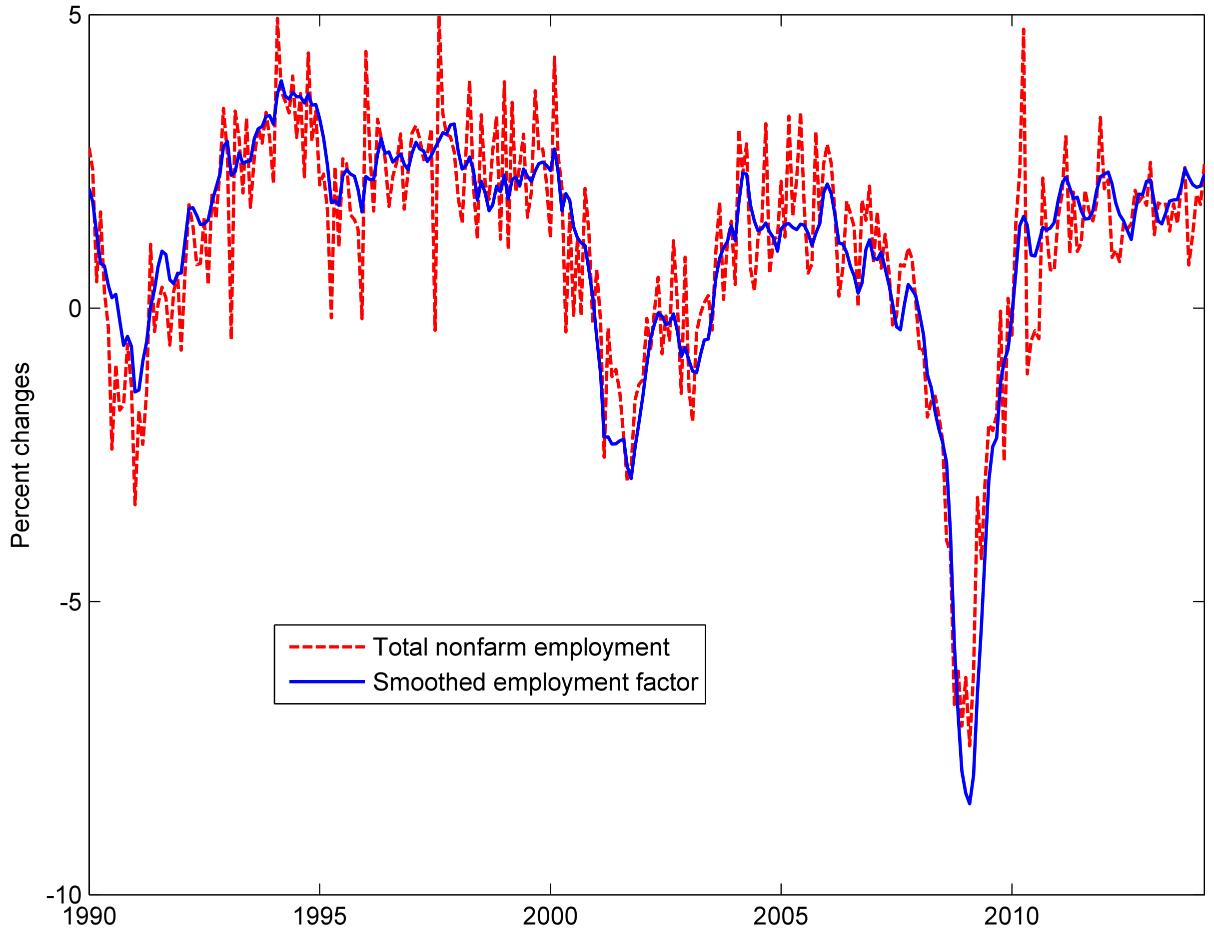

Figure 3: Total nonfarm employment and smoothed employment factor (annualised monthly frequency)

Table 1: Sample of NAICS 3-digits sectors for estimating the employment index

Oil and gas extraction (211)
Mining, except oil and gas (212)
Support activities for mining (213)
Construction of buildings (236)
Heavy and civil engineering construction (237)
Specialty trade contractors (238)
Wood products (321)
Nonmetallic mineral products (327)
Primary metals (331)
Fabricated metal products (332)
Machinery (333)
Computer and electronic products (334)
Electrical equipment and appliances (335)
Transportation equipment (336)
Furniture and related products (337)
Miscellaneous durable goods manufacturing (339)
Food manufacturing (311)
Textile mills (313)
Textile product mills (314)
Apparel (315)
Paper and paper products (322)
Printing and related support activities (323)
Petroleum and coal products (324)
Chemicals (325)
Plastics and rubber products (326)
Miscellaneous nondurable goods manufacturing (312,6)
Wholesale trade, durable goods (423)
Wholesale trade, nondurable goods (424)
Electronic markets and agents and brokers (425)
Motor vehicle and parts dealers (441)
Furniture and home furnishings stores (442)
Electronics and appliance stores (443)
Building material and garden supply stores (444)
Food and beverage stores (445)
Health and personal care stores (446)
Gasoline stations (447)
Clothing and clothing accessories stores (448)
Sporting goods, hobby, book, and music stores (451)
General merchandise stores (452)
Miscellaneous store retailers (453)
Nonstore retailers (454)
Air transportation (481)
Rail transportation (482)
Water transportation (483)
Truck transportation (484)
Transit and ground passenger transportation (485)
Pipeline transportation (486)
Scenic and sightseeing transportation (487)
Support activities for transportation (488)
Couriers and messengers (492)
Warehousing and storage (493)
Publishing industries, except Internet (511)
Motion picture and sound recording industries (512)
Broadcasting, except Internet (515)
Telecommunications (517)
Data processing, hosting and related services (518)
Other information services (519)
Monetary authorities - central bank (521)
Credit intermediation and related activities (522)
Securities, commodity contracts, investments, *etc.* (523,5)
Insurance carriers and related activities (524)
Real estate (531)
Rental and leasing services (532)
Lessors of nonfinancial intangible assets (533)
Administrative and support services (561)
Waste management and remediation services (562)
Ambulatory health care services (621)
Hospitals (622)
Nursing and residential care facilities (623)
Social assistance (624)
Performing arts and spectator sports (711)
Museums, historical sites, and similar institutions (712)
Amusements, gambling, and recreation (713)
Accommodation (721)
Food services and drinking places (722)
Repair and maintenance (811)
Personal and laundry services (812)
Membership associations and organizations (813)
Federal, except U.S. Postal Service
State government, excluding education
Local government, excluding education

Notes: NAICS 3-digit codes in parentheses.

Table 2: Dynamic loadings estimates

Series	$c_{i,0}$	std.err.	$c_{i,1}$	std.err.	Series	$c_{i,0}$	std.err.	$c_{i,1}$	std.err.
1	0.510	(0.458)	-0.183	(0.458)	42	0.956	(0.401)	-0.469	(0.400)
2	0.606	(0.542)	-0.201	(0.542)	43	0.026	(0.394)	0.275	(0.395)
3	0.233	(0.657)	1.180	(0.663)	44	0.049	(0.748)	0.178	(0.748)
4	1.757	(0.335)	-0.957	(0.331)	45	1.080	(0.235)	-0.522	(0.232)
5	2.004	(0.499)	-1.343	(0.494)	46	0.279	(0.608)	-0.136	(0.608)
6	2.195	(0.316)	-1.351	(0.308)	47	-0.520	(0.461)	0.444	(0.461)
7	2.457	(0.385)	-1.445	(0.377)	48	-0.341	(1.523)	0.836	(1.524)
8	2.031	(0.297)	-1.226	(0.288)	49	0.572	(0.299)	0.011	(0.300)
9	1.582	(0.295)	-0.135	(0.300)	50	0.312	(0.663)	0.241	(0.665)
10	1.060	(0.141)	0.038	(0.149)	51	0.877	(0.266)	-0.486	(0.264)
11	0.720	(0.183)	0.479	(0.195)	52	0.290	(0.135)	0.247	(0.140)
12	0.447	(0.169)	0.236	(0.172)	53	1.518	(0.908)	-1.227	(0.906)
13	0.741	(0.226)	0.166	(0.232)	54	-0.245	(0.216)	0.654	(0.221)
14	1.839	(0.483)	-1.042	(0.478)	55	0.136	(0.232)	0.163	(0.233)
15	2.068	(0.243)	-1.060	(0.234)	56	0.822	(0.339)	-0.617	(0.338)
16	0.625	(0.168)	-0.224	(0.167)	57	0.622	(0.430)	0.067	(0.430)
17	0.159	(0.200)	-0.069	(0.200)	58	-0.018	(0.480)	0.106	(0.480)
18	2.786	(0.448)	-1.992	(0.437)	59	0.030	(0.139)	0.010	(0.139)
19	2.037	(0.412)	-1.298	(0.406)	60	0.212	(0.220)	0.337	(0.223)
20	1.760	(0.462)	-1.068	(0.458)	61	0.029	(0.109)	0.040	(0.109)
21	0.677	(0.154)	-0.289	(0.153)	62	0.416	(0.193)	-0.174	(0.192)
22	0.391	(0.180)	0.212	(0.185)	63	0.456	(0.268)	0.212	(0.272)
23	0.329	(0.451)	-0.316	(0.451)	64	-0.610	(0.745)	0.847	(0.746)
24	0.168	(0.132)	0.067	(0.133)	65	2.734	(0.272)	-1.932	(0.255)
25	1.542	(0.236)	-0.694	(0.232)	66	0.261	(0.323)	-0.031	(0.323)
26	0.568	(0.358)	-0.319	(0.357)	67	0.260	(0.085)	-0.267	(0.085)
27	0.614	(0.098)	-0.169	(0.098)	68	0.073	(0.063)	-0.051	(0.063)
28	0.593	(0.125)	-0.367	(0.123)	69	-0.092	(0.090)	0.038	(0.090)
29	0.778	(0.156)	-0.502	(0.154)	70	-0.195	(0.301)	0.273	(0.301)
30	1.187	(0.136)	-0.819	(0.132)	71	-1.157	(0.970)	1.437	(0.971)
31	2.035	(0.261)	-1.352	(0.252)	72	0.093	(0.429)	0.174	(0.429)
32	1.547	(0.437)	-0.925	(0.434)	73	0.736	(0.496)	-0.374	(0.495)
33	1.668	(0.275)	-1.202	(0.269)	74	0.953	(0.262)	-0.544	(0.261)
34	0.119	(0.136)	0.031	(0.137)	75	0.683	(0.162)	-0.479	(0.160)
35	0.171	(0.194)	-0.006	(0.194)	76	0.967	(0.204)	-0.587	(0.202)
36	0.205	(0.187)	-0.016	(0.187)	77	0.520	(0.161)	-0.310	(0.160)
37	1.621	(0.365)	-1.326	(0.362)	78	0.182	(0.149)	-0.134	(0.149)
38	1.238	(0.565)	-0.821	(0.563)	79	-0.717	(1.330)	0.442	(1.329)
39	0.512	(0.318)	-0.326	(0.318)	80	-0.156	(0.113)	0.203	(0.113)
40	0.828	(0.277)	-0.303	(0.276)	81	-0.155	(0.137)	0.195	(0.137)
41	0.581	(0.420)	-0.128	(0.420)					

Table 3: ARMA parameter estimates

Series	α	std.err.	β	std.err.	ψ	std.err.	Series	α	std.err.	β	std.err.	ψ	std.err.
x	0.969	(0.015)	-0.448	(0.092)	1.000								
1	0.974	(0.017)	0.828	(0.044)	60.096	(4.986)	42	0.695	(0.092)	0.347	(0.120)	40.336	(3.358)
2	0.722	(0.130)	0.528	(0.159)	78.369	(6.502)	43	0.000		0.000		46.122	(3.826)
3	0.903	(0.034)	0.470	(0.069)	112.757	(9.371)	44	-0.299	(0.276)	-0.473	(0.255)	144.746	(12.001)
4	0.941	(0.031)	0.754	(0.060)	29.177	(2.465)	45	0.000		0.000		15.325	(1.288)
5	0.000		0.000		70.792	(5.926)	46	0.468	(0.088)	0.830	(0.055)	211.881	(17.568)
6	0.961	(0.024)	0.811	(0.052)	24.725	(2.125)	47	0.000		0.000		63.006	(5.227)
7	0.898	(0.062)	0.767	(0.091)	37.122	(3.168)	48	0.000		0.000		689.215	(57.149)
8	0.000		0.000		22.695	(1.942)	49	-0.424	(0.221)	-0.203	(0.239)	32.114	(2.673)
9	0.980	(0.014)	0.819	(0.043)	22.489	(1.908)	50	-0.043	(0.188)	0.273	(0.181)	183.986	(15.264)
10	0.983	(0.012)	0.797	(0.044)	4.482	(0.394)	51	0.000		0.000		20.391	(1.702)
11	0.978	(0.014)	0.761	(0.047)	8.523	(0.725)	52	0.935	(0.042)	0.811	(0.069)	5.039	(0.422)
12	0.913	(0.030)	0.432	(0.067)	7.197	(0.601)	53	0.087	(0.320)	0.266	(0.309)	291.773	(24.221)
13	0.981	(0.015)	0.886	(0.038)	14.162	(1.188)	54	0.000		0.000		13.672	(1.140)
14	0.025	(0.302)	0.220	(0.295)	80.586	(6.733)	55	0.968	(0.018)	0.715	(0.050)	14.950	(1.240)
15	0.931	(0.047)	0.819	(0.075)	13.359	(1.172)	56	0.954	(0.028)	0.811	(0.054)	32.269	(2.686)
16	0.921	(0.100)	0.877	(0.124)	7.852	(0.657)	57	0.934	(0.026)	0.509	(0.062)	48.497	(4.027)
17	0.653	(0.175)	0.778	(0.146)	13.743	(1.140)	58	0.000		0.000		68.358	(5.667)
18	0.000		0.000		53.176	(4.522)	59	0.933	(0.026)	0.518	(0.062)	5.091	(0.422)
19	0.000		0.000		47.095	(3.964)	60	0.944	(0.029)	0.740	(0.059)	13.408	(1.115)
20	0.975	(0.020)	0.883	(0.042)	59.655	(4.991)	61	0.900	(0.037)	0.558	(0.071)	3.167	(0.263)
21	0.969	(0.027)	0.898	(0.048)	6.569	(0.551)	62	0.000		0.000		10.871	(0.904)
22	0.000		0.000		9.236	(0.773)	63	0.000		0.000		20.938	(1.745)
23	0.000		0.000		60.368	(5.006)	64	0.924	(0.057)	0.832	(0.082)	158.608	(13.160)
24	0.937	(0.046)	0.841	(0.071)	4.965	(0.412)	65	0.000		0.000		15.862	(1.431)
25	0.903	(0.057)	0.763	(0.086)	13.737	(1.174)	66	0.000		0.000		30.885	(2.562)
26	0.000		0.000		37.754	(3.135)	67	0.974	(0.016)	0.791	(0.045)	1.986	(0.166)
27	0.937	(0.029)	0.673	(0.062)	2.344	(0.200)	68	0.935	(0.028)	0.619	(0.061)	1.084	(0.090)
28	0.000		0.000		4.314	(0.363)	69	0.961	(0.021)	0.735	(0.052)	2.248	(0.186)
29	0.966	(0.019)	0.758	(0.049)	6.406	(0.541)	70	0.000		0.000		26.847	(2.227)
30	0.908	(0.033)	0.485	(0.070)	3.934	(0.356)	71	0.000		0.000		278.380	(23.109)
31	0.970	(0.022)	0.869	(0.046)	16.520	(1.434)	72	0.032	(0.386)	0.183	(0.379)	63.386	(5.257)
32	0.000		0.000		54.725	(4.571)	73	0.000		0.000		72.657	(6.031)
33	0.870	(0.081)	0.743	(0.111)	19.082	(1.626)	74	0.055	(0.368)	-0.107	(0.366)	17.384	(1.454)
34	0.000		0.000		5.503	(0.457)	75	0.000		0.000		7.423	(0.622)
35	0.000		0.000		11.180	(0.928)	76	0.843	(0.078)	0.665	(0.109)	10.737	(0.904)
36	0.000		0.000		10.322	(0.857)	77	0.000		0.000		7.504	(0.626)
37	0.000		0.000		37.501	(3.148)	78	0.879	(0.048)	0.604	(0.080)	6.007	(0.499)
38	0.000		0.000		93.621	(7.783)	79	0.000		0.000		525.057	(43.536)
39	0.000		0.000		29.886	(2.481)	80	0.942	(0.035)	0.803	(0.061)	3.619	(0.301)
40	0.000		0.000		22.070	(1.841)	81	0.000		0.000		5.515	(0.458)
41	0.000		0.000		52.014	(4.318)							