Fast ML estimation of dynamic bifactor models: an application to European inflation*

Gabriele Fiorentini
Università di Firenze and RCEA, Viale Morgagni 59, I-50134 Firenze, Italy
< fiorentini@disia.unifi.it >

Alessandro Galesi
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
< galesi@cemfi.edu.es >

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
< sentana@cemfi.es >

February 2015

Abstract

We generalise the spectral EM algorithm for dynamic factor models in Fiorentini, Galesi and Sentana (2014) to bifactor models with pervasive global factors complemented by regional ones. We exploit the sparsity of the loading matrices so that researchers can estimate those models by maximum likelihood with many series from multiple regions. We also derive convenient expressions for the spectral scores and information matrix, which allows us to switch to the scoring algorithm near the optimum. We explore the ability of a model with a global factor and three regional ones to capture inflation dynamics across 25 European countries over 1999-2014.

Keywords: Euro area, Inflation convergence, Spectral maximum likelihood, Wiener-Kolmogorov filter.
JEL: C32, C38, E37, F45

*We are grateful to Ángel Estrada and Albert Satorra, as well as to audiences at the Advances in Econometrics Conference on Dynamic Factor Models (Aarhus, 2014), the EC^2 Advances in Forecasting Conference (UPF, 2014) and the Italian Congress of Econometrics and Empirical Economics (Salerno, 2015) for helpful comments and suggestions. Financial support from MIUR through the project "Multivariate statistical models for risk assessment" (Fiorentini) is gratefully acknowledged.

1 Introduction

The dynamic factor models introduced by Geweke (1977) and Sargent and Sims (1977) constitute a flexible tool for capturing the cross-sectional and dynamic correlations between multiple series in a parsimonious way. Although single factor versions of those models prevail because their ease of interpretation and the fact that they provide a reasonable first approximation to many data sets, there is often the need to add more common factors to adequately capture the off-diagonal elements of the autocovariance matrices. When the cross-sectional dimension, N, is commensurate with the time series dimension, T, one popular solution is to rely on the approximate factor models structures originally introduced by Chamberlain and Rothschild (1983) in the static case, which allow for some mild contemporaneous and dynamic correlation between idiosyncratic terms (see e.g. Bai and Ng (2008) and the references therein). Unfortunately, the cross-sectional asymptotic boundedness conditions on the eigenvalues of the autocovariance matrices of the idiosyncratic terms underlying those approximate factor models are largely meaningless in empirical situations in which N is small relative to T. In those situations in which it is natural to group the N series into R homogeneous blocks, an attractive solution are bifactor models with two types of factors:

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

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 may be reluctant to use ML because of the heavy computational burden involved, which is disproportionately larger as the number of series considered increases.

In the context of standard dynamic factor models, 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 it can be successfully applied.

In a recent companion paper (Fiorentini, Galesi and Sentana (2014)), we introduced a frequency domain version of the EM algorithm for general dynamic factor models with latent Arma processes. We showed there that 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. The purpose of the current paper is to extend our methods to dynamic versions of bifactor models.

We illustrate our procedure with an empirical application in which we study the dynamics of European inflation rates since the creation of the European Monetary Union (EMU). Specifically, we consider a dynamic bifactor model with a single global factor and three regional factors representing core, new entrant and outside EMU countries.

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 and our conclusions in section 5. Auxiliary results are gathered in appendices.

2 Theoretical background

2.1 Dynamic bifactor models

Let $\mathbf{y}_t$ denote a finite dimensional vector of N observed series, which can be grouped into R different categories or blocks as follows

$$\mathbf{y}'_t = \begin{pmatrix} \mathbf{y}'_{1t} & \dots & \mathbf{y}'_{rt} & \dots & \mathbf{y}'_{Rt} \end{pmatrix},$$

where $\mathbf{y}_{1t}$ is of dimension N_1, $\mathbf{y}_{rt}$ of dimension N_r and $\mathbf{y}_{Rt}$ is of dimension N_R, with $N_1 + \ldots + N_r + \ldots + N_R = N$. Henceforth we shall refer to each category as a "region", even though they could represent alternative groupings.

To keep the notation to a minimum, we focus on models with a single global factor and a single factor per region, which suffice to illustrate our procedures. Specifically, we assume that

$\mathbf{y}_t$ can be defined in the time domain by the system of dynamic stochastic difference equations

$$\left.\begin{array}{c}\mathbf{y}_{rt} = \boldsymbol{\mu}_r + \mathbf{c}_{rg}(L)x_{gt} + \mathbf{c}_{rr}(L)x_{rt} + \mathbf{u}_{rt}, \quad r = 1, \ldots, R \\ \alpha_{x_g}(L)x_{gt} = \beta_{x_g}(L)f_{gt}, \\ \alpha_{x_r}(L)x_{rt} = \beta_{x_r}(L)f_{rt}, \quad r = 1, \ldots, R \\ \alpha_{u_i}(L)u_{i,t} = \beta_{u_i}(L)v_{i,t}, \quad i = 1, \ldots, N, \\ (f_{gt}, f_{1t}, \ldots, f_{Rt}, v_{1t}, \ldots, v_{Nt})|I_{t-1}; \boldsymbol{\mu}, \boldsymbol{\theta} \sim N[0, diag(1, 1, , \ldots, 1, \psi_1, \ldots, \psi_N)], \end{array}\right\} \tag{1}$$

where x_{gt} is the global factor, x_{rt} $(r = 1, \ldots, R)$ the r^{th} regional factor, $\mathbf{u}_t = (\mathbf{u}'_{1t}, \ldots, \mathbf{u}'_{rt}, \ldots, \mathbf{u}'_{Rt})'$ the N specific factors,

$$\mathbf{c}_{rg}(L) = \sum_{k=-m_g}^{n_g} \mathbf{c}_{rgk}L^k \tag{2}$$

$$\mathbf{c}_{rr}(L) = \sum_{l=-m_r}^{n_r} \mathbf{c}_{rrl}L^k \tag{3}$$

for $(r = 1, \ldots, R)$ are $N_R \times 1$ vectors of possibly two-sided polynomials in the lag operator $c_{ig}(L)$ and $c_{ir}(L)$, $\alpha_{x_g}(L)$, $\alpha_{x_r}(L)$ and $\alpha_{u_i}(L)$ are one-sided polynomials of orders p_{x_g}, p_{x_r} and p_{u_i}, respectively, while $\beta_{x_g}(L)$, $\beta_{x_r}(L)$ and $\beta_{u_i}(L)$ are one-sided polynomials of orders q_{x_g}, q_{x_r} and q_{u_i}, coprime with $\alpha_{x_g}(L)$, $\alpha_{x_r}(L)$ and $\alpha_{u_i}(L)$, respectively, I_{t-1} is an information set that contains the values of $\mathbf{y}_t$ and $\mathbf{f}_t = (f_{gt}, f_{1t}, \ldots, f_{Rt})'$ 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 for a series y_{it} in region r would be

$$\left.\begin{array}{c} y_{it} = \mu_i + c_{i0g}x_{gt} + c_{i1g}x_{gt-1} + c_{i0r}x_{rt} + c_{i1r}x_{rt-1} + u_{it} \\ x_{gt} = \alpha_{1x_g}x_{gt-1} + f_{gt} \\ x_{rt} = \alpha_{1x_r}x_{rt-1} + \alpha_{2x_r}x_{rt-2} + f_{rt} \\ u_{it} = \alpha_{1u_i}u_{it-1} + v_{it} \end{array}\right\}. \tag{4}$$

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

1. The serial correlation of the global and regional factors $\mathbf{x}'_t = (x_{gt}, x_{1t}, \ldots, x_{Rt})$
2. The serial correlation of the idiosyncratic factors $\mathbf{u}_t$
3. The heterogeneous dynamic impact of the global and regional factors on each of the observed variables through the country-specific distributed lag polynomials $c_{ig}(L)$ and $c_{ir}(L)$.

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 dynamics for the global and regional factors can be regarded as a parsimonious way of modelling infinite distributed lags.

The main difference with respect to the standard dynamic factor models considered in Fiorentini, Galesi and Sentana (2014) is the presence of regional factors, which allow for richer covariance relationships between series that belong to the same region (see e.g. Stock and Watson

(2009)).[1] As we shall see below, though, the covariance between series in different regions depends exclusively on the pervasive common factor.

Model (1) differs from the dynamic hierarchical factor model considered by Moench, Ng and Potter (2013) in an important aspect. In their model, the common factor affects the observed series only through its effect on the regional factor. As a result, the autocovariance matrices of each block have a single factor structure and the dynamic impact of the common factor in the observed variables must involve longer distributed lags than the dynamic impact of the regional factor. As usual, the increase in parsimony involves a reduction in flexibility.

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

$$\mathbf{G_{yy}}(\lambda)=\begin{bmatrix} \mathbf{G}_{\mathbf{y}_1\mathbf{y}_1}(\lambda) & \dots & \mathbf{G}_{\mathbf{y}_1\mathbf{y}_r}(\lambda) & \dots & \mathbf{G}_{\mathbf{y}_1\mathbf{y}_R}(\lambda) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{G}_{\mathbf{y}_r\mathbf{y}_1}(\lambda) & \dots & \mathbf{G}_{\mathbf{y}_r\mathbf{y}_r}(\lambda) & \dots & \mathbf{G}_{\mathbf{y}_r\mathbf{y}_R}(\lambda) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{G}_{\mathbf{y}_R\mathbf{y}_1}(\lambda) & \dots & \mathbf{G}_{\mathbf{y}_R\mathbf{y}_r}(\lambda) & \dots & \mathbf{G}_{\mathbf{y}_R\mathbf{y}_R}(\lambda) \end{bmatrix}=\mathbf{C}(e^{-i\lambda})\mathbf{G}_{xx}(\lambda)\mathbf{C}'(e^{i\lambda})+\mathbf{G_{uu}}(\lambda),$$

where

$$\mathbf{C}(z)=\begin{bmatrix} \mathbf{c}_{1g}(z) & \mathbf{c}_{11}(z) & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{c}_{rg}(z) & \mathbf{0} & \dots & \mathbf{c}_{rr}(z) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{c}_{Rg}(z) & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{c}_{RR}(z) \end{bmatrix}=\begin{bmatrix} \mathbf{c}_g(z) & \mathbf{C}_r(z) \end{bmatrix}, \tag{5}$$

$$\begin{aligned}
\mathbf{G_{xx}}(\lambda) &= diag[G_{x_gx_g}(\lambda), G_{x_1x_1}(\lambda), \dots, G_{xrx_r}(\lambda)\dots, G_{x_Rx_R}(\lambda)], \\
G_{x_gx_g}(\lambda) &= \frac{\beta_{x_g}(e^{-i\lambda})\beta_{x_g}(e^{i\lambda})}{\alpha_{x_g}(e^{-i\lambda})\alpha_{x_g}(e^{i\lambda})}, \quad G_{x_rx_r}(\lambda)=\frac{\beta_{x_r}(e^{-i\lambda})\beta_{x_r}(e^{i\lambda})}{\alpha_{xr}(e^{-i\lambda})\alpha_{x_r}(e^{i\lambda})},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{G_{uu}}(\lambda) &= diag[G_{u_1u_1}(\lambda), \dots, G_{u_Nu_N}(\lambda)], \\
G_{u_iu_i}(\lambda) &= \psi_i\frac{\beta_{u_i}(e^{-i\lambda})\beta_{u_i}(e^{i\lambda})}{\alpha_{u_i}(e^{-i\lambda})\alpha_{u_i}(e^{i\lambda})}.
\end{aligned}$$

Thus, the matrix $\mathbf{G_{yy}}(\lambda)$ inherits the restricted $(R+1)$-factor structure of the unconditional covariance matrix of a static bifactor model with a common global factor and an additional

[1] Static versions of bifactor models have a long tradition in psychometrics after their introduction by Holzinger and Swineford (1937) as an important special case of confirmatory factor analysis (see Reise (2012) for an up to date list of references).

factor per region. As a result, the cross-covariances between two series within one region will depend on the influence of both the global and regional factors on each of the series since

$$\mathbf{G}_{\mathbf{y}_r\mathbf{y}_r}(\lambda) = \mathbf{c}_{rg}(e^{-i\lambda})G_{x_gx_g}(\lambda)\mathbf{c}'_{rg}(e^{i\lambda}) + \mathbf{c}_{rr}(e^{-i\lambda})G_{x_rx_r}(\lambda)\mathbf{c}_{rr}(e^{i\lambda}) + \mathbf{G}_{\mathbf{u}_r\mathbf{u}_r}(\lambda).$$

In contrast, the cross-covariances between two series that belong to different regions will only depend on their dynamic sensitivities to the common factor because

$$\mathbf{G}_{\mathbf{y}_r\mathbf{y}_k}(\lambda) = \mathbf{c}_{rg}(e^{-i\lambda})G_{x_gx_g}(\lambda)\mathbf{c}'_{r'g}(e^{i\lambda}), \quad r \neq r'.$$

We can easily ensure the separate identification of the common and idiosyncratic components of $\mathbf{G}_{\mathbf{y}_r\mathbf{y}_r}(\lambda)$ when $\mathbf{G}_{\mathbf{u}_r\mathbf{u}_r}(\lambda)$ has full rank provided N_r is sufficiently large. The separate identification of $\mathbf{c}_{rg}(e^{-i\lambda})$, $\mathbf{c}_{rr}(e^{-i\lambda})$, $G_{x_gx_g}(\lambda)$ and $G_{x_rx_r}(\lambda)$ is trickier, but it can be guaranteed (up to scale and time shifts) as long as R is sufficiently large, the polynomials $c_{ir}(.)$ do not share a common root within block r, and the polynomials c_{ig} do not share a common root across all N countries (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 (4),

$$\begin{aligned} G_{x_gx_g}(\lambda) &= \frac{1}{\alpha_{x_g}(e^{-i\lambda})\alpha_{x_g}(e^{i\lambda})} = \frac{1}{1+\alpha_{1x_g}^2 - 2\alpha_{1x_g}\cos\lambda}, \\ G_{x_rx_r}(\lambda) &= \frac{1}{\alpha_{x_r}(e^{-i\lambda})\alpha_{x_r}(e^{i\lambda})} = \frac{1}{1+\alpha_{1x_r}^2 + \alpha_{2x_r}^2 - 2\alpha_{1x_r}(1-\alpha_{2x_r})\cos\lambda - 2\alpha_{2x_r}\cos 2\lambda}, \end{aligned}$$

where we have exploited the fact that the variances of f_{gt} and f_{rt} can be normalised to 1 for identification purposes.[2]

Similarly,

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

Finally,

$$\begin{aligned} c_{ig}(e^{-i\lambda}) &= c_{ig0} + c_{ig1}e^{-i\lambda}, \\ c_{ir}(e^{-i\lambda}) &= c_{ir0} + c_{ir1}e^{-i\lambda}. \end{aligned}$$

The fact that the idiosyncratic impact of the common factors on each of the observed variables is in principle dynamic implies that the spectral density matrix of $\mathbf{y}_t$ will generally be complex

[2] Other symmetric scaling assumptions would normalise the unconditional variance of x_{gt} and x_{rt} $(r = 1, \ldots, R)$, or some norm of the vectors of impact multipliers $\mathbf{c}_{g0} = (\mathbf{c}'_{1g0}, \ldots, \mathbf{c}'_{Rg0})$ and $\mathbf{c}_{rr0}$ $(r = 1, \ldots, R)$ or their long run counterparts $\mathbf{c}_g(1)$ and $\mathbf{c}_{rr}(1)$. Alternatively, we could asymmetrically fix one element of $\mathbf{c}_{g0}$ and $\mathbf{c}_{rr0}$ (or $\mathbf{c}_g(1)$ and $\mathbf{c}_{rr}(1)$) $(r = 1, \ldots, R)$ to 1.

but Hermitian, even though the spectral densities of x_{gt}, x_{rt} 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

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

denote the spectral decomposition of the observed vector process.

Assuming that $\mathbf{G}_{\mathbf{yy}}(\lambda)$ is not singular at any frequency, the Wiener-Kolmogorov two-sided filter for the $(R+1)$ "common" factors $\mathbf{x}_t$ at each frequency is given by

$$d\mathbf{Z}^{\mathbf{x}^K}(\lambda) = \mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda) d\mathbf{Z}^{\mathbf{y}}(\lambda), \tag{6}$$

where

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

is known as the transfer function of the common factors' smoother. As a result, the spectral density of the smoothed values of the common factors, $\mathbf{x}_{t|\infty}^K$, is

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

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

$$\mathbf{G}_{\mathbf{xx}}(\lambda) - \mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda) = \boldsymbol{\Omega}(\lambda).$$

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

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

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

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

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

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

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

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

The advantage of this expression is that $\mathbf{G}_{\mathbf{uu}}(\lambda)$ is a diagonal matrix and $\mathbf{\Omega}(\lambda)$ of dimension $(R+1)$, much smaller than N, which greatly simplifies the computations.

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

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

so

$$\begin{aligned}
\mathbf{G}_{\mathbf{x}^K\mathbf{x}^K}(\lambda) &= \mathbf{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda) = \mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{\Omega}(\lambda) \\
&= \mathbf{G}_{\mathbf{xx}}(\lambda)\left\{\mathbf{G}_{\mathbf{xx}}(\lambda) + [\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{-i\lambda})]^{-1}\right\}^{-1}\mathbf{G}_{\mathbf{xx}}(\lambda) = \mathbf{G}_{\mathbf{xx}}(\lambda) - \mathbf{\Omega}(\lambda), \qquad (7)
\end{aligned}$$

where we have used the fact that

$$\mathbf{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{-i\lambda}) = \mathbf{I}_{R+1} - \mathbf{\Omega}(\lambda)\mathbf{G}^{-1}_{\mathbf{xx}}(\lambda), \qquad (8)$$

which can be easily proved by premultiplying both sides by $\mathbf{\Omega}^{-1}(\lambda)$.

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

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

so

$$\mathbf{G}_{\mathbf{u}^K\mathbf{u}^K}(\lambda) = \mathbf{G}_{\mathbf{uu}}(\lambda) - \mathbf{C}(e^{-i\lambda})\mathbf{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda}). \qquad (9)$$

Finally,

$$\mathbf{G}_{\mathbf{x}^K\mathbf{u}^K}(\lambda) = \mathbf{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda}). \qquad (10)$$

In addition, we can exploit the special structure of the matrix $\mathbf{C}(z)$ in (5) to further speed up the calculations. Specifically, tedious algebraic manipulations show that the $(R+1)\times(R+1)$

Hermitian matrix $\boldsymbol{\Omega}^{-1}(\lambda) = \mathbf{G}^{-1}_{\mathbf{xx}}(\lambda) + \mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{-i\lambda})$ can be easily computed as

$$\begin{bmatrix} \omega^{gg}(\lambda) & \omega^{g1}(\lambda) & \cdots & \omega^{gr}(\lambda) & \cdots & \omega^{gR}(\lambda) \\ \omega^{1g}(\lambda) & \omega^{11}(\lambda) & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \omega^{rg}(\lambda) & 0 & \cdots & \omega^{rr}(\lambda) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \omega^{Rg}(\lambda) & 0 & \cdots & 0 & \cdots & \omega^{RR}(\lambda) \end{bmatrix} \tag{11}$$

with

$$\begin{aligned} \omega^{gg}(\lambda) &= G^{-1}_{x_g x_g}(\lambda) + \mathbf{c}'_{rg}(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{c}_{rg}(e^{-i\lambda}), \\ \omega^{rr}(\lambda) &= G^{-1}_{x_r x_r}(\lambda) + \mathbf{c}'_{rr}(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{u}_r\mathbf{u}_r}(\lambda)\mathbf{c}_{rr}(e^{-i\lambda}) \end{aligned}$$

and

$$\omega^{rg}(\lambda) = \mathbf{c}'_{rr}(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{u}_r\mathbf{u}_r}(\lambda)\mathbf{c}_{rg}(e^{-i\lambda}) = \omega^{gr*}(\lambda),$$

where * denotes the complex conjugate transpose.

Interestingly, we can write (11) as

$$\mathbf{A}(\lambda) + \mathbf{B}(\lambda)\mathbf{D}^*(\lambda),$$

where

$$\begin{aligned} \mathbf{A}(\lambda) &= diag\left[\omega^{gg}(\lambda), \omega^{11}(\lambda), \ldots, \omega^{rr}(\lambda), \ldots, \omega^{RR}(\lambda)\right] \\ \mathbf{B}(\lambda) &= \begin{bmatrix} 1 & 0 \\ 0 & \omega^{1g}(\lambda) \\ \vdots & \vdots \\ 0 & \omega^{rg}(\lambda) \\ \vdots & \vdots \\ 0 & \omega^{Rg}(\lambda) \end{bmatrix} \end{aligned}$$

and

$$\mathbf{D}^*(\lambda) = \begin{bmatrix} 0 & \omega^{g1}(\lambda) & \cdots & \omega^{gr}(\lambda) & \cdots & \omega^{gR}(\lambda) \\ 1 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$

are two rank 2 matrices.

The advantage of this formulation is that the Woodbury formula for complex matrices implies that

$$\boldsymbol{\Omega}(\lambda) = [\mathbf{A}(\lambda) + \mathbf{B}(\lambda)\mathbf{D}^*(\lambda)]^{-1} = \mathbf{A}^{-1}(\lambda) - \mathbf{A}^{-1}(\lambda)\mathbf{B}(\lambda)\mathbf{F}^{-1}(\lambda)\mathbf{D}^*(\lambda)\mathbf{A}^{-1}(\lambda),$$

where

$$\mathbf{F}(\lambda) = \mathbf{I}_2 + \mathbf{D}^*(\lambda)\mathbf{A}^{-1}(\lambda)\mathbf{B}(\lambda) = \begin{bmatrix} 1 & \omega_{+g}(\lambda) \\ \frac{1}{\omega^{gg}(\lambda)} & 1 \end{bmatrix},$$

with

$$\omega_{+g}(\lambda) = \sum_{r=1}^{R} \frac{\|\omega^{rg}(\lambda)\|^2}{\omega^{rr}(\lambda)}$$

where we have exploited the fact that $\omega^{rg}(\lambda)$ and $\omega^{gr}(\lambda)$ are complex conjugates so that the matrix $\mathbf{F}(\lambda)$ is actually real.

If we put all the pieces together we will end up with

$$\boldsymbol{\Omega}(\lambda) = \begin{bmatrix} \omega_{gg}(\lambda) & \omega_{g1}(\lambda) & \cdots & \omega_{gr}(\lambda) & \cdots & \omega_{gR}(\lambda) \\ \omega_{1g}(\lambda) & \omega_{11}(\lambda) & \cdots & \omega_{1r}(\lambda) & \cdots & \omega_{1R}(\lambda) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \omega_{rg}(\lambda) & \omega_{r1}(\lambda) & \cdots & \omega_{rr}(\lambda) & \cdots & \omega_{rR}(\lambda) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \omega_{Rg}(\lambda) & \omega_{R1}(\lambda) & \cdots & \omega_{Rr}(\lambda) & \cdots & \omega_{RR}(\lambda) \end{bmatrix} = \begin{bmatrix} \omega_{gg}(\lambda) & \boldsymbol{\omega}^*_{rg}(\lambda) \\ \boldsymbol{\omega}_{rg}(\lambda) & \boldsymbol{\Omega}_{rr}(\lambda) \end{bmatrix} \tag{12}$$

where

$$\begin{aligned} \omega_{gg}(\lambda) &= \frac{1}{\omega^{gg}(\lambda)} + \frac{1}{\omega^{gg}(\lambda)} \frac{\omega_{+g}(\lambda)}{\omega^{gg}(\lambda) - \omega_{+g}(\lambda)} = \frac{1}{\omega^{gg}(\lambda) - \omega_{+g}(\lambda)} \\ \omega_{rr}(\lambda) &= \frac{1}{\omega^{rr}(\lambda)} \left(1 + \frac{\|\omega^{rg}(\lambda)\|^2}{\omega^{rr}(\lambda)} \omega_{gg}(\lambda)\right) \\ \omega_{rg}(\lambda) &= -\frac{\omega^{rg}(\lambda)}{\omega^{rr}(\lambda)} \omega_{gg}(\lambda) = \omega^*_{rg}(\lambda) \end{aligned}$$

and

$$\omega_{rk}(\lambda) = \frac{\omega^{rg}(\lambda)\omega^{gk}(\lambda)}{\omega^{rr}(\lambda)\omega^{kk}(\lambda)} \omega_{gg}(\lambda) = \omega^*_{kr}(\lambda).$$

It is of some interest to compare these expressions to the corresponding expressions in the case of a model with a single global factor but no regional factors and a model with regional factors but no global factor.

In the first case, we would have

$$\omega(\lambda) = \frac{1}{\omega^{gg}(\lambda)}$$

while in the second case

$$\omega_{rr}(\lambda) = \frac{1}{\omega^{rr}(\lambda)}.$$

As expected, the existence of regional factors makes more difficult the estimation of the common factor and vice versa.

The Woodbury formula also implies that

$$|\boldsymbol{\Omega}(\lambda)| = |\mathbf{A}(\lambda)| \, |\mathbf{F}(\lambda)| \, ,$$

with

$$|\mathbf{F}(\lambda)| = 1 - \frac{\omega_{+g}(\lambda)}{\omega^{gg}(\lambda)}.$$

The bifactor structure can also be used to speed up the filtering procedure. Specifically,

$$\mathbf{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda}) = \begin{bmatrix} \omega_{gg}(\lambda) & \boldsymbol{\omega}^*_{rg}(\lambda) \\ \boldsymbol{\omega}_{rg}(\lambda) & \mathbf{\Omega}_{rr}(\lambda) \end{bmatrix} \begin{bmatrix} \mathbf{c}'_{rg}(e^{i\lambda}) \\ \mathbf{C}'_r(e^{i\lambda}) \end{bmatrix} = \begin{bmatrix} \omega_{gg}(\lambda)\mathbf{c}'_{rg}(e^{i\lambda}) + \boldsymbol{\omega}^*_{rg}(\lambda)\mathbf{C}'_r(e^{i\lambda}) \\ \boldsymbol{\omega}_{rg}(\lambda)\mathbf{c}'_{rg}(e^{i\lambda}) + \mathbf{\Omega}_{rr}(\lambda)\mathbf{C}'_r(e^{i\lambda}) \end{bmatrix}$$

and

$$\begin{aligned} \mathbf{C}(e^{-i\lambda})\mathbf{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda}) &= \mathbf{c}_{rg}(e^{i\lambda})\omega_{gg}(\lambda)\mathbf{c}'_{rg}(e^{i\lambda}) + \mathbf{C}_r(e^{-i\lambda})\mathbf{\Omega}_{rr}(\lambda)\mathbf{C}'_r(e^{i\lambda}) \\ &\quad + \mathbf{c}_{rg}(e^{-i\lambda})\boldsymbol{\omega}^*_{rg}(\lambda)\mathbf{C}'_r(e^{i\lambda}) + \mathbf{C}_r(e^{-i\lambda})\boldsymbol{\omega}_{rg}(\lambda)\mathbf{c}'_{rg}(e^{i\lambda}), \end{aligned}$$

which can be computed rather quickly by exploiting the block diagonal nature of $\mathbf{C}_r(z)$ in (5).

2.4 The minimal sufficient statistics for $\{\mathbf{x}_t\}$

Define $\mathbf{x}^G_{t|\infty}$ as the spectral GLS estimator of $\mathbf{x}_t$ through the transformation

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

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

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

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

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

and the (2,2) block

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

whose rank is $N-(R+1)$.

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 $\mathbf{x}^G_{t|\infty}$, which is of dimension $(R+1)$, and the log-likelihood function of $\mathbf{u}^G_{t|\infty}$. Importantly, the parameters characterising $\mathbf{G}_{\mathbf{xx}}(\lambda)$ only enter through the first component. In contrast, the remaining parameters affect both components. Moreover, we can easily show that

1. $\mathbf{x}^G_{t|\infty} = \mathbf{x}_t + \boldsymbol{\zeta}^G_{t|\infty}$, with $\mathbf{x}_t$ and $\boldsymbol{\zeta}^G_{t|\infty}$ orthogonal at all leads and lags.

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

This confirms that $\mathbf{x}^G_{t|\infty}$ constitute minimal sufficient statistics for $\mathbf{x}_t$, 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 Gouriéroux, Monfort and Renault (1991)). In addition, the degree of unobservability of $\mathbf{x}_t$ 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 $\mathbf{G}_{\mathbf{xx}}(\lambda)$ (see Sentana (2004) for a closely related discussion).

2.5 Maximum likelihood estimation in the frequency domain

Let

$$\mathbf{I}_{\mathbf{yy}}(\lambda) = \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_s - \boldsymbol{\mu})' e^{-i(t-s)\lambda} \tag{13}$$

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 singular at any of those frequencies, the so-called Whittle (discrete) spectral approximation to the log-likelihood function is[3]

$$N\varkappa - \frac{1}{2}\sum_{j=0}^{T-1}\ln|\mathbf{G}_{\mathbf{yy}}(\lambda_j)| - \frac{1}{2}\sum_{j=0}^{T-1}\operatorname{tr}\left\{\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)[2\pi\mathbf{I}_{\mathbf{yy}}(\lambda_j)]\right\}, \tag{14}$$

with $\varkappa = -(T/2)\ln(2\pi)$ (see e.g. Hannan (1973) and Dunsmuir and Hannan (1976)).

Expression (13), 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 (14) becomes

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

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

But since $\mathbf{z}^{\mathbf{y}}_j$ 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 $\mathbf{z}^{\mathbf{y}}_0 = (\bar{\mathbf{y}}_T - \boldsymbol{\mu})$, where $\bar{\mathbf{y}}_T$ is the sample mean of $\mathbf{y}_t$, it immediately follows that the ML of $\boldsymbol{\mu}$ will be $\bar{\mathbf{y}}_T$, so in what follows we focus on demeaned

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

variables. As for the remaining parameters, the score function will be given by:

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

$$\begin{aligned}\mathbf{d}(\lambda_j;\boldsymbol{\theta}) &= \frac{1}{2}\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda_j)\right]}{\partial\boldsymbol{\theta}}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\right] vec\left[2\pi\mathbf{z}^{\mathbf{y}c}_j\mathbf{z}^{\mathbf{y}\prime}_j - \mathbf{G}'_{\mathbf{yy}}(\lambda_j)\right] \\ &= \frac{1}{2}\frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda_j)]}{\partial\boldsymbol{\theta}}\mathbf{M}(\lambda_j)\mathbf{m}(\lambda_j), \end{aligned} \tag{15}$$

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

$$\mathbf{m}(\lambda_j) = vec\left[2\pi\mathbf{z}^{\mathbf{y}c}_j\mathbf{z}^{\mathbf{y}\prime}_j - \mathbf{G}'_{\mathbf{yy}}(\lambda_j)\right] \tag{16}$$

and

$$\mathbf{M}(\lambda_j) = \mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j). \tag{17}$$

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

$$\mathbf{Q}(\boldsymbol{\theta}) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathbf{Q}(\lambda;\boldsymbol{\theta})d\lambda = \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{\partial\boldsymbol{\theta}}\mathbf{M}(\lambda)\left\{\frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{\partial\boldsymbol{\theta}}\right\}^* d\lambda, \tag{18}$$

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.[4]

Appendix A provides detailed expressions for the Jacobian of $vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]$ and the spectral score of dynamic bifactor models, while appendix B includes numerically reliable and efficient formulae for their information matrix. Those expressions make extensive use of the complex version of the Woodbury formula described in section 2.3. We can also exploit the same formula to compute the quadratic form $\mathbf{z}^{\mathbf{y}*}_j\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{z}^{\mathbf{y}}_j$ as

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

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

where

$$\mathbf{z}_j^{\mathbf{x}^K}(\boldsymbol{\theta}) = E[\mathbf{z}_j^{\mathbf{x}}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] = \mathbf{G}_{\mathbf{xx}}(\lambda_j)\mathbf{C}'(e^{i\lambda_j})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{z}_j^{\mathbf{y}} = \boldsymbol{\Omega}(\lambda_j)\mathbf{C}'(e^{i\lambda_j})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda_j)\mathbf{z}_j^{\mathbf{y}} \tag{19}$$

denotes the filtered value of $\mathbf{z}_j^{\mathbf{x}}$ given the observed series and the current parameter values from (6).

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 described 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, while Demos and Sentana (1998) adapted it to conditionally heteroskedastic factor models in which the common factors followed GARCH-type processes.

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. This idea led us to propose a spectral version of the EM algorihtm for dynamic factor models with only pervasive factors in a companion paper (see Fiorentini, Galesi and Sentana (2014)). In order to apply the same idea to bifactor models, we need to do some additional algebra.

3.1 Complete log-likelihood function

Consider a situation in which the $(R+1)$ common factors $\mathbf{x}_t$ were also observed. The joint spectral density of $\mathbf{y}_t$ and $\mathbf{x}_t$, which is given by

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

could be diagonalised as

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

with

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

and

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

Let us define as $[\mathbf{Z}^{\mathbf{y}}|\mathbf{Z}^{\mathbf{x}}]$ as the Fourier transform of the $T \times (N+1+R)$ matrix

$$[\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{x}_g, \mathbf{x}_1, \ldots, \mathbf{x}_R] = [\mathbf{Y}|\mathbf{X}],$$

so that the joint periodogram of $\mathbf{y}_t$ and $\mathbf{x}_t$ at frequency λ_j could be quickly computed as

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

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

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

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

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

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

$$= \sum_{i=1}^{N}\left[\varkappa - \frac{1}{2}\sum_{j=0}^{T-1} \ln|G_{u_iu_i}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{u_iu_i}(\lambda_j) z_j^{u_i} z_j^{u_i*}\right] \tag{20}$$

$$+\varkappa - \frac{1}{2}\sum_{j=0}^{T-1} \ln\left|G_{x_gx_g}(\lambda_j)\right| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{x_gx_g}(\lambda_j) z_j^{x_g} z_j^{x_g*} \tag{21}$$

$$+\sum_{r=1}^{R}\left[\varkappa - \frac{1}{2}\sum_{j=0}^{T-1} \ln|G_{x_rx_r}(\lambda_j)| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{x_rx_r}(\lambda_j) z_j^{u_r} z_j^{u_r*}\right] \tag{22}$$

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

where[5] if country i belongs to region r we have that

$$z_j^{u_i} = z_j^{y_i} - c_{ig}(e^{-i\lambda_j})z_j^{x_g} - -c_{ir}(e^{-i\lambda_j})z_j^{x_r} = z_j^{y_i} - \sum_{k=-m_g}^{n_g} c_{ikg}e^{-ik\lambda} z_j^{x_g} - \sum_{l=-m_r}^{n_r} c_{ilr}e^{-il\lambda} z_j^{x_r}, \tag{23}$$

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

so that

$$z_j^{u_i} z_j^{u_i*} = z_j^{y_i} z_j^{y_i*} - c_{ig}(e^{-i\lambda_j}) z_j^{x_g} z_j^{y_i*} - c_{ir}(e^{-i\lambda_j}) z_j^{x_r} z_j^{y_i*} - c_{ig}(e^{i\lambda_j}) z_j^{y_i} z_j^{x_g*} - c_{ir}(e^{i\lambda_j}) z_j^{y_i} z_j^{x_r*}$$
$$+c_{ig}(e^{-i\lambda_j}) c_{ig}(e^{i\lambda_j}) z_j^{x_g} z_j^{x_g*} + c_{ir}(e^{-i\lambda_j}) c_{ir}(e^{i\lambda_j}) z_j^{x_r} z_j^{x_r*}$$
$$+c_{ig}(e^{-i\lambda_j}) c_{ir}(e^{i\lambda_j}) z_j^{x_g} z_j^{x_r*} + c_{ir}(e^{-i\lambda_j}) c_{ig}(e^{i\lambda_j}) z_j^{x_r} z_j^{x_g*}$$
$$= I_{y_i y_i}(\lambda_j) - c_{ig}(e^{-i\lambda_j}) I_{x_g y_i}(\lambda_j) - c_{ir}(e^{-i\lambda_j}) I_{x_r y_i}(\lambda_j) - c_{ig}(e^{i\lambda_j}) I_{y_i x_g}(\lambda_j) - c_{ir}(e^{i\lambda_j}) I_{y_i x_r}(\lambda_j)$$
$$+c_{ig}(e^{-i\lambda_j}) c_{ig}(e^{i\lambda_j}) I_{x_g x_g}(\lambda_j) + c_{ir}(e^{-i\lambda_j}) c_{ir}(e^{i\lambda_j}) I_{x_r x_r}(\lambda_j)$$
$$+c_{ig}(e^{-i\lambda_j}) c_{ir}(e^{i\lambda_j}) I_{x_g x_r}(\lambda_j) + c_{ir}(e^{-i\lambda_j}) c_{ig}(e^{i\lambda_j}) I_{x_r x_g}(\lambda_j) = I_{u_i u_i}(\lambda_j).$$

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 $\mathbf{x}$, $l(\mathbf{X})$, and the log-likelihood function of $\mathbf{y}_1, \ldots, \mathbf{y}_N$ given $\mathbf{x}$, $l(\mathbf{Y}|\mathbf{X})$. In turn, each of those components can be decomposed as the sum of univariate log-likelihoods. Specifically, $l(\mathbf{Y}|\mathbf{X})$ can be computed as in (20) by exploiting the diagonality of $\mathbf{G_{uu}}(\lambda_j)$, while $l(\mathbf{X})$ coincides with the sum of (21) and (22) by the diagonality of $\mathbf{G_{xx}}(\lambda_j)$.

Importantly, all the above expressions can be computed using real arithmetic only since

$$\begin{aligned}
c_{ig}(e^{-i\lambda_j}) I_{x_g y_i}(\lambda_j) + c_{ig}(e^{i\lambda_j}) I_{y_i x_g}(\lambda_j) &= 2\Re\left[c_{ig}(e^{-i\lambda_j}) I_{x_g y_i}(\lambda_j)\right], \\
c_{ir}(e^{-i\lambda_j}) I_{x_r y_i}(\lambda_j) + c_{ir}(e^{i\lambda_j}) I_{y_i x_r}(\lambda_j) &= 2\Re\left[c_{ir}(e^{-i\lambda_j}) I_{x_r y_i}(\lambda_j)\right], \\
c_{ig}(e^{-i\lambda_j}) c_{ir}(e^{i\lambda_j}) I_{x_g x_r}(\lambda_j) + c_{ir}(e^{-i\lambda_j}) c_{ig}(e^{i\lambda_j}) I_{x_r x_g}(\lambda_j) &= 2\Re\left[c_{ig}(e^{-i\lambda_j}) c_{ir}(e^{i\lambda_j}) I_{x_g x_r}(\lambda_j)\right], \\
c_{ig}(e^{-i\lambda_j}) c_{ig}(e^{i\lambda_j}) I_{x_g x_g}(\lambda_j) &= \left\| c_{ig}(e^{-i\lambda_j}) \right\|^2 I_{x_g x_g}(\lambda_j)
\end{aligned}$$

and

$$c_{ir}(e^{-i\lambda_j}) c_{ir}(e^{i\lambda_j}) I_{x_r x_r}(\lambda_j) = \left\| c_{ir}(e^{-i\lambda_j}) \right\|^2 I_{x_r x_r}(\lambda_j).$$

Let us classify the parameters into three blocks:

1. the parameters that characterise the spectral density of $\mathbf{x}_t : \boldsymbol{\theta}_x = (\boldsymbol{\theta}'_{x_g}, \boldsymbol{\theta}'_{x_1},, \ldots, \boldsymbol{\theta}'_{x_R})'$

2. the parameters that characterise the spectral density of u_{it} $(i = 1, \ldots, N) : \boldsymbol{\psi} = (\psi_1, \ldots, \psi_N)'$ 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 global and regional factor on each observed variable: $\mathbf{c}_{ig} = (c_{i,-m_g,g}, \ldots, c_{i,0,g}, \ldots, c_{i,n_g,g})'$ and $\mathbf{c}_{ir} = (c_{i,-m_r,r}, \ldots, c_{i,0,r}, \ldots, c_{i,n_r,r})'$.

Importantly, $\boldsymbol{\theta}_{x_g}$ only appear in (21), $\boldsymbol{\theta}_{x_r}$ in (22), while $\boldsymbol{\theta}_{\mathbf{u}_i}$, $\mathbf{c}_{ig}$ and $\mathbf{c}_{ir}$ appear in (20). This sequential cut on the joint spectral density confirms that z^{x_g} and z^{x_r}, and therefore x_{gt} and x_{rt}, would be weakly exogenous for ψ_i, $\boldsymbol{\theta}_{\mathbf{u}_i}$, $\mathbf{c}_{ig}$ and $\mathbf{c}_{ir}$ (see Engle, Hendry and Richard (1983)).

Moreover, the fact that f_{gt} and f_{rt} are uncorrelated at all leads and lags with v_{it} implies that x_{gt} and x_{rt} would be strongly exogenous too.

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

$$\frac{\partial l(\mathbf{Y},\mathbf{x})}{\partial \boldsymbol{\theta}_{x_g}} = \frac{\partial l(\mathbf{x}_g)}{\partial \boldsymbol{\theta}_{x_g}} = \frac{1}{2}\sum_{j=0}^{T-1} \frac{\partial G_{x_g x_g}(\lambda_j)}{\partial \boldsymbol{\theta}_{x_g}} G^{-2}_{x_g x_g}(\lambda_j)\left[2\pi z_j^{x_g} z_j^{x_g *} - G_{x_g x_g}(\lambda_j)\right], \tag{24a}$$

$$\frac{\partial l(\mathbf{Y},\mathbf{x})}{\partial \boldsymbol{\theta}_{x_r}} = \frac{\partial l(\mathbf{x}_r)}{\partial \boldsymbol{\theta}_{x_r}} = \frac{1}{2}\sum_{j=0}^{T-1} \frac{\partial G_{x_r x_r}(\lambda_j)}{\partial \boldsymbol{\theta}_{x_r}} G^{-2}_{x_r x_r}(\lambda_j)\left[2\pi z_j^{x_r} z_j^{x_r *} - G_{x_r x_r}(\lambda_j)\right] \tag{24b}$$

$$\frac{\partial l(\mathbf{Y},\mathbf{x})}{\partial \boldsymbol{\theta}_{u_i}} = \frac{\partial l(\mathbf{y}_i|\mathbf{X})}{\partial \boldsymbol{\theta}_{u_i}} = \frac{1}{2}\sum_{j=0}^{T-1} \frac{\partial G_{u_i u_i}(\lambda_j)}{\partial \boldsymbol{\theta}_{u_i}} G^{-2}_{u_i u_i}(\lambda_j)\left[2\pi z_j^{u_i} z_j^{u_i *} - G_{u_i u_i}(\lambda_j)\right] \tag{24c}$$

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

$$= \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{u_i u_i}(\lambda_j)\left[\begin{array}{c} \left(z_j^{y_i} - \sum_{k=-m_g}^{n_g} c_{ikg} e^{-ik\lambda} z_j^{x_g} - \sum_{l=-m_r}^{n_r} c_{ilr} e^{-il\lambda} z_j^{x_r}\right) e^{ik\lambda_j} z_j^{x_g *} \\ + e^{-ik\lambda_j} z_j^{x_g}\left(z_j^{y_i *} - \sum_{k=-m_g}^{n_g} c_{ikg} e^{ik\lambda} z_j^{x_g *} - \sum_{l=-m_r}^{n_r} c_{ilr} e^{il\lambda} z_j^{x_r *}\right) \end{array}\right] \tag{24d}$$

$$\frac{\partial l(\mathbf{Y},\mathbf{x})}{\partial c_{ilr}} = \frac{\partial l(\mathbf{y}_i|\mathbf{X})}{\partial c_{ikr}} = \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{u_i u_i}(\lambda_j)\left[z_j^{u_i} e^{il\lambda_j} z_j^{x_r *} + e^{-il\lambda_j} z_j^{xr} z_j^{u_i *}\right]$$

$$= \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{u_i u_i}(\lambda_j)\left[\begin{array}{c} \left(z_j^{y_i} - \sum_{k=-m_g}^{n_g} c_{ikg} e^{-ik\lambda} z_j^{x_g} - \sum_{l=-m_r}^{n_r} c_{ilr} e^{-il\lambda} z_j^{x_r}\right) e^{il\lambda_j} z_j^{x_r *} \\ + e^{-il\lambda_j} z_j^{x_r}\left(z_j^{y_i *} - \sum_{k=-m_g}^{n_g} c_{ikg} e^{ik\lambda} z_j^{x_g *} - \sum_{l=-m_r}^{n_r} c_{ilr} e^{il\lambda} z_j^{x_r *}\right) \end{array}\right] \tag{24e}$$

where we have used the fact that

$$\begin{aligned} \frac{\partial z_j^{u_i}}{\partial c_{ikg}} &= -e^{-ik\lambda} z_j^{x_g} \\ \frac{\partial z_j^{u_i}}{\partial c_{ilr}} &= -e^{-il\lambda} z_j^{x_r} \end{aligned}$$

in view of (23).

Expression (24a) confirms that the MLE of $\boldsymbol{\theta}_{x_g}$ would be obtained from a univariate time series model for x_{gt}, and the same applies to $\boldsymbol{\theta}_{x_r}$. However, since $G_{x_g x_g}(\lambda_j)$ also depends on $\boldsymbol{\theta}_{x_g}$, there are no closed form solutions for models with MA components. Although it would be straightforward to adapt the indirect inference procedures we have developed in our companion paper (see Fiorentini, Galesi and Sentana (2014)) to deal with general ARMA processes without resorting to the numerical maximisation of (21), in what follows we only consider pure autoregressions. Obviously, the same comments apply to $\boldsymbol{\theta}_{x_r}$.

In this regard, if we consider the AR(2) example for x_r in (4), the derivatives of $G_{x_r x_r}(\lambda)$

with respect to α_{1x_r} and α_{2x_g} would be

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

Hence, the log-likelihood scores would become

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

and

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

where we have exploited the Yule-Walker equations to show that

$$\begin{aligned}&\sum_{j=0}^{T-1}\frac{(\cos\lambda - \alpha_{1x_r} - \alpha_{2x_r}\cos\lambda)}{(1+\alpha_{1x_r}^2+\alpha_{2x_r}^2 - 2\alpha_{1x_r}(1-\alpha_{2x_r})\cos\lambda - 2\alpha_{2x_r}\cos 2\lambda)}\\ &\qquad = \gamma_{x_r x_r}(1) - \alpha_{1x_r}\gamma_{x_r x_r}(0) - \alpha_{2x_r}\gamma_{x_r x_r}(1) = 0,\\ &\sum_{j=0}^{T-1}\frac{(\cos 2\lambda - \alpha_{1x_r}\cos\lambda - \alpha_{2x_r})}{(1+\alpha_{1x_r}^2+\alpha_{2x_r}^2 - 2\alpha_{1x_r}(1-\alpha_{2x_r})\cos\lambda - 2\alpha_{2x_r}\cos 2\lambda)}\\ &\qquad = \gamma_{x_r x_r}(2) - \alpha_{1x_r}\gamma_{x_r x_r}(1) - \alpha_{2x_r}\gamma_{x_r x_r}(0) = 0.\end{aligned}$$

As a result, when we set both scores to 0 we would be left with the system of equations

$$\sum_{j=0}^{T-1}\left[z_j^{x_r} z_j^{x_r *}\otimes\begin{pmatrix}1 & \cos\lambda_j\\ \cos\lambda_j & 1\end{pmatrix}\right]\begin{pmatrix}\hat{\alpha}_{1x_r}\\ \hat{\alpha}_{2x_r}\end{pmatrix} = \sum_{j=0}^{T-1}\left[z_j^{x_r} z_j^{x_r *}\otimes\begin{pmatrix}\cos\lambda_j\\ \cos 2\lambda_j\end{pmatrix}\right].$$

But since

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

we would have that

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

and

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

which are the sample (circulant) autocovariances of x_{rt} of orders 0, 1 and 2, respectively. Therefore, the spectral estimators for $\hat{\alpha}_{1x_r}$ and $\hat{\alpha}_{2x_r}$ are (almost) identical to the ones we would obtain in the time domain, which will be given by the solution to the system of equations

$$\begin{pmatrix} \hat{\gamma}_{x_r x_r}(0) & \hat{\gamma}_{x_r x_r}(1) \\ \hat{\gamma}_{x_r x_r}(1) & \hat{\gamma}_{x_r x_r}(0) \end{pmatrix}\begin{pmatrix} \hat{\alpha}_{1x_r} \\ \hat{\alpha}_{2x_r} \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_{x_r x_r}(1) \\ \hat{\gamma}_{x_r x_r}(2) \end{pmatrix},$$

because both $\hat{\gamma}_{x_r x_r}(T-1) = T^{-1}x_{rT}x_{r1}$ and $\hat{\gamma}_{x_r x_r}(T-2) = T^{-1}(x_{rT}x_{r2}+x_{rT-1}x_{r1})$ are $o_p(1)$.

Similar expressions would apply to the dynamic parameters that appear in $\boldsymbol{\theta}_{u_i}$ for a given value of $\mathbf{c}_{ig}$ and $\mathbf{c}_{ir}$ in view of (24c), 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 in (4), the partial derivatives of $G_{u_i u_i}(\lambda)$ with respect to ψ_i and α_{1u_i} would be

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

Hence, the corresponding log-likelihood scores would be

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

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

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

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, (24d) and (24e) would allow us to obtain the ML estimators of $\mathbf{c}_{ig}$ and $\mathbf{c}_{ir}$ for given values of $\boldsymbol{\theta}_{u_i}$. In particular, if we write together the derivatives for c_{ikg} $(k=-m_g,\dots,0,\dots,n_g)$ and c_{ikr} $(k=-m_r,\dots,0,\dots,n_r)$ 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_g\lambda_j} z_j^{x_g} z_j^{x_g*} e^{-im_g\lambda_j} + e^{im_g\lambda_j} z_j^{x_g} z_j^{x_g*} e^{-im_g\lambda_j} & \cdots & e^{-in_g\lambda_j} z_j^{x_g} z_j^{x_g*} e^{-im_g\lambda_j} + e^{im_g\lambda_j} z_j^{x_g} z_j^{x_g*} e^{in_g\lambda_j} & e^{im_g\lambda_j} z_j^{x_g} z_j^{x_r*} e^{-im_r\lambda_j} + e^{im_r\lambda_j} z_j^{x_r} z_j^{x_g*} e^{-im_g\lambda_j} & \cdots & e^{im_g\lambda_j} z_j^{x_g} z_j^{x_r*} e^{in_r\lambda_j} + e^{-in_r\lambda_j} z_j^{x_r} z_j^{x_g*} e^{-im_g\lambda_j} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
e^{im_g\lambda_j} z_j^{x_g} z_j^{x_g*} e^{in_g\lambda_j} + e^{-in_g\lambda_j} z_j^{x_g} z_j^{x_g*} e^{-im_g\lambda_j} & \cdots & e^{-in_g\lambda_j} z_j^{x_g} z_j^{x_g*} e^{in_g\lambda_j} + e^{-in_g\lambda_j} z_j^{x_g} z_j^{x_g*} e^{in_g\lambda_j} & e^{-in_g\lambda_j} z_j^{x_g} z_j^{x_r*} e^{-im_r\lambda_j} + e^{im_r\lambda_j} z_j^{x_r} z_j^{x_g*} e^{in_g\lambda_j} & \cdots & e^{-in_g\lambda_j} z_j^{x_g} z_j^{x_r*} e^{in_r\lambda_j} + e^{-in_r\lambda_j} z_j^{x_r} z_j^{x_g*} e^{in_g\lambda_j} \\
e^{im_g\lambda_j} z_j^{x_g} z_j^{x_r*} e^{-im_r\lambda_j} + e^{im_r\lambda_j} z_j^{x_r} z_j^{x_g*} e^{-im_g\lambda_j} & \cdots & e^{-in_g\lambda_j} z_j^{x_g} z_j^{x_r*} e^{-im_r\lambda_j} + e^{im_r\lambda_j} z_j^{x_r} z_j^{x_g*} e^{in_g\lambda_j} & e^{im_r\lambda_j} z_j^{x_r} z_j^{x_r*} e^{-im_r\lambda_j} + e^{im_r\lambda_j} z_j^{x_r} z_j^{x_r*} e^{-im_r\lambda_j} & \cdots & e^{im_r\lambda_j} z_j^{x_r} z_j^{x_r*} e^{in_r\lambda_j} + e^{-in_r\lambda_j} z_j^{x_r} z_j^{x_r*} e^{-im_r\lambda_j} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
e^{im_g\lambda_j} z_j^{x_g} z_j^{x_r*} e^{in_r\lambda_j} + e^{-in_r\lambda_j} z_j^{x_r} z_j^{x_g*} e^{-im_g\lambda_j} & \cdots & e^{-in_g\lambda_j} z_j^{x_g} z_j^{x_r*} e^{in_r\lambda_j} + e^{-in_r\lambda_j} z_j^{x_r} z_j^{x_g*} e^{in_g\lambda_j} & e^{-in_r\lambda_j} z_j^{x_r} z_j^{x_r*} e^{-im_r\lambda_j} + e^{im_r\lambda_j} z_j^{x_r} z_j^{x_r*} e^{in_r\lambda_j} & \cdots & e^{-in_r\lambda_j} z_j^{x_r} z_j^{x_r*} e^{in_r\lambda_j} + e^{-in_r\lambda_j} z_j^{x_r} z_j^{x_r*} e^{in_r\lambda_j}
\end{pmatrix}\right]\begin{pmatrix}\tilde{c}_{i,-m_g,g}\\ \vdots \\ \tilde{c}_{i,n_gg}\\ \tilde{c}_{i,-m_r,r}\\ \vdots \\ \tilde{c}_{i,n_r,r}\end{pmatrix}
$$

$$
= \sum_{j=0}^{T-1} G_{u_iu_i}^{-1}(\lambda_j)\begin{pmatrix}
z_j^{y_i} z_j^{x_g*} e^{-im_g\lambda_j} + z_j^{y_i*} z_j^{x_g} e^{im_g\lambda_j}\\
\vdots\\
z_j^{y_i} z_j^{x_g*} e^{in_g\lambda_j} + z_j^{y_i*} z_j^{x_g} e^{-in_g\lambda_j}\\
z_j^{y_i} z_j^{x_r*} e^{-im_r\lambda_j} + z_j^{y_i*} z_j^{x_r} e^{im_r\lambda_j}\\
\vdots\\
z_j^{y_i} z_j^{x_r*} e^{in_r\lambda_j} + z_j^{y_i*} z_j^{x_r} e^{-in_r\lambda_j}
\end{pmatrix}.
$$

Thus, unrestricted MLE's of $\mathbf{c}_{ig}$ and $\mathbf{c}_{ir}$ could be obtained from N univariate distributed lag weighted least squares regressions of each y_{it} on x_{gt} and the appropriate x_{rt} 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 (4). In that case, the matrix on the left hand of the normal equations becomes

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

while the vector on the right hand side will be

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

In principle, we could carry out a zig-zag procedure that would estimate $\mathbf{c}_{ig}$ and $\mathbf{c}_{ir}$ for given $\boldsymbol{\theta}_{u_i}$, and then $\boldsymbol{\theta}_{u_i}$ for a given $\mathbf{c}_{ig}$ and $\mathbf{c}_{ir}$. 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. In that case, the above equations could be written in terms of the elements of the covariance and the first autocovariance matrices of y_t, x_{gt} and x_{rt}.

3.2 Expected log-likelihood function

In practice, of course, we do not observe $\mathbf{x}_t$. 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 $l(\mathbf{Y}|\mathbf{X}) + l(\mathbf{X})$ conditional on $\mathbf{Y}$ and the current parameter estimates, $\boldsymbol{\theta}^{(n)}$. The rationale stems from the fact that $l(\mathbf{Y}, \mathbf{X})$ can also be factorized as $l(\mathbf{Y}) + l(\mathbf{X}|\mathbf{Y})$. 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 $l(\mathbf{Y}, \mathbf{X})$ must represent an increase in $l(\mathbf{Y})$. This is the generalised EM principle.

In the E step we must compute

$$\begin{aligned}
E[l(\mathbf{x}_g)|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}^{(n)}] &= \varkappa - \frac{1}{2}\sum_{j=0}^{T-1}\ln\left|G_{x_g x_g}(\lambda_j)\right| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{x_g x_g}(\lambda_j)E[z_j^{x_g} z_j^{x_g *}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}^{(n)}],\\
E[l(\mathbf{x}_r)|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}^{(n)}] &= \varkappa - \frac{1}{2}\sum_{j=0}^{T-1}\ln\left|G_{x_r x_r}(\lambda_j)\right| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{x_r x_r}(\lambda_j)E[z_j^{x_r} z_j^{x_r *}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}^{(n)}],\\
E[l(\mathbf{y}_i|\mathbf{X})|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}^{(n)}] &= \varkappa - \frac{1}{2}\sum_{j=0}^{T-1}\ln\left|G_{u_i u_i}(\lambda_j)\right| - \frac{2\pi}{2}\sum_{j=0}^{T-1} G^{-1}_{u_i u_i}(\lambda_j)E[z_j^{u_i} z_j^{u_i *}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}^{(n)}].
\end{aligned}$$

But

$$\begin{aligned}
E[\mathbf{z}_j^{\mathbf{x}}\mathbf{z}_j^{\mathbf{x}*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}^{(n)}] &= \mathbf{z}_j^{\mathbf{x}^K}(\boldsymbol{\theta}^{(n)})\mathbf{z}_j^{\mathbf{x}^K *}(\boldsymbol{\theta}^{(n)}) + E\left\{[\mathbf{z}_j^{\mathbf{x}} - \mathbf{z}_j^{\mathbf{x}^K}(\boldsymbol{\theta}^{(n)})][\mathbf{z}_j^{\mathbf{x}*} - \mathbf{z}_j^{\mathbf{x}^K *}(\boldsymbol{\theta}^{(n)})]|\mathbf{z}_j^{\mathbf{y}},\boldsymbol{\theta}^{(n)}\right\}\\
&= \mathbf{I}^{(n)}_{\mathbf{x}^K\mathbf{x}^K}(\lambda_j) + \boldsymbol{\Omega}^{(n)}(\lambda_j),
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{I}_{\mathbf{x}^K\mathbf{x}^K}(\lambda) &= 2\pi\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\\
&= 2\pi\boldsymbol{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{-i\lambda})\boldsymbol{\Omega}(\lambda). \qquad (25)
\end{aligned}$$

is the periodogram of the smoothed values of the $R+1$ common factors $\mathbf{x}$ and

$$E\left\{[\mathbf{z}_j^{\mathbf{x}} - \mathbf{z}_j^{\mathbf{x}^K}(\boldsymbol{\theta})][\mathbf{z}_j^{\mathbf{x}*} - \mathbf{z}_j^{\mathbf{x}^K *}(\boldsymbol{\theta})]|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}\right\} = \boldsymbol{\Omega}(\lambda_j).$$

In turn, if we define

$$\mathbf{I}_{\mathbf{yx}^K}(\lambda) = \mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda) = \mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{-i\lambda})\boldsymbol{\Omega}(\lambda)$$

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

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

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.

The assumed bifactor structure implies that for the i^{th} series, the above expression reduces to

$$\begin{aligned}
I^{(\mathrm{N})}_{u_i u_i}(\lambda_j) &= E[z_j^{u_i} z_j^{u_i *}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}^{(n)}] = I_{y_i y_i}(\lambda_j)\\
&-c_{ig}(e^{-i\lambda_j})I^{(n)}_{x_g^K y_i}(\lambda_j) - c_{ir}(e^{-i\lambda_j})I^{(n)}_{x_r^K y_i}(\lambda_j) - I^{(n)}_{y_i x_g^K}(\lambda_j)c_{ig}(e^{i\lambda_j}) - I^{(n)}_{y_i x_r^K}(\lambda_j)c_{ir}(e^{i\lambda_j})\\
&+[I^{(n)}_{x_g^K x_g^K}(\lambda_j) + \omega^{(n)}_{gg}(\lambda_j)]c_{ig}(e^{-i\lambda_j})c_{ig}(e^{i\lambda_j}) + [I^{(n)}_{x_r^K x_r^K}(\lambda_j) + \omega^{(n)}_{rr}(\lambda_j)]c_{ir}(e^{-i\lambda_j})c_{ir}(e^{i\lambda_j})\\
&+[I^{(n)}_{x_g^K x_r^K}(\lambda_j) + \omega^{(n)}_{gr}(\lambda_j)]c_{ig}(e^{-i\lambda_j})c_{ir}(e^{i\lambda_j}) + [I^{(n)}_{x_r^K x_g^K}(\lambda_j) + \omega^{(n)}_{rg}(\lambda_j)]c_{ir}(e^{-i\lambda_j})c_{ig}(e^{i\lambda_j}).
\end{aligned}$$

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

$$E[l(\mathbf{x}_g)|\mathbf{Y},\boldsymbol{\theta}^{(n)}] = \varkappa - \frac{1}{2}\sum_{j=0}^{T-1}\ln\left|G_{x_gx_g}(\lambda_j)\right| - \frac{2\pi}{2}\sum_{j=0}^{T-1}G^{-1}_{x_gx_g}(\lambda_j)\left[I^{(n)}_{x_g^Kx_g^K}(\lambda_j)+\omega^{(n)}_{gg}(\lambda_j)\right], \quad (26)$$

$$E[l(\mathbf{x}_r)|\mathbf{Y},\boldsymbol{\theta}^{(n)}] = \varkappa - \frac{1}{2}\sum_{j=0}^{T-1}\ln\left|G_{x_rx_r}(\lambda_j)\right| - \frac{2\pi}{2}\sum_{j=0}^{T-1}G^{-1}_{x_rx_r}(\lambda_j)\left[I^{(n)}_{x_r^Kx_r^K}(\lambda_j)+\omega^{(n)}_{rr}(\lambda_j)\right], \quad (27)$$

$$E[l(\mathbf{y}_i|\mathbf{X})|\mathbf{Y},\boldsymbol{\theta}^{(n)}] = \varkappa - \frac{1}{2}\sum_{j=0}^{T-1}\ln\left|G_{u_iu_i}(\lambda_j)\right| - \frac{2\pi}{2}\sum_{j=0}^{T-1}G^{-1}_{u_iu_i}(\lambda_j)I^{(\mathrm{N})}_{u_iu_i}(\lambda_j). \quad (28)$$

We can then maximise $E[l(\mathbf{x}_g)|\mathbf{Y},\boldsymbol{\theta}^{(n)}]$ in (26) with respect to $\boldsymbol{\theta}_{x_g}$ to update those parameters, and the same applies to (27) and $\boldsymbol{\theta}_{x_r}$. Similarly, we can maximise $E[l(\mathbf{y}_i|\mathbf{X})|\mathbf{Y},\boldsymbol{\theta}^{(n)}]$ with respect to $\mathbf{c}_{ig}$, $\mathbf{c}_{ir}$, 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 (26) and (21), it is easy to see that

$$\frac{\partial E[l(\mathbf{x}_g)|\mathbf{Y},\boldsymbol{\theta}^{(n)}]}{\partial\boldsymbol{\theta}_{x_g}} = \frac{1}{2}\sum_{j=0}^{T-1}\frac{\partial G_{x_gx_g}(\lambda_j)}{\partial\boldsymbol{\theta}_{x_g}}G^{-2}_{x_gx_g}(\lambda_j)\left\{2\pi\left[I^{(n)}_{x_g^Kx_g^K}(\lambda_j)+\omega^{(n)}_{gg}(\lambda_j)\right]-G_{x_gx_g}(\lambda_j)\right\},$$

which, not surprisingly, coincides with the the expected value of (24a) given $\mathbf{Y}$ and the current parameter estimates, $\boldsymbol{\theta}^{(n)}$. As a result, for the AR(1) process for x_g in (4) we will have

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

whence

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

Likewise, we will have that

$$\frac{\partial E[l(\mathbf{x}_r)|\mathbf{Y},\boldsymbol{\theta}^{(n)}]}{\partial\boldsymbol{\theta}_{x_r}} = \frac{1}{2}\sum_{j=0}^{T-1}\frac{\partial G_{x_rx_r}(\lambda_j)}{\partial\boldsymbol{\theta}_{x_r}}G^{-2}_{x_rx_r}(\lambda_j)\left\{2\pi\left[I^{(n)}_{x_r^Kx_r^K}(\lambda_j)+\omega^{(n)}_{rr}(\lambda_j)\right]-G_{x_rx_r}(\lambda_j)\right\}.$$

Hence, in the case of the AR(2) process for x_{rt} in (4), the expected log-likelihood scores become

$$\begin{aligned}\frac{\partial E[l(\mathbf{x}_r)|\mathbf{Y},\boldsymbol{\theta}^{(n)}]}{\partial\alpha_{1x_r}} &= 2\pi\sum_{j=0}^{T-1}(\cos\lambda_j-\alpha_{1x_r}-\alpha_{2x_r}\cos\lambda_j)\left[I^{(n)}_{x_r^Kx_r^K}(\lambda_j)+\omega^{(n)}_{rr}(\lambda_j)\right],\\ \frac{\partial E[l(\mathbf{x}_r)|\mathbf{Y},\boldsymbol{\theta}^{(n)}]}{\partial\alpha_{2x_r}} &= 2\pi\sum_{j=0}^{T-1}2(\cos 2\lambda_j-\alpha_{1x_r}\cos\lambda_j-\alpha_{2x_r})\left[I^{(n)}_{x_r^Kx_r^K}(\lambda_j)+\omega^{(n)}_{rr}(\lambda_j)\right],\end{aligned}$$

so that the updated autoregressive coefficients will be the solution to the system of equations

$$\begin{aligned}
&\sum_{j=0}^{T-1}\left\{\left[I^{(n)}_{x_r^K x_r^K}(\lambda_j)+\omega^{(n)}_{rr}(\lambda_j)\right]\otimes\begin{pmatrix}1 & \cos\lambda_j\\ \cos\lambda_j & 1\end{pmatrix}\right\}\begin{pmatrix}\hat{\alpha}_{1x_r}\\ \hat{\alpha}_{2x_r}\end{pmatrix}\\
=&\sum_{j=0}^{T-1}\left\{\left[I^{(n)}_{x_r^K x_r^K}(\lambda_j)+\omega^{(n)}_{rr}(\lambda_j)\right]\otimes\begin{pmatrix}\cos\lambda_j\\ \cos 2\lambda_j\end{pmatrix}\right\}.
\end{aligned}$$

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

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

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

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

Finally, the derivatives of (28) with respect to c_{ikg} ($k=-m_g,\ldots,0,\ldots,n_g$) and c_{ilr} ($l=-m_r,\ldots,0,\ldots,n_r$) for fixed values of $\boldsymbol{\theta}_{u_i}$ will give rise to a set of modified "weighted" normal equations analogous to the ones in the previous section but with cross-product terms of the form $z_j^{x_g}z_j^{x_r*}$ replaced by $[I^{(n)}_{x_g^K x_r^K}(\lambda_j)+\omega^{(n)}_{gr}(\lambda_j)]$.

For the example in (4), the matrix on the left hand of the normal equations becomes

$$2\sum_{j=0}^{T-1}G^{-1}_{u_iu_i}(\lambda_j)\left(\begin{array}{cccc}
[I^{(n)}_{x_g^Kx_g^K}(\lambda_j)+\omega^{(n)}_{gg}(\lambda_j)] & \cos\lambda_j[I^{(n)}_{x_g^Kx_g^K}(\lambda_j)+\omega^{(n)}_{gg}(\lambda_j)] & \Re[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)+\omega^{(n)}_{gr}(\lambda_j)] & \cos\lambda_j\Re[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)]-\sin\lambda_j\Im[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)]\\
\cos\lambda_j[I^{(n)}_{x_g^Kx_g^K}(\lambda_j)+\omega^{(n)}_{gg}(\lambda_j)] & [I^{(n)}_{x_g^Kx_g^K}(\lambda_j)+\omega^{(n)}_{gg}(\lambda_j)] & \cos\lambda_j\Re[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)]+\sin\lambda_j\Im[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)] & \Re[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)+\omega^{(n)}_{gr}(\lambda_j)]\\
\Re[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)+\omega^{(n)}_{gr}(\lambda_j)] & \cos\lambda_j\Re[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)]+\sin\lambda_j\Im[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)] & [I^{(n)}_{x_r^Kx_r^K}(\lambda_j)+\omega^{(n)}_{rr}(\lambda_j)] & \cos\lambda_j[I^{(n)}_{x_r^Kx_r^K}(\lambda_j)+\omega^{(n)}_{rr}(\lambda_j)]\\
\cos\lambda_j\Re[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)]-\sin\lambda_j\Im[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)] & \Re[I^{(n)}_{x_g^Kx_r^K}(\lambda_j)+\omega^{(n)}_{gr}(\lambda_j)] & \cos\lambda_j[I^{(n)}_{x_r^Kx_r^K}(\lambda_j)+\omega^{(n)}_{rr}(\lambda_j)] & [I^{(n)}_{x_r^Kx_r^K}(\lambda_j)+\omega^{(n)}_{rr}(\lambda_j)]
\end{array}\right)$$

while the vector on the right hand side will be

$$2\sum_{j=0}^{T-1} G_{u_iu_i}^{-1}(\lambda_j)\begin{pmatrix} \Re[I_{y_ix_g^K}^{(n)}(\lambda_j)] \\ \cos\lambda_j\Re[I_{y_ix_g^K}^{(n)}(\lambda_j)] - \sin\lambda_j\Im[I_{y_ix_g^K}^{(n)}(\lambda_j)] \\ \Re[I_{y_ix_r^K}^{(n)}(\lambda_j)] \\ \cos\lambda_j\Re[I_{y_ix_r^K}^{(n)}(\lambda_j)] - \sin\lambda_j\Im[I_{y_ix_r^K}^{(n)}(\lambda_j)] \end{pmatrix}$$

In principle, we could carry out a zig-zag procedure that would estimate $\mathbf{c}_{ig}$, $\mathbf{c}_{ir}$ and ψ_i for given $\boldsymbol{\theta}_{u_i}$ and $\boldsymbol{\theta}_{u_i}$ for given $\mathbf{c}_{ig}$, $\mathbf{c}_{ir}$ 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.

3.3 Alternative marginal scores

As is well known, the EM algorithm slows down considerably near the optimum. At that point, the best practical strategy would be to switch to a first derivative-based method. Fortunately, 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

$$\begin{aligned}
\frac{\partial l(\mathbf{Y})}{\partial\boldsymbol{\theta}_{x_g}} &= \frac{1}{2}\sum_{j=0}^{T-1}\frac{\partial G_{x_gx_g}(\lambda_j)}{\partial\boldsymbol{\theta}_{x_g}}G_{x_gx_g}^{-2}(\lambda_j)\left[2\pi E[z_j^{x_g}z_j^{x_g*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] - G_{x_gx_g}(\lambda_j)\right], \\
\frac{\partial l(\mathbf{Y})}{\partial\boldsymbol{\theta}_{x_r}} &= \frac{1}{2}\sum_{j=0}^{T-1}\frac{\partial G_{x_rx_r}(\lambda_j)}{\partial\boldsymbol{\theta}_{x_r}}G_{x_rx_r}^{-2}(\lambda_j)\left[2\pi E[z_j^{x_r}z_j^{x_r*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] - G_{xx}(\lambda_j)\right], \\
\frac{\partial l(\mathbf{Y})}{\partial\boldsymbol{\theta}_{u_i}} &= \frac{1}{2}\sum_{j=0}^{T-1}\frac{\partial G_{u_iu_i}(\lambda_j)}{\partial\boldsymbol{\theta}_{u_i}}G_{u_iu_i}^{-2}(\lambda_j)\left[2\pi E[z_j^{u_i}z_j^{u_i*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] - G_{u_iu_i}(\lambda_j)\right], \\
\frac{\partial l(\mathbf{Y})}{\partial c_{ikg}} &= \frac{2\pi}{2}\sum_{j=0}^{T-1}G_{u_iu_i}^{-1}(\lambda_j)\left[e^{ik\lambda_j}E[z_j^{u_i}z_j^{x_g*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] + e^{-ik\lambda_j}E[z_j^{x_g}z_j^{u_i*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]\right], \\
\frac{\partial l(\mathbf{Y})}{\partial c_{ilr}} &= \frac{2\pi}{2}\sum_{j=0}^{T-1}G_{u_iu_i}^{-1}(\lambda_j)\left[e^{il\lambda_j}E[z_j^{u_i}z_j^{x_r*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}] + e^{-il\lambda_j}E[z_j^{x_r}z_j^{u_i*}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]\right]
\end{aligned}$$

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

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

where

$$\begin{gathered}
\mathbf{z}_j^{\mathbf{u}^K}=E[\mathbf{z}_j^{\mathbf{u}}|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]=\mathbf{G}_{\mathbf{uu}}(\lambda_j)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{z}_j^{\mathbf{y}}=\mathbf{z}_j^{\mathbf{y}}-\mathbf{C}(e^{-i\lambda})\mathbf{z}_j^{\mathbf{x}^K},\\
E[(\mathbf{z}_j^{\mathbf{u}}-\mathbf{z}_j^{\mathbf{u}^K})(\mathbf{z}_j^{\mathbf{u}*}-\mathbf{z}_j^{\mathbf{u}^K*})|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]=\mathbf{C}(e^{-i\lambda_j})\boldsymbol{\Omega}(\lambda_j)\mathbf{C}'(e^{i\lambda_j}),\\
E[(\mathbf{z}_j^{\mathbf{u}}-\mathbf{z}_j^{\mathbf{u}_K})(\mathbf{z}_j^{\mathbf{x}*}-\mathbf{z}_j^{\mathbf{x}^K*})|\mathbf{Z}^{\mathbf{y}},\boldsymbol{\theta}]=\mathbf{C}(e^{-i\lambda_j})\boldsymbol{\Omega}(\lambda_j),
\end{gathered}$$

$$\begin{gathered}
\mathbf{I}_{\mathbf{u}^K\mathbf{u}^K}(\lambda)=2\pi\mathbf{G}_{\mathbf{uu}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{G}_{\mathbf{uu}}(\lambda)\\
=2\pi\left[\mathbf{I}_N-\mathbf{C}(e^{-i\lambda})\boldsymbol{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\right]\mathbf{I}_{\mathbf{yy}}(\lambda)\left[\mathbf{I}_N-\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{C}(e^{i\lambda})\boldsymbol{\Omega}(\lambda)\mathbf{C}'(e^{-i\lambda})\right]
\end{gathered}\tag{29}$$

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

$$\begin{gathered}
\mathbf{I}_{\mathbf{x}^K\mathbf{u}^K}(\lambda)=2\pi\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{G}_{\mathbf{uu}}(\lambda)\\
=2\pi\boldsymbol{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{I}_{\mathbf{yy}}(\lambda)\left[\mathbf{I}_N-\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{C}(e^{-i\lambda})\boldsymbol{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda})\right]
\end{gathered}\tag{30}$$

is the co-periodogram between $\mathbf{x}_{t|\infty}^K$ 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 $\boldsymbol{\theta}$.

4 Inflation dynamics across European countries

Increasing economic and financial integration implies that nowadays countries are more sensitive to shocks originating outside their frontiers. In particular, national price levels may be affected by external shocks such as fluctuations in global commodity prices, shifts in global demand, exchange rate swings, or variations in the prices of competing countries. Understanding the extent to which foreign factors determine movements in domestic inflation is a key question for macroeconomic policy.

A recent growing literature tackles this question by employing factor analysis techniques. Ciccarelli and Mojon (2009) estimate a static single factor model for 22 OECD economies over the period 1960-2008 and document that the estimated global factor accounts for about 70 percent of the variance of CPI inflation in those countries. Mumtaz and Surico (2012) estimate a dynamic factor model with drifting coefficients and stochastic volatility for a panel of 164 inflation indicators for the G7 countries, Australia, New Zealand and Spain. These authors find that the historical decline in the level of inflation is shared by most countries in their sample, which is consistent with the idea that a global factor drives the bulk of inflation movements across economies.

At the same time, the inflation rates of closely integrated economies tend to be more correlated with each other than with other countries, which is difficult to square with a single factor model. Motivated by this, we explore the ability of the dynamic bifactor models discussed in section 2.1 to capture inflation dynamics across European countries. The European case is of particular interest because whether EMU has played a decisive role in the observed convergence of inflation rates across its member economies remains an open question. In this regard, Estrada, Galí and López-Salido (2013) examine the extent to which the inflation rates of the original 11 euro area countries and other OECD economies have become synchronised over the period 1999-2012, reporting strong evidence of convergence towards low inflation rates. They also show that other advanced non-euro countries experience similar levels of convergence, which suggests that EMU may not be responsible for the generalised decline in inflation.

We use monthly data on Harmonised Indices of Consumer Prices (HICP) for 25 European economies over the period 1998:1-2014:12.[6] In particular, we consider three groups of countries:

1. the original[7] euro area members: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal and Spain;

2. the new euro area participants: Cyprus, Estonia, Latvia, Lithuania, Malta and Slovakia;

3. other non-EMU countries: Bulgaria, Denmark, Iceland, Norway, Poland, Sweden and United Kingdom.

We focus on year-on-year growth rates of HICP indices excluding energy and unprocessed food, which are widely viewed as the relevant measure to track for inflation targeting purposes; see for example Galí (2002). As a result, we are left with $T = 192$ time series observations. Figure 1, which contains the inflation rates for each country (solid blue line) together with the

[6]Since our aim is to maximise the time span of our sample, we exclude several countries for which data start at later dates: Czech Republic and Slovenia (1999:12-), Hungary and Romania (2000:12-), and Croatia and Switzerland (2004:12-).

[7]We include Greece among the original euro area even though its accession year was 2001.

inflation rate of the European Union (dashed black line), confirms the generalised downward trend in inflation.

For modelling purposes, we assume that the (demeaned) inflation rate of each country is driven by a global factor which affects all European countries, an orthogonal region-specific factor which affects all countries within a region, and an idiosyncratic factor. We also assume that the global and regional factors affect the inflation rate of a country not only through their contemporaneous values but also via their one-month lagged values with country-specific loadings. Further, we assume that all factors (global, regional, and idiosyncratic) follow orthogonal AR(1) processes. Despite the apparent simplicity of our model, each series is effectively the sum of three components: an ARMA(1,1) global component, another ARMA(1,1) regional component and an idiosyncratic AR(1) term.

We estimate our dynamic bifactor model using the EM algorithm developed in previous sections. As starting values, we assume unit loadings on the contemporaneous and lagged values of both common and regional factors, unit specific variances, autoregressive coefficients set to 0.5 for both common and idiosyncratic factors, and 0.3 for regional factors. Importantly, the scoring algorithm fails to achieve convergence from these initial values. To speed up the EM iterations, we employ just five Cochrane-Orcutt iterations instead of continuing until convergence. Despite the large amount of parameters involved (154), the algorithm performs remarkably well, as shown in Figure 2. The first EM iteration yields a massive increase in the log-likelihood function, while subsequent iterations also provide noticeable gains. As expected, though, after 200 iterations the improvements become minimal. For that reason, we switched to a scoring algorithm with line searches at that stage, which converged rather smoothly to parameter estimates reported in Tables 1 and 2, together with standard errors obtained on the basis of the analytical expressions for the information matrix in appendix B.

Table 3 contains the results of joint significance tests for the dynamic loading coefficients associated to the global (columns 1 and 2) and regional (columns 3 and 4) factors for each country. Those tests confirm that with the possible exception of Iceland, all countries in our sample are dynamically correlated. More importantly, they also show that some clusters of countries are more correlated with each other than what a single factor model would allow for, thereby confirming the need for a bifactor model. This is particularly noticeable for the Baltic countries, but it also affects Norway, Sweden and the UK among those countries which have never belonged to EMU.

From an empirical point of view, it is of substantive interest to look at the evolution and persistence of those latent factors. Unfortunately, it is well known that 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 standard Kalman fixed interval smoother in the time domain with exact initial conditions derived from the stationary distribution of the 33 state variables (2 for the common factor and each of the regional factors and 1 for each of the idiosyncratic ones; see appendix C for details).[8]

Smoothed versions of the global and regional factors are displayed in Figure 3. In panel (a) we plot the estimated global factor jointly with the unweighted average of inflation rates across countries in our sample, and the inflation rate of the European Union countries. For ease of comparison, we re-scale both the global factor and the equally weighted inflation average to have the same mean and variance as the European Union inflation. The smoothed global factor, which with an estimated autocorrelation of 0.97 is rather persistent, tracks fairly well these two measures over the sample. The main exception is the period 1999-2002, when the global factor is significantly higher than the inflation rate of the European Union countries. Such discrepancies are explained by two facts: (i) the European Union HICP is a consumption-weighed average of country-specific price indices, and (ii) there are differences between our sample of countries and the set of economies used to construct the European Union HICP.[9] Since 2002, the global factor generally trends downwards, in line with the other two measures. The other panels of Figure 3 plot the estimated regional factors, which are scaled so that their innovations have unit variance. Interestingly, the factor for the new entrants to the euro area is even more persistent than the global factor (its autocorrelation is 0.98). In contrast, we do not observe statistically significant persistence in the evolution of the other two regional factors. These results suggest that some of the new entrant economies share a regional factor which drives the medium term trends in inflation, while other regional factors have a predominant role at higher frequencies. We revisit this question below.

Given the estimated factors and factor loadings, we can compute the contributions of global, regional and idiosyncratic factors in driving the observed changes in prices across countries. Figure 4 plots the results for all the countries in our sample. The global factor clearly drives the downward trend in inflation for many countries, including Cyprus, Denmark, France, Italy, Poland, Slovakia and Spain, among others. We also observe a sizeable role for the regional factor

[8] The main difference between the Wiener-Kolmogorov filtered values, $\mathbf{x}^K_{t|\infty}$, and the Kalman filter smoothed values, $\mathbf{x}^K_{t|T}$, 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.

[9] Specifically, the weight of a country is its share of household final monetary consumption expenditure in the total. The European Union HICP is constructed as the weighed average of the original 12 countries until 2004, then it extends to 15 countries until 2006, 27 countries until 2013, and finally 28 countries until the end of the sample.

for Estonia, Latvia, and Lithuania. For these Baltic economies, inflation dramatically swings over the period 2005-2011. Conversely, the regional factor only plays a marginal role for the other new entrants (Cyprus, Malta, and Slovakia), which did not experience such swings over the same period. In this regard, it is worth noticing that the Baltic countries adopted the euro in the late part of the sample (Estonia in 2011, Latvia in 2014 and Lithuania in 2015), while the other entrants joined the euro area earlier (Cyprus and Malta in 2008, Slovakia in 2009). This evidence, although far from conclusive, suggests that EMU may have had a dampening effect on inflation fluctuations for all the new entrant countries.

We complement our time domain results by decomposing the spectral density of each country inflation series into the corresponding global, regional, and idiosyncratic components. Figure 5 show for each frequency the fraction of variance explained by each of those components. To aid in the interpretation of the results, we have added vertical lines at those frequencies which capture movements in the series at 2 and 1 years, and 6 and 3 months. As can be seen, the global factor explains an important fraction of variance across many economies, especially at lower frequencies. This result confirms the view that most countries experience a common downward trend in inflation. Nevertheless, we also observe that the global factor plays virtually no role in other economies such as Norway, Sweden, and United Kingdom, whose correlations are mostly driven by the third regional factor. This somewhat surprising result may be partly explained by the fact that energy and food components are by construction excluded from our analysis. The regional factor of new entrants affects particularly Estonia, Latvia, and Lithuania, which confirms our previous time domain findings. In contrast, regional factors do not seem to influence medium term trends for most other countries.

Finally, we conducted two robustness exercises. First, we considered a version of the model with just a global factor and no regional factors, which hardly surprisingly leads to a markedly worse fit. More importantly, we have also experimented with a subdivision of the core euro area region to single out those countries which experienced the most dramatic drops in interest rates prior to their accession to EMU. This is an important distinction to explore as there has been considerable debate on whether the conduct of monetary policy by the ECB since its inception has resulted in unwanted effects on those economies; see Estrada and Saurina (2014) for a discussion of the Spanish case. By looking at the evolution of real interest differentials between 1995 and 1999, we interestingly find that the additional group is composed by Portugal, Ireland, Italy, Greece and Spain (the so-called PIIGS). However, we find that a dynamic bifactor model with four regions, including two within the core euro area, does not lead to a substantial improvement in fit.

5 Conclusions

We generalise the frequency domain version of the EM algorithm for dynamic factor models in Fiorentini, Galesi and Sentana (2014) to bifactor models in which pervasive common factors are complemented by block factors. We explain how to efficiently exploit the sparsity of the loading matrix to reduce the computational burden so much that researchers can estimate such models by maximum likelihood with a large number of series from multiple regions. We find that the EM algorithm leads to substantial likelihood gains starting from arbitrary initial values. Unfortunately, it slows down considerably near the optimum. For that reason, we also derive convenient expressions for the frequency domain scores and information matrix that allow us to switch to the scoring method at that point.

In an empirical application we explore the ability of a bifactor model to capture inflation dynamics across European countries. Specifically, we apply our procedure to year-on-year core inflation rates for 25 European countries over the period 1999:1-2014:12. We estimate a model with a common factor and three regional factors: original euro area members, new entrants and others. Overall, our results suggest that a global factor drives the medium-long term trends of inflation across most European economies, which is consistent with the evidence in the previous literature. But we also find a persistent regional factor driving the inflation trends of the Baltic countries, which are new entrants to the euro area. In contrast, we find that the regional factors for most other countries affect mainly their short run movements.

References

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

Ciccarelli, M., and Mojon, B. (2010): "Global inflation", *Review of Economics and Statistics* 92, 524-535.

Chamberlain, G. and Rothschild, M. (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets", *Econometrica* 51, 1281-1304.

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

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): "Testing for GARCH effects: a one-sided approach", *Journal of Econometrics* 86, 97-127.

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.

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.

Estrada, A., Galí, J. and López-Salido, D. (2013): "Patterns of convergence and divergence in the euro area", *IMF Economic Review* 61, 601-630.

Estrada, A. and Saurina, J. (2014): "Spanish boom and bust: some lessons for macroprudential policy", mimeo, Bank of Spain.

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., Galesi, A. and Sentana, E. (2014): "A spectral EM algorithm for dynamic factor models", CEMFI Working Paper 1411.

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

Galí, J. (2002): "Monetary policy in the early years of EMU", in Buti, M. and Sapir, A. (eds) *Economic and Monetary Union and Economic Policy in Europe*, pp. 41-72, Edward Elgar, Cheltenham.

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*, pp. 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.

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.

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

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

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

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

Holzinger, K.J. and Swineford, S. (1937): “The bi-factor method”, *Psychometrika* 47, 41–54.

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

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

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

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

Moench, E., Ng, S. and Potter, S (2013): “Dynamic hierarchical factor models”, *Review of Economics and Statistics* 95, 1811-1817.

Mumtaz, H. and Surico, P. (2012): “Evolving international inflation dynamics: world and country-specific factors”, *Journal of the European Economic Association* 10, 716-734.

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.

Reise S.P. (2012): “The rediscovery of bifactor measurement models”, *Multivariate Behav-*

ioral Research 47, 667-696.

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.

Stock, J.H. and Watson, M. (2009): "The evolution of national and regional factors in U.S. housing construction", in *Volatility and time series econometrics: essays in honour of Robert F. Engle*, T.Bollerslev, J. Russell and M. Watson (eds.), Oxford University 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.

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 (15). Since

$$\begin{aligned} d\mathbf{G}_{\mathbf{yy}}(\lambda) &= [d\mathbf{C}(e^{-i\lambda})]\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda}) + \mathbf{C}(e^{-i\lambda})[d\mathbf{G}_{\mathbf{xx}}(\lambda)]\mathbf{C}'(e^{i\lambda}) \\ &\quad + \mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)[d\mathbf{C}'(e^{i\lambda})] + d\mathbf{G}_{\mathbf{uu}}(\lambda) \end{aligned}$$

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

$$\begin{aligned} dvec\,[\mathbf{G}_{\mathbf{yy}}(\lambda)] &= \left[\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right] dvec[\mathbf{C}(e^{-i\lambda})] \\ &\quad + \left[\mathbf{I}_N\otimes\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\right]\mathbf{K}_{N,R+1} dvec[\mathbf{C}(e^{i\lambda})] \\ &\quad + \left[\mathbf{C}(e^{i\lambda})\otimes\mathbf{C}(e^{-i\lambda})\right]\mathbf{E}_{R+1} dvecd\,[\mathbf{G}_{\mathbf{xx}}(\lambda)] + \mathbf{E}_N dvecd\,[\mathbf{G}_{\mathbf{uu}}(\lambda)] \\ &= \left[\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right] dvec[\mathbf{C}(e^{-i\lambda})] + \mathbf{K}_{NN}\left[\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right] dvec[\mathbf{C}(e^{i\lambda})] \\ &\quad + \left[\mathbf{C}(e^{i\lambda})\otimes\mathbf{C}(e^{-i\lambda})\right]\mathbf{E}_{R+1} dvecd\,[\mathbf{G}_{\mathbf{xx}}(\lambda)] + \mathbf{E}_N dvecd\,[\mathbf{G}_{\mathbf{uu}}(\lambda)], \end{aligned}$$

where

$$\begin{array}{c} \mathbf{E}'_m = (\mathbf{e}_{1m}\mathbf{e}'_{1m}|\ldots|\mathbf{e}_{mm}\mathbf{e}'_{mm}), \\ (\mathbf{e}_{1m}|\ldots|\mathbf{e}_{mm}) = \mathbf{I}_m, \end{array} \tag{A1}$$

is the unique $m^2\times m$ "diagonalisation" matrix that transforms $vec(\mathbf{A})$ into $vecd(\mathbf{A})$ as $vecd(\mathbf{A}) = \mathbf{E}'_m vec(\mathbf{A})$ and $\mathbf{K}_{mn}$ is the commutation matrix of orders m and n (see Magnus (1988)). Further, we can use (5) to express $dvec\,[\mathbf{C}(z)]$ in terms of its non-zero elements $d\mathbf{c}(z)$ by means of the following linear transformation

$$
\begin{pmatrix}
d\mathbf{c}_{1g}(z) \\ \vdots \\ d\mathbf{c}_{rg}(z) \\ \vdots \\ d\mathbf{c}_{Rg}(z) \\ d\mathbf{c}_{11}(z) \\ \vdots \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \vdots \\ d\mathbf{c}_{rr}(z) \\ \vdots \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \vdots \\ d\mathbf{c}_{RR}(z)
\end{pmatrix}
=
\begin{pmatrix}
\mathbf{I}_{N_1} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\
\vdots & \ddots & \vdots & \ddots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{I}_{N_r} & \dots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{0} & \ddots & \vdots & \ddots & \vdots & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{I}_{N_R} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_{N_1} & \dots & \vdots & \dots & \vdots \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{N_r} & \dots & \mathbf{0} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{I}_{N_R}
\end{pmatrix}
\begin{pmatrix}
d\mathbf{c}_{1g}(z) \\ \vdots \\ d\mathbf{c}_{rg}(z) \\ \vdots \\ d\mathbf{c}_{Rg}(z) \\ d\mathbf{c}_{11}(z) \\ \vdots \\ d\mathbf{c}_{rr}(z) \\ \vdots \\ d\mathbf{c}_{RR}(z)
\end{pmatrix}
$$

$$
= \left(\mathfrak{e}_{1g}, \dots, \mathfrak{e}_{rg}, \dots, \mathfrak{e}_{Rg}, \mathfrak{e}_{11}, \dots, \mathfrak{e}_{rr}, \dots, \mathfrak{e}_{RR}\right) d\mathbf{c}(z) = \mathfrak{E} d\mathbf{c}(z),
$$

where $\mathfrak{E}$ contains a block analogue to the diagonalisation matrix above. Consequently, the Jacobian of $vec\left[\mathbf{G_{yy}}(\lambda)\right]$ will be

$$
\begin{aligned}
\frac{\partial vec\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_x'} &= \left[\mathbf{C}(e^{i\lambda}) \otimes \mathbf{C}(e^{-i\lambda})\right] \mathbf{E}_{R+1} \frac{\partial vecd\left[\mathbf{G_{xx}}(\lambda)\right]}{\partial \boldsymbol{\theta}_x'} \\
\frac{\partial vec\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \boldsymbol{\psi}'} &= \mathbf{E}_N \frac{\partial vecd\left[\mathbf{G_{uu}}(\lambda)\right]}{\partial \boldsymbol{\psi}'} \\
\frac{\partial vec\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_{\mathbf{u}}'} &= \mathbf{E}_N \frac{\partial vecd\left[\mathbf{G_{uu}}(\lambda)\right]}{\partial \boldsymbol{\theta}_{\mathbf{u}}'} \\
\frac{\partial vec\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \mathbf{c}_{rgk}'} &= \left\{ \begin{array}{c} \left[e^{-ik\lambda}\mathbf{C}(e^{i\lambda})\mathbf{G_{xx}}(\lambda) \otimes \mathbf{I}_N\right] \\ +\mathbf{K}_{NN}\left[e^{ik\lambda}\mathbf{C}(e^{-i\lambda})\mathbf{G_{xx}}(\lambda) \otimes \mathbf{I}_N\right] \end{array} \right\} \mathfrak{e}_{rg} \\
\frac{\partial vec\left[\mathbf{G_{yy}}(\lambda)\right]}{\partial \mathbf{c}_{rrl}'} &= \left\{ \begin{array}{c} \left[e^{-il\lambda}\mathbf{C}(e^{i\lambda})\mathbf{G_{xx}}(\lambda) \otimes \mathbf{I}_N\right] \\ +\mathbf{K}_{NN}\left[e^{il\lambda}\mathbf{C}(e^{-i\lambda})\mathbf{G_{xx}}(\lambda) \otimes \mathbf{I}_N\right] \end{array} \right\} \mathfrak{e}_{rr}
\end{aligned}
$$

where we have used the fact that

$$\frac{\partial vec\left[\mathbf{C}(z)\right]}{\partial \mathbf{c}'_{rgk}} = \mathfrak{E}\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{I}_{N_r} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} z^k = \mathfrak{e}_{rg} z^k$$

and

$$\frac{\partial vec\left[\mathbf{C}(z)\right]}{\partial \mathbf{c}'_{rrl}} = \mathfrak{E}\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \vdots \\ \mathbf{I}_{N_r} \\ \vdots \\ \mathbf{0} \end{pmatrix} z^l = \mathfrak{e}_{rr} z^l$$

since

$$\begin{aligned} \frac{\partial \mathbf{c}_{rg}(z)}{\partial \mathbf{c}'_{rgk}} &= z^k \mathbf{I}_{N_r} \\ \frac{\partial \mathbf{c}_{rr}(z)}{\partial \mathbf{c}'_{rrl}} &= z^l \mathbf{I}_{N_r} \end{aligned}$$

in view of (2) and (3).

If we combine those expressions with the fact that

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

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}_{\boldsymbol{\theta}_{\mathbf{x}}}(\lambda;\boldsymbol{\theta}) &= \frac{\partial vecd'[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{x}}}\mathbf{E}'_{R+1}\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] \\
&= \frac{\partial vecd'[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{x}}} vecd\left[\begin{array}{c} 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}) \end{array}\right] \\
2\mathbf{d}_{\boldsymbol{\psi}}(\lambda;\boldsymbol{\theta}) &= \frac{\partial vecd'[\mathbf{G}_{\mathbf{uu}}(\lambda)]}{\partial\boldsymbol{\psi}} vecd\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] \\
2\mathbf{d}_{\boldsymbol{\theta}_{\mathbf{u}}}(\lambda;\boldsymbol{\theta}) &= \frac{\partial vecd'[\mathbf{G}_{\mathbf{uu}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{u}}} vecd\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] \\
2\mathbf{d}_{\mathbf{c}_{rgk}}(\lambda;\boldsymbol{\theta}) &= \mathfrak{e}'_{rg}\left\{\begin{array}{c} \left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})e^{-ik\lambda}\otimes\mathbf{I}_N\right] \\ +\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})e^{ik\lambda}\otimes\mathbf{I}_N\right]\mathbf{K}_{NN} \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] \\
&= \mathfrak{e}'_{rg}\left\{\begin{array}{c} e^{-ik\lambda}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})\mathbf{G}_{\mathbf{xx}}(\lambda) \\ -\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda) \end{array}\right] \\ +e^{ik\lambda}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})\mathbf{G}_{\mathbf{xx}}(\lambda) \\ -\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda) \end{array}\right] \end{array}\right\} \\
2\mathbf{d}_{\mathbf{c}_{rrl}}(\lambda;\boldsymbol{\theta}) &= \mathfrak{e}'_{rr}\left\{\begin{array}{c} \left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})e^{-il\lambda}\otimes\mathbf{I}_N\right] \\ +\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})e^{il\lambda}\otimes\mathbf{I}_N\right]\mathbf{K}_{NN} \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] \\
&= \mathfrak{e}'_{rr}\left\{\begin{array}{c} e^{-il\lambda}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})\mathbf{G}_{\mathbf{xx}}(\lambda) \\ -\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda) \end{array}\right] \\ +e^{il\lambda}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})\mathbf{G}_{\mathbf{xx}}(\lambda) \\ -\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda) \end{array}\right] \end{array}\right\},
\end{aligned}$$

where we have used the fact that $\mathbf{K}'_{NN} = \mathbf{K}_{NN} = \mathbf{K}^{-1}_{NN}$ (see again Magnus (1988)).

Let us now try to interpret the different components of this expression. To do so, it is convenient to further assume that $\mathbf{G}_{\mathbf{xx}}(\lambda) > 0$ and $\mathbf{G}_{\mathbf{uu}}(\lambda) > \mathbf{0}$.

The first thing to note is that

$$\begin{aligned}
&2\pi\mathbf{C}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{I}'_{\mathbf{yy}}(\lambda)\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{C}(e^{i\lambda}) - \mathbf{C}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\mathbf{C}(e^{i\lambda}) \\
&\quad = \mathbf{G}^{-1}_{\mathbf{xx}}(\lambda)\left[2\pi\mathbf{I}'_{\mathbf{x}^K\mathbf{x}^K}(\lambda) - \mathbf{G}'_{\mathbf{x}^K\mathbf{x}^K}(\lambda)\right]\mathbf{G}^{-1}_{\mathbf{xx}}(\lambda).
\end{aligned}$$

Given that

$$\frac{\partial vecd\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\right]}{\partial\boldsymbol{\theta}'_{x_g}} = \frac{\partial G_{x_g x_g}(\lambda)}{\partial\boldsymbol{\theta}'_{x_g}}\mathbf{e}_{1,R+1},$$

the component of the score associated to the parameters that determine $G_{x_g x_g}(\lambda)$ will be the cross-product across frequencies of the product of the derivatives of the spectral density of x_{gt} with the difference between the periodogram and spectrum of x_{gt}^K inversely weighted by the squared spectral density of x_{gt}. Thus, we can interpret this term as arising from a marginal log-likelihood function for x_{gt} that takes into account the unobservability of x_{gt}. Exactly the same comments apply to the scores of the parameters that determine $G_{x_r x_r}(\lambda)$ for $r = 1, \ldots, R$ in view of the fact that

$$\frac{\partial vecd\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\right]}{\partial\boldsymbol{\theta}'_{x_r}} = \frac{\partial G_{x_r x_r}(\lambda)}{\partial\boldsymbol{\theta}'_{x_r}}\mathbf{e}_{r+1,R+1}.$$

Similarly, given that

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

$$\frac{\partial vecd\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right]}{\partial\psi_i}=\frac{\partial G_{u_iu_i}(\lambda)}{\partial\psi_i}\mathbf{e}_{iN}$$

and

$$\frac{\partial vecd\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right]}{\partial\boldsymbol{\theta}'_{u_i}}=\frac{\partial G_{u_iu_i}(\lambda)}{\partial\boldsymbol{\theta}'_{u_i}}\mathbf{e}_{iN},$$

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^K_{it} 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 $\mathbf{x}_t$ that takes into account the unobservability of u_{t_i}.

Finally, to interpret the scores of the distributed lag coefficients it is worth noting that

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

and

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

are complex conjugates because $\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)$ is Hermitian and the conjugate of a product is the product of the conjugates, so it suffices to analyse one of them. On this basis, if we write

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

the components of the score associated to $\mathbf{c}_{rgk}$ and will be the sum across frequencies of terms of the form

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

(and their conjugate transposes), which capture the difference between the cross-periodogram and cross-spectrum of x^K_{gt-r} and u^K_{it} inversely weighted by the spectral density of u_{it}. Exactly the same comments apply to the scores of $\mathbf{c}_{rrl}$. Therefore, we can understand those terms as arising from the normal equation in the spectral regression of y_{it} onto $x_{g,t+m_g},\ldots x_{g,t-n_g}$ and $x_{r,t+m_r},\ldots x_{rt-n_r}$ but taking into account the unobservability of the regressors.

As usual, we can exploit the Woodbury formula, as in expressions (7), (9), (10), (25), (29) and (30), to greatly speed up the computations.

B Spectral information matrix

Given the expression for the Jacobian matrix in derived in appendix A, we will have that

$$\begin{aligned}
\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_x} &= \frac{\partial vecd'\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\right]}{\partial \boldsymbol{\theta}_x}\mathbf{E}'_{R+1}\left[\mathbf{C}'(e^{i\lambda})\otimes\mathbf{C}'(e^{-i\lambda})\right] \\
\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \boldsymbol{\psi}} &= \frac{\partial vecd'\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right]}{\partial \boldsymbol{\psi}'}\mathbf{E}'_{N} \\
\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_{\mathbf{u}}} &= \frac{\partial vecd'\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right]}{\partial \boldsymbol{\theta}_{\mathbf{u}}}\mathbf{E}'_{N} \\
\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \mathbf{c}_{rgk}} &= \mathfrak{e}'_{rg}\left\{\begin{array}{c}\left[e^{-ik\lambda}\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\otimes\mathbf{I}_N\right] \\ +\left[e^{ik\lambda}\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})\otimes\mathbf{I}_N\right]\mathbf{K}_{NN}\end{array}\right\} \\
\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \mathbf{c}_{rrl}} &= \mathfrak{e}'_{rr}\left\{\begin{array}{c}\left[e^{-il\lambda}\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\otimes\mathbf{I}_N\right] \\ +\left[e^{il\lambda}\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})\otimes\mathbf{I}_N\right]\mathbf{K}_{NN}\end{array}\right\}
\end{aligned}$$

and

$$\begin{aligned}
\left\{\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_x}\right\}^* &= \left[\mathbf{C}(e^{-i\lambda})\otimes\mathbf{C}(e^{i\lambda})\right]\mathbf{E}_{R+1}\frac{\partial vecd\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\right]}{\partial \boldsymbol{\theta}'_x} \\
\left\{\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \boldsymbol{\psi}}\right\}^* &= \mathbf{E}_N\frac{\partial vecd\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right]}{\partial \boldsymbol{\psi}'} \\
\left\{\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_{\mathbf{u}}}\right\}^* &= \mathbf{E}_N\frac{\partial vecd\left[\mathbf{G}_{\mathbf{uu}}(\lambda)\right]}{\partial \boldsymbol{\theta}'_{\mathbf{u}}} \\
\left\{\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \mathbf{c}_{rgk}}\right\}^* &= \left\{\begin{array}{c}\left[e^{ik\lambda}\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right] \\ +\mathbf{K}_{NN}\left[e^{-ik\lambda}\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\end{array}\right\}\mathfrak{e}_{rg} \\
\left\{\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \mathbf{c}_{rrl}}\right\}^* &= \left\{\begin{array}{c}\left[e^{il\lambda}\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right] \\ +\mathbf{K}_{NN}\left[e^{-il\lambda}\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\end{array}\right\}\mathfrak{e}_{rr}
\end{aligned}$$

Hence, it is straightforward to see that the elements of the block of the information matrix (18) corresponding to the autoregressive parameters for the common factors will be

$$\begin{aligned}
\mathbf{Q}_{\boldsymbol{\theta}_{\mathbf{x}}\boldsymbol{\theta}_{\mathbf{x}}}(\lambda;\boldsymbol{\theta}) &= \frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_x}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right]\left\{\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_x}\right\}^* \\
&= \frac{\partial vecd'[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial \boldsymbol{\theta}_x}\mathbf{E}'_{R+1}\left[\mathbf{C}'(e^{i\lambda})\otimes\mathbf{C}'(e^{-i\lambda})\right]\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right] \\
&\qquad\times\left[\mathbf{C}(e^{-i\lambda})\otimes\mathbf{C}(e^{i\lambda})\right]\mathbf{E}_{R+1}\frac{\partial vecd[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial \boldsymbol{\theta}'_x} \\
&= \frac{\partial vecd'[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial \boldsymbol{\theta}_x}\left\{\left[\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{C}(e^{-i\lambda})\right]\odot\left[\mathbf{C}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{C}(e^{i\lambda})\right]\right\}\frac{\partial vecd[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial \boldsymbol{\theta}'_x},
\end{aligned}$$

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

Similarly,

$$\begin{aligned}
\mathbf{Q}_{\boldsymbol{\theta}_{\mathbf{u}}\boldsymbol{\theta}_{\mathbf{u}}}(\lambda;\boldsymbol{\theta}) &= \frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_{\mathbf{u}}}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right]\left\{\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \boldsymbol{\theta}_{\mathbf{u}}}\right\}^* \\
&= \frac{\partial vecd'[\mathbf{G}_{\mathbf{uu}}(\lambda)]}{\partial \boldsymbol{\theta}_{\mathbf{u}}}\mathbf{E}'_N\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right]\mathbf{E}_N\frac{\partial vecd[\mathbf{G}_{\mathbf{uu}}(\lambda)]}{\partial \boldsymbol{\theta}'_{\mathbf{u}}} \\
&= \frac{\partial vecd'[\mathbf{G}_{\mathbf{uu}}(\lambda)]}{\partial \boldsymbol{\theta}_{\mathbf{u}}}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda)\odot\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right]\frac{\partial vecd[\mathbf{G}_{\mathbf{uu}}(\lambda)]}{\partial \boldsymbol{\theta}'_{\mathbf{u}}},
\end{aligned}$$

with an almost identical expression for $\mathbf{Q}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\lambda;\boldsymbol{\theta})$.

Also,

$$\mathbf{Q}_{\boldsymbol{c}_{rgk}\boldsymbol{c}_{rrl}}(\lambda;\boldsymbol{\theta}) = \frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{\partial \boldsymbol{c}_{rgk}}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\left\{\frac{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial \mathbf{c}_{rrl}}\right\}^*$$

$$= \mathfrak{e}_{rg}'\left\{\begin{array}{c}\left[e^{-ik\lambda}\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\otimes\mathbf{I}_N\right]\\ +\left[e^{ik\lambda}\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})\otimes\mathbf{I}_N\right]\mathbf{K}_{NN}\end{array}\right\}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]$$

$$\times\left\{\begin{array}{c}\left[e^{il\lambda}\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\\ +\mathbf{K}_{NN}\left[e^{-il\lambda}\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\end{array}\right\}\mathfrak{e}_{rr}$$

$$= \mathfrak{e}_{rg}'\left\{\begin{array}{c} e^{-i(k+l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\otimes\mathbf{I}_N\right]\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\mathbf{K}_{NN}\left[\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\\ e^{i(k+l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})\otimes\mathbf{I}_N\right]\mathbf{K}_{NN}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\left[\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\\ e^{-i(k-l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\otimes\mathbf{I}_N\right]\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\left[\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\\ e^{i(k-l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})\otimes\mathbf{I}_N\right]\mathbf{K}_{NN}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\mathbf{K}_{NN}\left[\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\end{array}\right\}\mathfrak{e}_{rr}$$

$$= \mathfrak{e}_{rg}'\left\{\begin{array}{c} e^{-i(k+l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\otimes\mathbf{I}_N\right]\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\left[\mathbf{I}_N\otimes\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\right]\mathbf{K}_{N,R+1}\\ e^{i(k+l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})\otimes\mathbf{I}_N\right]\left[\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\right]\left[\mathbf{I}_N\otimes\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\right]\mathbf{K}_{N,R+1}\\ e^{-i(k-l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\otimes\mathbf{I}_N\right]\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\left[\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\\ e^{i(k-l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})\otimes\mathbf{I}_N\right]\left[\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\right]\left[\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\end{array}\right\}\mathfrak{e}_{rr}$$

$$= \mathfrak{e}_{rg}'\left\{\begin{array}{c} e^{-i(k+l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\right]\mathbf{K}_{N,R+1}\\ e^{i(k+l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\right]\mathbf{K}_{N,R+1}\\ e^{-i(k-l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j)\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\\ e^{i(k-l)\lambda}\left[\mathbf{G}_{\mathbf{xx}}(\lambda)\mathbf{C}'(e^{-i\lambda})\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda_j)\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\right]\end{array}\right\}\mathfrak{e}_{rr},$$

where we have made use of the properties of the commutation matrix.

Further

$$\mathbf{Q}_{\boldsymbol{\theta}_{\mathbf{x}}\boldsymbol{\theta}_{\mathbf{u}}}(\lambda;\boldsymbol{\theta}) = \frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{\partial\boldsymbol{\theta}_x}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\frac{\partial vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial\boldsymbol{\theta}_{\mathbf{u}}'}$$

$$\frac{\partial vecd'[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial\boldsymbol{\theta}_x}\mathbf{E}_{R+1}'\left[\mathbf{C}'(e^{i\lambda})\otimes\mathbf{C}'(e^{-i\lambda})\right]\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\mathbf{E}_N\frac{\partial vecd[\mathbf{G}_{\mathbf{uu}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{u}}'}$$

$$= \frac{\partial vecd'[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial\boldsymbol{\theta}_x}\left[\mathbf{C}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\odot\mathbf{C}'(e^{-i\lambda})\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\frac{\partial vecd[\mathbf{G}_{\mathbf{uu}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{u}}'},$$

$$\mathbf{Q}_{\boldsymbol{\theta}_{\mathbf{x}}\mathbf{c}_{rrl}}(\lambda;\boldsymbol{\theta}) = \frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{x}}}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\frac{\partial vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial\mathbf{c}_{rrl}'}$$

$$= \frac{\partial vecd'[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{x}}}\mathbf{E}_{R+1}'\left[\mathbf{C}'(e^{i\lambda})\otimes\mathbf{C}'(e^{-i\lambda})\right]$$

$$\times\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\left\{\begin{array}{c}\left[e^{il\lambda}\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\\ +\mathbf{K}_{NN}\left[e^{-il\lambda}\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\end{array}\right\}\mathfrak{e}_{rr}$$

$$= \frac{\partial vecd'[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{x}}}\mathbf{E}_{R+1}'\left[\begin{array}{c} e^{il\lambda}\mathbf{C}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{C}'(e^{-i\lambda})\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\\ +e^{-il\lambda}\mathbf{C}'(e^{-i\lambda})\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{C}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\end{array}\right]\mathfrak{e}_{rr},$$

and

$$\mathbf{Q}_{\boldsymbol{\theta}_{\mathbf{u}}\mathbf{c}_{rrl}}(\lambda;\boldsymbol{\theta}) = \frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{u}}}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\frac{\partial vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]}{\partial\mathbf{c}_{rrl}'}$$

$$= \frac{\partial vecd'[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{u}}}\mathbf{E}_N'\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\right]\left\{\begin{array}{c}\left[e^{il\lambda}\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\\ +\mathbf{K}_{NN}\left[e^{-il\lambda}\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{I}_N\right]\end{array}\right\}\mathfrak{e}_{rr}$$

$$= \frac{\partial vecd'[\mathbf{G}_{\mathbf{xx}}(\lambda)]}{\partial\boldsymbol{\theta}_{\mathbf{u}}}\mathbf{E}_N'\left[\begin{array}{c} e^{il\lambda}\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\\ +e^{-il\lambda}\mathbf{G}_{\mathbf{yy}}^{\prime -1}(\lambda)\mathbf{C}(e^{i\lambda})\mathbf{G}_{\mathbf{xx}}(\lambda)\otimes\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\end{array}\right]\mathfrak{e}_{rr},$$

where we have used the properties of the diagonalisation and commutation matrices, and in particular, that $\mathbf{E}'_m\mathbf{K}_{mm} = \mathbf{E}'_m$. In fact, further simplification can be achieved by exploiting (A1). The formulae for the remaining elements are entirely analogous. In this regard, it is important to note that all the above expressions can be written as the sum of some matrix and its complex conjugate transpose, as one would expect given that the information matrix is real.

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

Given that

$$\begin{aligned}\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j) &= \left[\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda) - \mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\right], \\ \mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda_j) &= \left[\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda) - \mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{i\lambda})\mathbf{\Omega}'(\lambda)\mathbf{C}'(e^{-i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\right],\end{aligned}$$

we will have that

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

where we have used the fact that

$$\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{\Omega}(\lambda) = \mathbf{I}_{R+1} - \mathbf{G}^{-1}_{\mathbf{xx}}(\lambda)\mathbf{\Omega}(\lambda)$$

and

$$\mathbf{C}'(e^{-i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{i\lambda})\mathbf{\Omega}'(\lambda) = \mathbf{I}_{R+1} - \mathbf{G}^{-1}_{\mathbf{xx}}(\lambda)\mathbf{\Omega}'(\lambda).$$

As a result,and

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

and

$$\mathbf{G_{xx}}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{C}(e^{-i\lambda})\mathbf{G_{xx}}(\lambda) = \mathbf{\Omega}(\lambda)\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{-i\lambda})\mathbf{G_{xx}}(\lambda).$$

In addition, the special structure of $\mathbf{C}(z)$ in (5) can also be successfully exploited to speed up the calculations. In particular,

$$\mathbf{C}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{C}(e^{i\lambda}) = \mathbf{\Omega}^{-1}(\lambda) - \mathbf{G}^{-1}_{\mathbf{xx}}(\lambda),$$

where $\mathbf{\Omega}^{-1}(\lambda)$ has been defined in (11). Further speed gains can be achieved by noticing that

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

C State space representation of dynamic bifactor models with AR(1) factors

There are several ways of casting the dynamic factor model in (4) into state-space format, but the most straightforward one is to consider a state vector of dimension $2(R+1)+N$ in which the AR(1) processes for both global and regional factors are written as a bivariate VAR(1) in (x_t, x_{t-1}), and the N AR(1) processes for the specific factors are written as first order ARs in u_{it}. As a result, we can write the measurement equation without an error term as

$$\mathbf{y}_t = \mathbf{Z}\boldsymbol{\alpha}_t,$$

where the state vector is

$$\begin{aligned}
\boldsymbol{\alpha}_t &= (\mathbf{x}'_t, \mathbf{x}'_{t-1}, \mathbf{u}'_t)', \\
\mathbf{x}_t &= (x_{gt}, x_{1t}, \dots, x_{Rt})', \\
\mathbf{u}_t &= (u_{1t}, \dots, u_{it}, \dots, u_{Nt})',
\end{aligned}$$

and $\mathbf{Z}$ is the $N \times (N+2R+2)$ matrix

$$\mathbf{Z} = [\mathbf{C}_0|\mathbf{C}_1|\mathbf{I}_N],$$

with $\mathbf{C}_0, \mathbf{C}_1$being $N \times (R+1)$ sparse matrices of contemporaneous and lagged loadings.

Consequently, the transition equation is simply

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \mathbf{u}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\rho}_{\mathbf{x}} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{R+1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\rho}_{\mathbf{u}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \mathbf{u}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{f}_t \\ \mathbf{0} \\ \mathbf{v}_t \end{bmatrix},$$

with

$$\begin{aligned}
\boldsymbol{\rho}_{\mathbf{x}} &= diag(\rho_{x_g}, \rho_{x_1}, \dots, \rho_{x_R}), \\
\boldsymbol{\rho}_{\mathbf{u}} &= diag(\rho_{u_1}, \dots, \rho_{u_N}),
\end{aligned}$$

$$\begin{aligned}
Cov(\mathbf{f}_t) &= \mathbf{I}_{R+1}, \\
Cov(\mathbf{v}_t) &= \mathbf{\Psi} = diag(\psi_1, \dots, \psi_N).
\end{aligned}$$

Given our covariance stationarity conditions, the initial condition for the state variables will trivially be $\boldsymbol{\alpha}_{1|0} = \mathbf{0}_{(N+2R+2)x1}$, and

$$\mathbf{P}_{1|0} = \begin{bmatrix} \mathbf{Q_{x}}_0 & \mathbf{Q_{x}}_1 & \mathbf{0} \\ \mathbf{Q_{x}}_1 & \mathbf{Q_{x}}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q_{u}}_0 \end{bmatrix},$$

where $\mathbf{Q_{x}}_0$ and $\mathbf{Q_{u}}_0$ are diagonal matrices with the unconditional variance of the corresponding AR(1) processes along the main diagonal, while $\mathbf{Q_{x1}}$ is also diagonal with the first autocovariance of the global and regional factors AR(1) processes on the main diagonal.

Table 1: Dynamic Loadings Estimates

Country	$c_{gi,0}$	std.err.	$c_{gi,1}$	std.err.	$c_{ri,0}$	std.err.	$c_{ri,1}$	std.err.
Core euro area								
Austria	-0.024	(0.017)	0.021	(0.017)	-0.058	(0.018)	0.021	(0.019)
Belgium	0.041	(0.021)	0.000	(0.021)	-0.170	(0.026)	0.000	(0.033)
Finland	-0.001	(0.016)	0.054	(0.016)	-0.043	(0.016)	0.054	(0.016)
France	0.041	(0.012)	0.011	(0.012)	0.019	(0.011)	0.011	(0.012)
Germany	-0.001	(0.018)	0.013	(0.018)	-0.006	(0.020)	0.013	(0.020)
Greece	0.357	(0.039)	-0.070	(0.039)	0.083	(0.036)	-0.070	(0.036)
Ireland	0.160	(0.023)	0.022	(0.023)	0.049	(0.022)	0.022	(0.022)
Italy	0.117	(0.017)	-0.001	(0.017)	0.047	(0.021)	-0.001	(0.021)
Luxembourg	-0.153	(0.019)	0.206	(0.020)	0.044	(0.020)	0.206	(0.020)
Netherlands	0.093	(0.019)	-0.005	(0.019)	-0.065	(0.019)	-0.005	(0.019)
Portugal	0.185	(0.026)	0.021	(0.026)	0.014	(0.026)	0.021	(0.026)
Spain	0.187	(0.023)	0.007	(0.023)	0.036	(0.023)	0.007	(0.023)
New entrants euro area								
Cyprus	0.286	(0.036)	-0.145	(0.036)	-0.063	(0.047)	-0.145	(0.047)
Estonia	0.269	(0.031)	-0.033	(0.030)	0.117	(0.049)	-0.033	(0.046)
Latvia	0.148	(0.037)	0.086	(0.037)	0.215	(0.076)	0.086	(0.087)
Lithuania	0.162	(0.034)	0.013	(0.033)	0.166	(0.059)	0.013	(0.057)
Malta	0.148	(0.036)	-0.015	(0.036)	0.019	(0.050)	-0.015	(0.050)
Slovakia	0.390	(0.035)	0.000	(0.035)	-0.022	(0.042)	0.000	(0.041)
Outside euro area								
Bulgaria	0.472	(0.060)	-0.098	(0.060)	0.036	(0.065)	-0.098	(0.064)
Denmark	0.077	(0.015)	0.028	(0.015)	0.035	(0.018)	0.028	(0.018)
Iceland	0.078	(0.065)	0.063	(0.065)	0.038	(0.074)	0.063	(0.073)
Norway	-0.006	(0.021)	-0.006	(0.021)	-0.046	(0.031)	-0.006	(0.027)
Poland	0.546	(0.043)	-0.149	(0.043)	-0.005	(0.044)	-0.149	(0.042)
Sweden	-0.019	(0.017)	0.025	(0.017)	0.007	(0.025)	0.025	(0.021)
United Kingdom	0.026	(0.016)	-0.019	(0.015)	0.038	(0.027)	-0.019	(0.021)

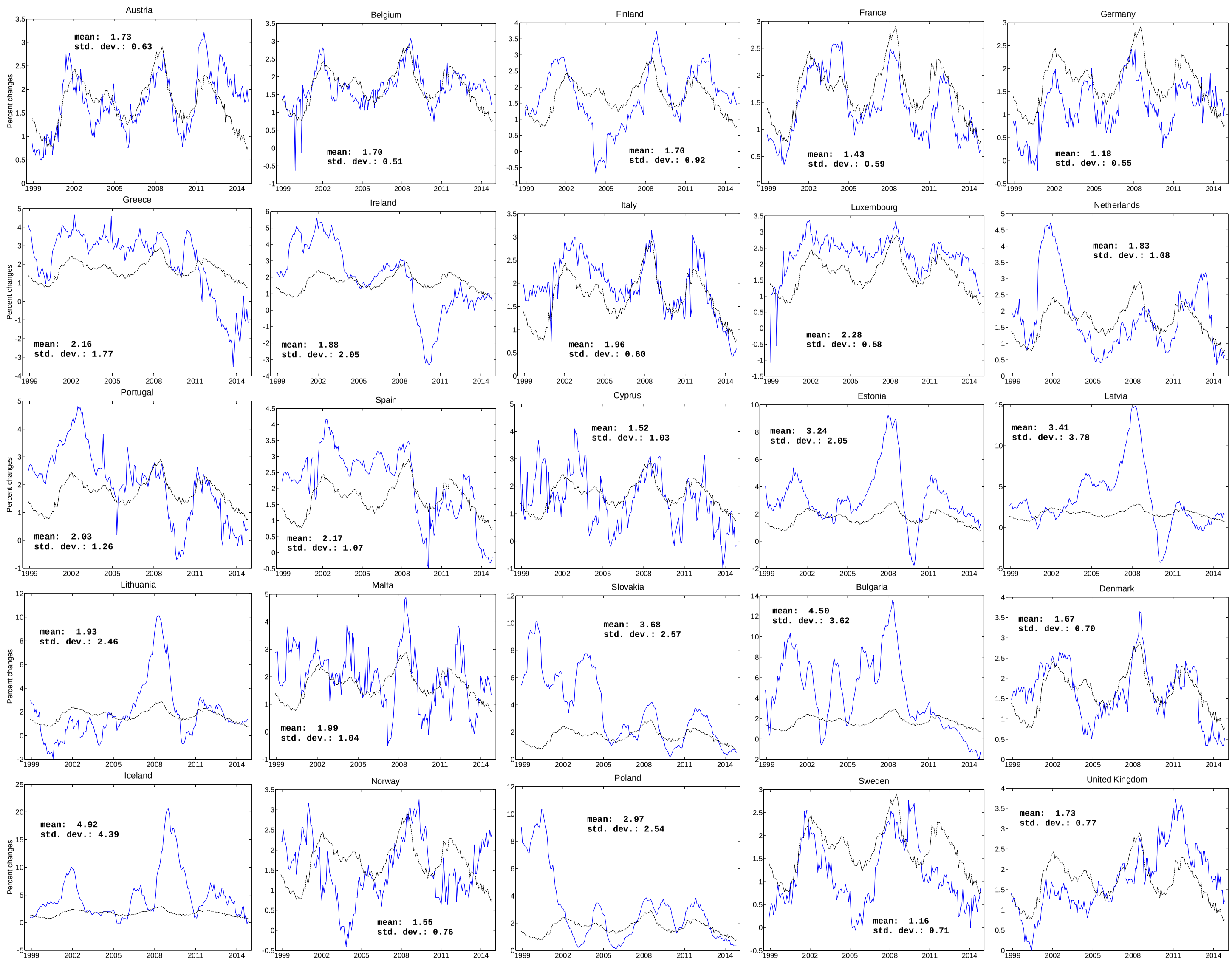
Table 2: Autoregressive Coefficients Estimates

Country	α	std.err.	ψ	std.err.
Global	0.9736	(0.017)	1.000	
Core euro area	0.2810	(0.207)	1.000	
New entrants euro area	0.9828	(0.013)	1.000	
Outside euro area	-0.1392	(0.302)	1.000	
Core euro area				
Austria	0.936	(0.025)	0.049	(0.005)
Belgium	0.912	(0.033)	0.033	(0.007)
Finland	0.974	(0.016)	0.041	(0.004)
France	0.948	(0.023)	0.022	(0.002)
Germany	0.887	(0.033)	0.063	(0.006)
Greece	0.941	(0.025)	0.194	(0.022)
Ireland	0.983	(0.011)	0.079	(0.009)
Italy	0.663	(0.071)	0.051	(0.006)
Luxembourg	0.852	(0.039)	0.049	(0.006)
Netherlands	0.970	(0.017)	0.055	(0.006)
Portugal	0.898	(0.034)	0.107	(0.011)
Spain	0.899	(0.035)	0.080	(0.009)
New entrants euro area				
Cyprus	0.805	(0.046)	0.213	(0.024)
Estonia	0.956	(0.028)	0.106	(0.013)
Latvia	0.977	(0.024)	0.113	(0.027)
Lithuania	0.960	(0.026)	0.147	(0.018)
Malta	0.799	(0.045)	0.268	(0.028)
Slovakia	0.981	(0.013)	0.135	(0.016)
Outside euro area				
Bulgaria	0.968	(0.018)	0.505	(0.055)
Denmark	0.918	(0.030)	0.036	(0.004)
Iceland	0.980	(0.013)	0.705	(0.072)
Norway	0.940	(0.025)	0.066	(0.009)
Poland	0.986	(0.010)	0.171	(0.023)
Sweden	0.953	(0.022)	0.044	(0.005)
United Kingdom	0.973	(0.016)	0.032	(0.004)

Table 3: Significance of Dynamic Loadings

Country	$H_0 : c_{gi,0} = c_{gi,1} = 0$ Wald test	p-value	$H_0 : c_{ri,0} = c_{ri,1} = 0$ Wald test	p-value
Core euro area				
Austria	3.07	(0.216)	15.44	(0.000)
Belgium	5.38	(0.068)	56.38	(0.000)
Finland	11.26	(0.004)	7.92	(0.019)
France	13.88	(0.001)	4.29	(0.117)
Germany	0.55	(0.760)	5.83	(0.054)
Greece	86.60	(0.000)	5.99	(0.050)
Ireland	47.22	(0.000)	6.40	(0.041)
Italy	61.32	(0.000)	12.23	(0.002)
Luxembourg	119.75	(0.000)	6.42	(0.041)
Netherlands	23.51	(0.000)	16.88	(0.000)
Portugal	53.15	(0.000)	0.42	(0.812)
Spain	65.92	(0.000)	5.68	(0.058)
New entrants euro area				
Cyprus	64.54	(0.000)	2.21	(0.330)
Estonia	78.72	(0.000)	25.96	(0.000)
Latvia	17.35	(0.000)	66.20	(0.000)
Lithuania	22.60	(0.000)	30.37	(0.000)
Malta	19.21	(0.000)	0.40	(0.817)
Slovakia	125.00	(0.000)	0.47	(0.790)
Outside euro area				
Bulgaria	64.18	(0.000)	0.88	(0.644)
Denmark	30.05	(0.000)	5.75	(0.057)
Iceland	2.36	(0.308)	0.68	(0.710)
Norway	0.18	(0.915)	13.52	(0.001)
Poland	164.30	(0.000)	2.51	(0.285)
Sweden	3.18	(0.204)	8.32	(0.016)
United Kingdom	3.78	(0.151)	11.84	(0.003)

Figure 1: European Inflation Rates



Notes: Inflation series are HICP excluding energy and unprocessed food. Dashed black line refers to HICP Inflation of European Union (EU12 until 2004, EU15 until 2006, EU27 until 2013, then EU28). Mean and standard deviations refer to country-specific series. Mean and standard deviation for European Union inflation are 1.69 and 0.50, respectively.

Figure 2: EM Algorithm Log-likelihood Evolution

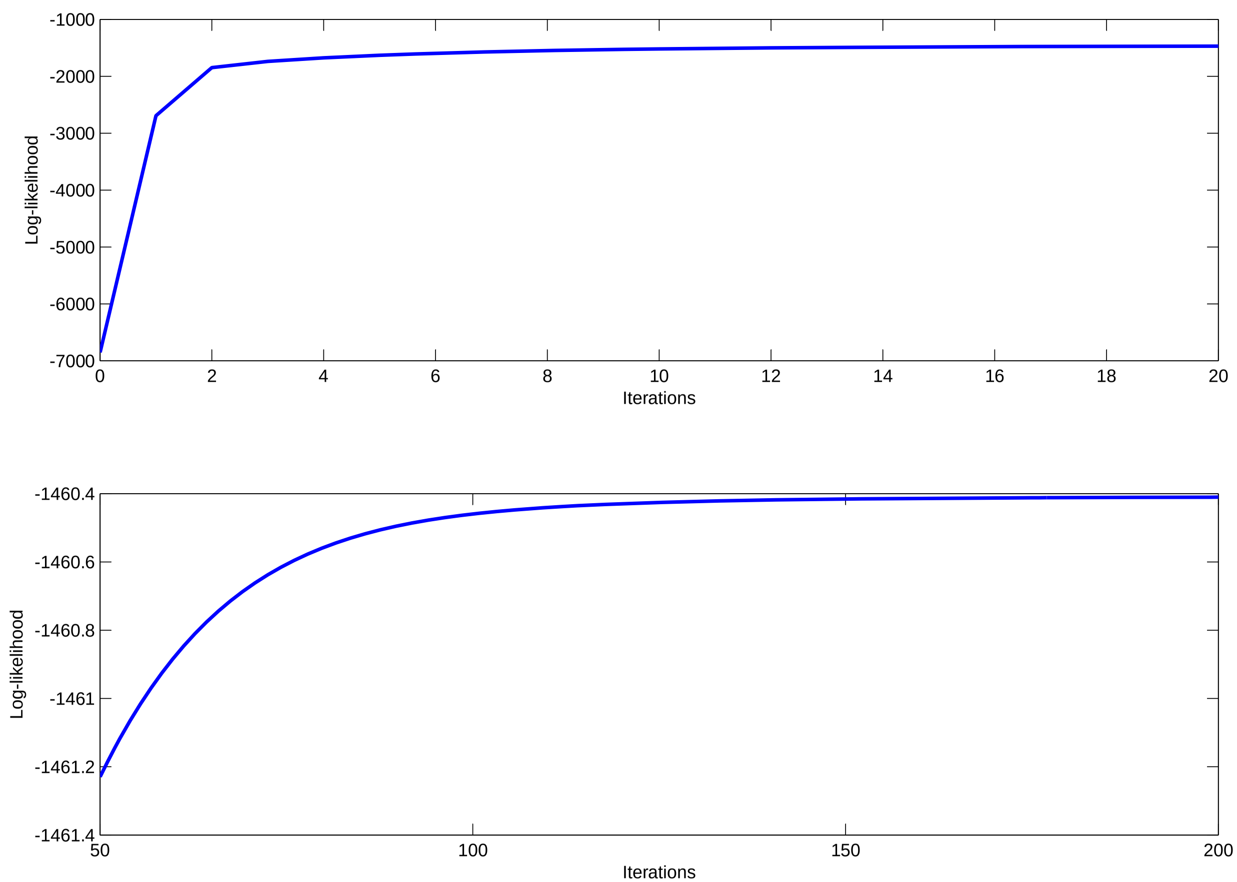
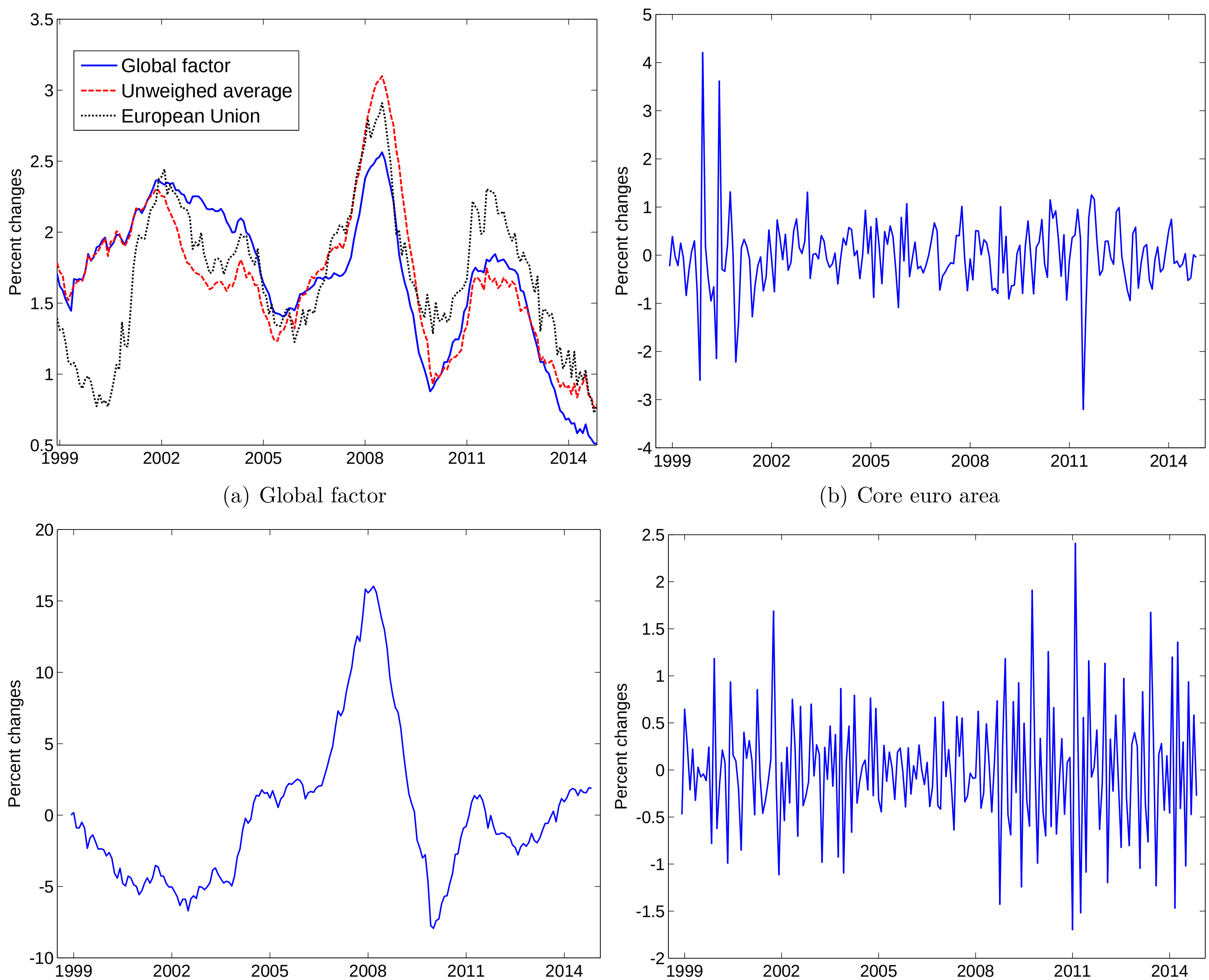


Figure 3: Smoothed Inflation Factors



(a) Global factor

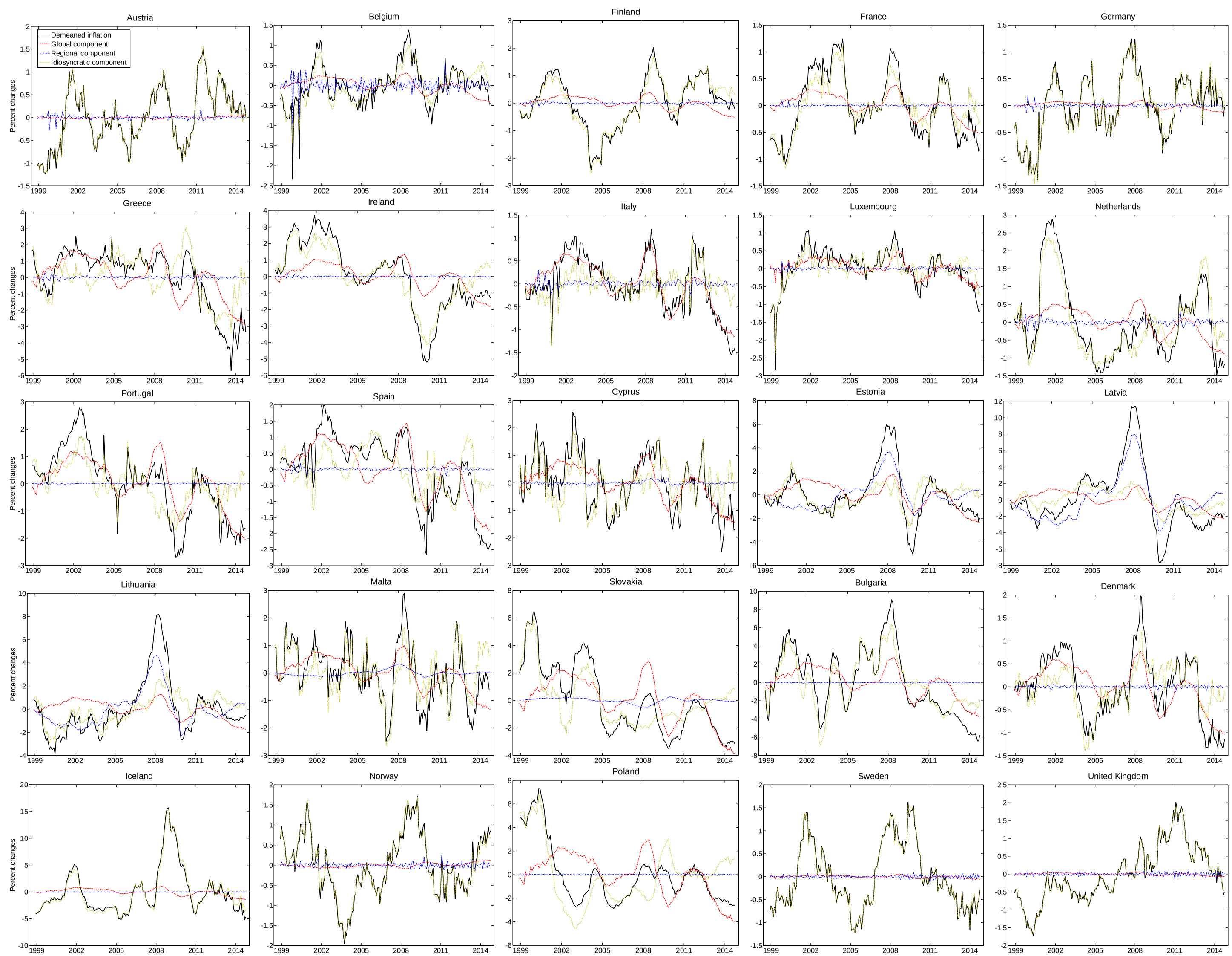
(b) Core euro area

(c) New entrants euro area

(d) Outside euro area

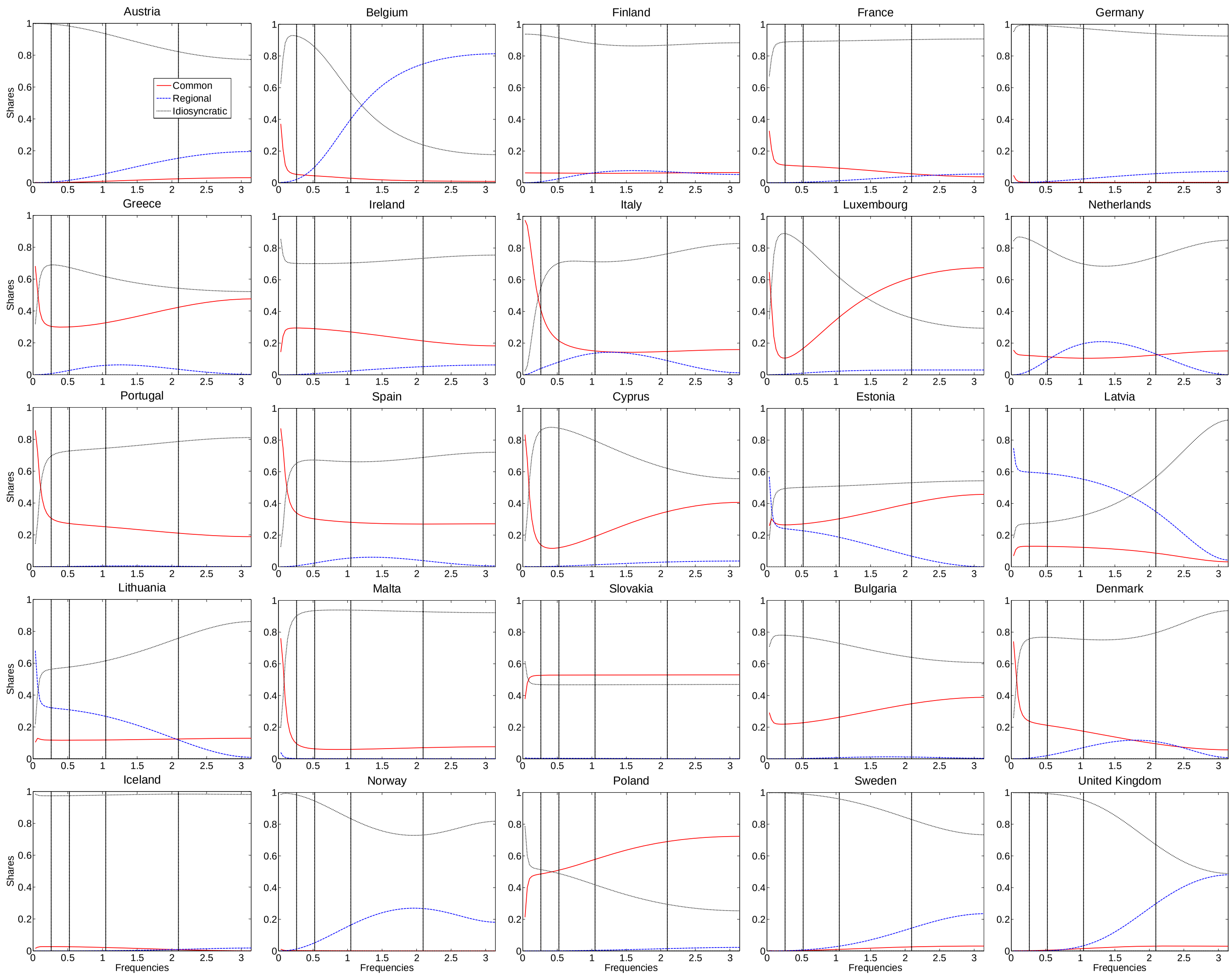
Notes: The series Global factor and Unweighed average are re-scaled to have same mean and variance as the European Union inflation. Regional factors are re-scaled so that their innovations have unit variance.

Figure 4: Contributions of Global, Regional, and Idiosyncratic Factors to Observed HICP Inflation



Notes: Inflation series are HICP excluding energy and unprocessed food.

Figure 5: Spectral Decompositions



Notes: The vertical lines correspond to those frequencies which reflect movements in the series at cycles of 2 and 1 years, and 6 and 3 months.