Dynamic specification tests for dynamic factor models*

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

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

June 2013
Revised: October 2018

Abstract

We derive computationally simple expressions for score tests of misspecification in parametric dynamic factor models using frequency domain techniques. We interpret those diagnostics as time domain moment tests which assess whether certain autocovariances of the smoothed latent variables match their theoretical values under the null of correct model specification. We also reinterpret reduced-form residual tests as checking specific restrictions on structural parameters. Our Gaussian tests are robust to nonnormal, independent innovations. Monte Carlo exercises confirm the finite sample reliability and power of our proposals. Finally, we illustrate their empirical usefulness in an application that constructs a US coincident indicator.

Keywords: Kalman filter, LM tests, Spectral maximum likelihood, Wiener-Kolmogorov filter.

JEL: C32, C38, C52, C12, C13

*This paper is a substantially revised version of Fiorentini and Sentana (2013), which in turn was based on the misspecification analysis of static factor models in Fiorentini and Sentana (2009). We are grateful to Tincho Almuzara, Dante Amengual, William Dunsmuir and Andrew Harvey, as well as to seminar audiences at Bologna, Cambridge, Columbia, ECARES/ULB, EUI, Georgetown, GRIPS, Penn, Princeton, Salento, Toulouse, the Oxford-Man Institute Time Series Econometrics Conference in honour of Andrew Harvey, the fifth ICEEE Congress (Genova), the first Barcelona GSE Summer Forum, ESEM (Gothenburg), the Lugano Large-Scale Factor Models in Finance Conference, and two RCEA Conferences (Toronto and Rimini) for helpful comments, discussions and suggestions. We are also grateful to Máximo Camacho and Gabriel Pérez-Quirós for allowing us to use their data and assisting us in reproducing their results. An anonymous referee has provided very useful feedback. Of course, the usual caveat applies. Financial support from MIUR through the project "Multivariate statistical models for risk assessment" (Fiorentini) and the Spanish Ministries of Science & Innovation and Economy & Competitiveness through grants ECO 2011-26342 and ECO 2014-59262 (Sentana) is also gratefully acknowledged.

1 Introduction

Dynamic factor models have been extensively used in macroeconomics and finance since their introduction by Geweke (1977) and Sargent and Sims (1977) as a way of capturing the cross-sectional and dynamic correlations between multiple series in a parsimonious way. A far from comprehensive list of early and more recent applications include not only traditional topics such as business cycle analysis (see Litterman and Sargent (1979), Stock and Watson (1989, 1991, 1993), Diebold and Rudebusch (1996), Gregory et al (1997), Mariano and Murosawa (2003), Aruoba et al (2009), Bańbura and Rünstler (2011)) and bond yields (Singleton (1981), Jegadeesh and Pennacchi (1996), Dungey et al (2000) or Diebold et al (2006)), but also investment (Sargent (1989)), wages (Engle and Watson (1981)), employment (Quah and Sargent (1993)), commodity prices (Peña and Box (1987)) house prices (Del Negro and Otrok (2007)), financial contagion (Mody and Taylor (2007)) or mortality rates (French and O'Hare (2013)).

Like its static counterpart, dynamic factor analysis can be either *exploratory* or *confirmatory* in nature. The goal of exploratory analysis is to maximise the fraction of covariance explained by the model without using any guidance on the economic interpretation of the common factors. In confirmatory analysis, in contrast, researchers often rely on economic and finance theory or previous studies to specify a priori the number of factors, their correlation structure and the dynamic impact they have on the observable variables. For that reason, Geweke and Singleton (1981) forcefully argued that in dynamic confirmatory factor models for economic time series "hypotheses about the relations between the observable time series and the latent factors and about the properties of the latent factors themselves may be tested". The purpose of our paper is precisely to derive such specification tests. In particular, we provide diagnostics for neglected serial correlation in common and specific factors, as well as for misspecification of the dynamic impact the former have on the observed variables. We focus on Lagrange Multiplier (LM) tests, which only require estimation of the model under the null. As is well known, Likelihood ratio (LR), Wald and LM tests are asymptotically equivalent under the null and sequences of local alternatives, and therefore they share their optimality properties. In addition to computational considerations, which are particularly relevant when one is concerned about several alternatives, an important advantage of LM tests is that they are often easy to interpret as moment tests, so that rejections provide a clear indication of the directions along which modelling efforts should focus. We emphasise this moment test interpretation for all our proposals.

Earlier work on specification testing in models with a finite number of series include Engle and Watson (1980), who explained how to apply the LM testing principle in the time domain for dynamic factor models with static factor loadings, Geweke and Singleton (1981), who studied LR

and Wald tests in the frequency domain, and Fernández (1990), who applied the LM principle in the frequency domain to a multivariate "structural time series model" (see Harvey (1989) for a comparison of time domain and frequency domain testing methods in that context).

Aside from considering a more general class of models, our main contribution is that our proposed tests are very simple to implement with a few lines of code, and even simpler to interpret. Once the parameters of a dynamic factor model have been estimated, its correct specification becomes the hypothesis of interest. Individual or joint score tests focusing on several departures from this null can then be computed from the auto- or cross-covariances of the smoothed values of the innovations in the state variables regardless of the complexity of the model estimated under the null. In this regard, our model validation proposal is entirely analogous to the battery of LM-based regression diagnostics that accompany the usual OLS output in most econometric packages. And even though our theoretical derivations make extensive use of spectral methods for time series, we provide both time domain and frequency domain interpretations of the relevant scores, so researchers who strongly prefer one method over the other could apply them without abandoning their favourite estimation techniques.

We also explicitly relate our proposals to alternative tests based on one-period-ahead prediction errors, which should be white noise under correct dynamic specification. In particular, we express those reduced form tests in terms of homogeneous restrictions on the dynamic factor loadings and idiosyncratic components, which allows us to make them robust to parameter uncertainty and study their relative power.

To keep the notation to a minimum, we focus on single factor models throughout, which suffice to illustrate our main results, although extensions to multiple factors could be easily entertained, as in Fiorentini and Sentana (2009). And even though we initially focus on Gaussian factor models with a diagonal idiosyncratic dynamic covariance structure for pedagogical reasons, we relax both these assumptions later on.

The rest of the paper is organised as follows. In section 2, we review the properties of dynamic factor models and their filters. Then, we derive our testing procedures in section 3 under an increasingly weaker set of assumptions, and present a Monte Carlo evaluation of their finite sample behaviour in section 4. This is followed in section 5 by an empirical illustration that assesses the dynamic factor model used by Camacho et al (2015) to construct a coincident indicator for the US. Finally, our conclusions, together with several interesting extensions, can be found in section 6. Auxiliary results are gathered in appendices.

2 Theoretical background

2.1 Dynamic factor models

A dynamic, exact, single factor model for a finite dimensional vector of N observed series, $\mathbf{y}_t$, can be defined in the time domain by the system of stochastic difference equations

$$\left.\begin{array}{c} \mathbf{y}_t = \boldsymbol{\mu} + \mathbf{c}(L)x_t + \mathbf{u}_t, \quad \alpha_x(L)x_t = \beta_x(L)f_t, \quad \mathbf{A}(L)\mathbf{u}_t = \mathbf{B}(L)\mathbf{v}_t, \\ \mathbf{A}(L) = diag[\alpha_{u_1}(L), \ldots, \alpha_{u_N}(L)], \quad \mathbf{B}(L) = diag[\beta_{u_1}(L), \ldots, \beta_{u_N}(L)], \\ (f_t, v_{1,t}, \ldots, v_{N,t})|I_{t-1}; \boldsymbol{\mu}, \boldsymbol{\phi} \sim N[0, diag(\gamma_f, \gamma_{v_1}, \ldots, \gamma_{v_N})], \end{array}\right\} \tag{1}$$

where x_t is the common factor, $\mathbf{u}_t$ the N specific factors, $\mathbf{c}(L) = \sum_{k=-m}^{n} \mathbf{c}_k L^k$ a vector of N possibly two-sided polynomials in the lag operator $c_i(L)$ (with m or n potentially unbounded), $\alpha_x(L)$ and $\alpha_{u_i}(L)$ are one-sided polynomials of finite orders p_x and p_{u_i}, respectively, while $\beta_x(L)$ and $\beta_{u_i}(L)$ are one-sided polynomials of finite orders q_x and q_{u_i} coprime with $\alpha_x(L)$ and $\alpha_{u_i}(L)$, respectively, I_{t-1} is an information set that contains the values of $\mathbf{v}_t$ and f_t up to, and including time $t-1$, $\boldsymbol{\mu}$ is the unconditional mean vector and $\boldsymbol{\phi}$ refers to all the d remaining static and dynamic second moment parameters, which we assume variation-free.

The dynamic nature of model (1) is the result of three different characteristics:

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

To some extent, characteristics 1 and 3 overlap, as one could always write any dynamic factor model in terms of white noise common factors as follows:

$$\mathbf{y}_t = \boldsymbol{\mu} + \alpha_x^{-1}(L)\beta_x(L)\mathbf{c}(L)f_t + \mathbf{A}^{-1}(L)\mathbf{B}(L)\mathbf{v}_t = \boldsymbol{\mu} + \boldsymbol{\Delta}(L)\boldsymbol{\xi}_t, \tag{2}$$

$$\boldsymbol{\Delta}(L) = \begin{bmatrix} c_1(L)\beta_x(L)\alpha_x^{-1}(L) & \beta_{u_1}(L)\alpha_{u_1}^{-1}(L) & 0 & \cdots & 0 \\ c_2(L)\beta_x(L)\alpha_x^{-1}(L) & 0 & \beta_{u_2}(L)\alpha_{u_2}^{-1}(L) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_N(L)\beta_x(L)\alpha_x^{-1}(L) & 0 & 0 & \cdots & \beta_{u_N}(L)\alpha_{u_N}^{-1}(L) \end{bmatrix}, \tag{3}$$

$$\boldsymbol{\xi}_t|I_{t-1}; \boldsymbol{\mu}, \boldsymbol{\phi} \sim N(\mathbf{0}, \boldsymbol{\Gamma}), \quad \boldsymbol{\xi}_t' = (f_t, v_{1,t}, \ldots, v_{N,t}), \quad \boldsymbol{\Gamma} = diag(\gamma_f, \gamma_{v_1}, \ldots, \gamma_{v_N}).$$

Thus, the assumption of ARMA(p_x, q_x) dynamics for the common factor can be regarded as a parsimonious way of modelling an infinite distributed lag for the dynamic factor loadings (see sections 3.1 and 3.6 for further details). In any case, we would need to shut down all three sources to go back to a traditional static factor model (see Lawley and Maxwell (1971)). Cancelling only one or two of those channels still results in a dynamic factor model. For example, Engle and Watson (1981) considered models with static factor loadings, while Peña and Box (1987) further assumed that the specific factors were white noise.

Nevertheless, when both m and n are finite, the dynamic factor model (1) can be written as a "static" factor model with $m+n+1$ common factors $(x_{t+m},\ldots,x_{t-n})$, a representation that has been exploited in the recent literature to justify the use of inefficient estimation methods for static factor models (see Bai and Ng (2008) or Stock and Watson (2011)). As we shall see in section 3.6, though, our proposed testing procedures can also deal with models with no static factor representation because the dynamic loadings $c_i(L)$ are rational functions.

The main difference between the exact model in (1) and the generalised dynamic factor models considered by Forni, et al (2000), Forni and Lippi (2001, 2011) and Forni et al (2015) is that it rules out any contemporaneous or dynamic cross-correlation between the idiosyncratic terms. We revisit this issue in section 3.10. and our concluding remarks.

2.2 Reduced form representation

To obtain the reduced form of a dynamic factor model, we can pre-multiplying both sides of the first equation in (1) by $\alpha_x(L)\mathbf{A}(L)$, which yields

$$\alpha_x(L)\mathbf{A}(L)(\mathbf{y}_t-\boldsymbol{\mu})=\mathbf{A}(L)\mathbf{c}(L)\beta_x(L)f_t+\alpha_x(L)\mathbf{B}(L)\mathbf{v}_t=\mathbf{m}_t. \tag{4}$$

The left hand side corresponds to a diagonal VAR, while the right hand side to a restricted VMA with the following dynamic factor structure: an MA common factor $\beta_x(L)f_t$, dynamic loadings $\mathbf{A}(L)\mathbf{c}(L)$, and MA specific factors $\alpha_x(L)\mathbf{B}(L)\mathbf{v}_t$.[1] Finding the Wold representation of $\mathbf{m}_t$, i.e.

$$\mathbf{m}_t=(\mathbf{I}_N+\mathbf{D}_1L+\ldots+\mathbf{D}_sL^s)\mathbf{w}_t=\mathbf{D}(L)\mathbf{w}_t, \tag{5}$$

with $\mathbf{w}_t|\mathbf{y}_{t-1},\mathbf{y}_{t-2},\ldots\sim N(\mathbf{0},\boldsymbol{\Sigma})$ and the roots of $|\mathbf{D}(L)|=1$ on or outside the unit circle, is by no means an easy task. The following multivariate version of the AR(p) signal plus white noise (WN) studied by Dunsmuir (1979) and Dzhaparidze (1986) among others:

$$\mathbf{y}_t=\boldsymbol{\mu}+\mathbf{c}x_t+\mathbf{u}_t,\ (1-\alpha_1L-\ldots-\alpha_pL^p)x_t=f_t,\quad \mathbf{u}_t=\mathbf{v}_t, \tag{6}$$

is a notable exception. Equation (4) implies that the Wold decomposition of $\mathbf{y}_t$ corresponds to a restricted VARMA(p,p) process with scalar autoregressive polynomial and a dynamic single factor model for the VMA(p) part. More importantly, we show in appendix B that both the reduced form innovation matrix $\boldsymbol{\Sigma}$ and all the Wold coefficient matrices $\mathbf{D}_j$ $(j=1,\ldots,p)$ are overidentified, with a common exact single factor structure whose factor loadings and idiosyncratic variances are proportional to $\mathbf{c}$ and $\boldsymbol{\Gamma}_u$, respectively. Therefore, the reduced form will be dynamically misspecified when any of three dynamic characteristics of model (1) is misspecified.

[1] As usual, cancellation between the VAR and VMA polynomials can occur for some parameter configurations; e.g. when the AR polynomials of the common factor and all the specific ones share some common roots.

2.3 Spectral density matrix and Wiener-Kolmogorov filter

Under the assumption that $\mathbf{y}_t$ in (1) is a covariance stationary process, possibly after suitable differencing or cointegration transformations, as in section 5, its spectral decomposition will be

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

with a spectral density matrix given by

$$\mathbf{G}_{\mathbf{yy}}(\lambda) = \mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda}) + \mathbf{G}_{\mathbf{uu}}(\lambda), \tag{7}$$

$$G_{xx}(\lambda) = \frac{\beta_x(e^{-i\lambda})\beta_x(e^{i\lambda})}{\alpha_x(e^{-i\lambda})\alpha_x(e^{i\lambda})}\gamma_f,$$

$$\mathbf{G}_{\mathbf{uu}}(\lambda) = \mathbf{A}^{-1}(e^{-i\lambda})\mathbf{B}(e^{-i\lambda})\boldsymbol{\Gamma}_u\mathbf{B}'(e^{i\lambda})\mathbf{A}^{-1\prime}(e^{i\lambda}) = diag[G_{u_1u_1}(\lambda), \ldots, G_{u_Nu_N}(\lambda)],$$

$$G_{u_iu_i}(\lambda) = \frac{\beta_{u_i}(e^{-i\lambda})\beta_{u_i}(e^{i\lambda})}{\alpha_{u_i}(e^{-i\lambda})\alpha_{u_i}(e^{i\lambda})}\gamma_{v_i}.$$

Thus, $\mathbf{G}_{\mathbf{yy}}(\lambda)$ is the sum of the rank 1 matrix $\mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})$ and the diagonal matrix $\mathbf{G}_{\mathbf{uu}}(\lambda)$, thereby inheriting the exact single factor structure of the unconditional covariance matrix of a static factor model. The fact that the idiosyncratic impact of the common factor on each of the observed variables is in principle dynamic implies that the spectral density matrix of $\mathbf{y}_t$ will generally be complex but Hermitian, even though the spectral densities of x_t and u_{it} are all real because they correspond to univariate processes.

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

$$dZ^{x^K}(\lambda) = G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)d\mathbf{Z}^{\mathbf{y}}(\lambda), \tag{8}$$

where $G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)$ is known as the transfer function of the common factor smoother. As a result, the spectral density of the smoothed values of the common factors, $x_{t|\infty}^K$, will be

$$G_{x^Kx^K}(\lambda) = G_{xx}^2(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}) \tag{9}$$

while the spectral density of the final estimation error $x_t - x_{t|\infty}^K$ will be given by

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

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

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

Hence, the spectral density matrix of the smoothed values of the specific factors is 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

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

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

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

Having obtained these, we can easily obtain the smoother for the innovations in common and specific factors, $f_{t|\infty}^K$, and $v_{it|\infty}^K$, respectively, by applying to $x_{t|\infty}^K$ and $u_{it|\infty}^K$ the one-sided filters $\alpha_x(e^{-i\lambda})/\beta_x(e^{-i\lambda})$ and $\alpha_{u_i}(e^{-i\lambda})/\beta_{u_i}(e^{-i\lambda})$. Thus, we can derive their joint spectral density, which plays an important role in our tests, as well as the joint spectral density of their final estimation errors $f_t - f_{t|\infty}^K$ and $v_{it} - v_{it|\infty}^K$.

Finally, we can obtain the autocovariances of $x_{t|\infty}^K$, $f_{t|\infty}^K$, $\mathbf{u}_{t|\infty}^K$, $\mathbf{v}_{t|\infty}^K$ and their final estimation errors by applying the usual inverse Fourier transformation $cov(z_{at}, z_{bt-k}) = \int_{-\pi}^{\pi} e^{i\lambda k} G_{z_a z_b}(\lambda) d\lambda$.

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

$$\begin{aligned} |\mathbf{G_{yy}}(\lambda)| &= |\mathbf{G_{uu}}(\lambda)| G_{xx}(\lambda)\omega(\lambda), \\ \mathbf{G_{yy}^{-1}}(\lambda) &= \mathbf{G_{uu}^{-1}}(\lambda) - \omega(\lambda)\mathbf{G_{uu}^{-1}}(\lambda)\mathbf{c}(e^{-i\lambda})\mathbf{c}'(e^{i\lambda})\mathbf{G_{uu}^{-1}}(\lambda), \\ \omega(\lambda) &= [G_{xx}^{-1}(\lambda) + \mathbf{c}'(e^{i\lambda})\mathbf{G_{uu}^{-1}}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}. \end{aligned} \tag{10}$$

The computational gains arise because $\mathbf{G_{uu}}(\lambda)$ is a diagonal matrix and $\omega(\lambda)$ a scalar (see appendix A of Fiorentini et al (2018) for further details).

If both N is very large and (1) admits a finite static factor representation, one can accurately estimate the latent factors using simpler procedures (see Bai and Ng (2008) or Stock and Watson (2011) and the references therein). But when the cross-sectional dimension is small, the filtered estimates of the state variables are likely to be heavily influenced by the dynamic specification of the model, which thus becomes a first order issue. The objective of our paper is precisely to provide diagnostics for misspecification in $G_{xx}(\lambda)$, $\mathbf{G_{uu}}(\lambda)$ and $\mathbf{c}(e^{-i\lambda})$ in those small N models.

3 Inference procedures

3.1 Identification

The identification by means of homogeneous restrictions of linear dynamic models with latent variables such as (1) was discussed by Geweke (1977) and Geweke and Singleton (1981), and more recently by Scherrer and Deistler (1998) and Heaton and Solo (2004) (see also Forni and Lippi (2001, 2011) and Bai and Wang (2014) for related results). These authors extend

well known results from static factor models and simultaneous equation systems to the spectral density matrix (7) on a frequency by frequency basis. Thus, two models will be observationally equivalent for a Gaussian log-likelihood function if and only if they generate exactly the same spectral density matrix for the observed variables at all frequencies. As in the traditional case, there are two different identification issues:

1. the nonparametric identification of common and specific components,
2. the parametric identification of dynamic loadings and factor dynamics within the common components.

The answer to the first question is easy when $\mathbf{G_{uu}}(\lambda)$ is a diagonal, full rank matrix, as in (1).[2] We can show that for the dynamic single factor model (1), nonparametric identification of common and idiosyncratic terms is guaranteed when $N \geq 3$ provided that at least three series load on the common factor. The intuition is as follows. We know that the condition above coincides with the so-called Ledermann bound for single factor models (see e.g. Bekker and ter Berge (1997) or Scherrer and Deistler (1998), as well as footnote 19). Since it is not possible to transfer variance from the common to the idiosyncratic components (or vice versa) in those circumstances, and any model with more than one factor will lead to some singular idiosyncratic variance, we can uniquely decompose $\mathbf{G_{yy}}(\lambda)$ into the rank one matrix $\mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})$ and the full rank matrix $\mathbf{G_{uu}}(\lambda)$ in this way.

The separate identification of $\mathbf{c}(e^{-i\lambda})$ and $G_{xx}(\lambda)$ is trickier, as we can always write any dynamic factor model (up to time shifts) in terms of white noise common factors, as in (2). But it can be guaranteed (up to scaling and sign changes) if in addition the dynamic loading polynomials $c_i(L)$ are one-sided of finite order and jointly coprime, so they do not share a common root across all N series (see theorem 3 in Heaton and Solo (2004) for a more formal argument along these lines).[3] The only remaining issue is the unconditional scaling of the common factor, which we can achieve by normalising the variance of f_t to 1.[4]

In this paper we mostly focus in hypothesis tests of $p_x = d_x$ vs $p_x = d_x + 1$ or $p_{u_i} = d_{u_i}$ vs $p_{u_i} = d_{u_i} + 1$, or the analogous hypotheses for q_x and q_{u_i}, although we can easily handle higher order alternatives, as we explain at the end of section 3.4. In addition, we are interested in testing the null hypothesis that the maximum lag order of the dynamic loadings is n instead of $n + 1$. For that reason, we henceforth maintain the assumption that the model is identified both under the null and under the AR and MA versions of all those different alternatives (see Fiorentini

[2] Scherrer and Deistler (1998) refer to this situation as the Frisch case.

[3] The one-sided restriction is without loss of generality in models with m and n finite because any shift in the dating of the common factor can be exactly matched by an opposite shift in the timing of the dynamic loadings.

[4] Other symmetric scaling assumptions would normalise $V(x_t)$, or some norm of the loadings vector $\mathbf{c}_0$ or their long run counterparts $\mathbf{c}(1)$. Alternatively, we could asymmetrically fix one element of $\mathbf{c}_0$ or $\mathbf{c}(1)$ to 1.

and Sentana (2016) for the effects of lack of identification under some analogous alternatives in UCARIMA models). This will indeed be the case for the models we consider in the Monte Carlo experiment and empirical application in sections 4 and 5, respectively, for which the conditions in this section guarantee identification both under the null and under all those alternatives.

3.2 Maximum likelihood estimation in the frequency domain

Let

$$\mathbf{I}_{\mathbf{yy}}(\lambda;\boldsymbol{\mu}) = \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\mathbf{y}_t-\boldsymbol{\mu})(\mathbf{y}_s-\boldsymbol{\mu})'e^{-i(t-s)\lambda} \tag{11}$$

denote the Hermitian positive semidefinite periodogram matrix of $\mathbf{y}_t$ and $\lambda_j = 2\pi j/T$ ($j = 0,\dots T-1$) the usual Fourier frequencies. If we assume that $\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})$ is positive definite at all frequencies,[5] the so-called Whittle (1962) discrete,[6] spectral approximation to the log-likelihood function is $L_T(\boldsymbol{\mu},\boldsymbol{\phi}) = \sum_{j=0}^{T-1}\ell(\lambda_j;\boldsymbol{\mu},\boldsymbol{\phi})$, where

$$\ell(\lambda;\boldsymbol{\mu},\boldsymbol{\phi}) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})| - \frac{1}{2}\mathrm{tr}[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda;\boldsymbol{\phi})2\pi\mathbf{I}_{\mathbf{yy}}(\lambda;\boldsymbol{\mu})]. \tag{12}$$

Expression (11), 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 $\mathbf{Y} = (\mathbf{y}_1,\dots,\mathbf{y}_t,\dots,\mathbf{y}_T)'$, the FFT creates the centred and orthogonalised $T\times N$ complex data matrix $\mathbf{Z}^{\mathbf{y}} = (\mathbf{z}_0^{\mathbf{y}},\dots,\mathbf{z}_j^{\mathbf{y}},\dots,\mathbf{z}_{T-1}^{\mathbf{y}})'$ by effectively premultiplying $\mathbf{Y}-\boldsymbol{\iota}_T\boldsymbol{\mu}'$ by the $T\times T$ Fourier matrix $\mathbf{W}$. On this basis, we can easily compute $\mathbf{I}_{\mathbf{yy}}(\lambda_j)$ as $2\pi\mathbf{z}_j^{\mathbf{y}}\mathbf{z}_j^{\mathbf{y}*}$, where * denotes complex conjugate transpose. Hence, $L_T(\boldsymbol{\mu},\boldsymbol{\phi})$ becomes

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

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

But since $\mathbf{z}_j^{\mathbf{y}}$ does not depend on $\boldsymbol{\mu}$ for $j = 1,\dots,T-1$ because $\boldsymbol{\iota}_T$ is proportional to the first column of the orthogonal Fourier matrix and $\mathbf{z}_0^{\mathbf{y}} = (\bar{\mathbf{y}}_T-\boldsymbol{\mu})$, where $\bar{\mathbf{y}}_T$ is the sample mean of $\mathbf{y}_t$, it immediately follows that the MLE of $\boldsymbol{\mu}$ will be $\bar{\mathbf{y}}_T$, so from now on we focus on demeaned variables, maximising the criterion function $L_T(\bar{\mathbf{y}}_T,\boldsymbol{\phi})$ with respect to all the remaining static and dynamic second moment parameters in $\boldsymbol{\phi}$ over the admissible parameter space $\boldsymbol{\Phi}\subseteq\mathbb{R}^d$. It immediately follows that the score with respect to those parameters is

$$\begin{aligned}
\mathbf{s}_{\phi T}(\boldsymbol{\phi}) &= \sum_{j=0}^{T-1}\mathbf{s}_{\phi j}(\boldsymbol{\phi}),\\
\mathbf{s}_{\phi j}(\boldsymbol{\phi}) &= \frac{1}{2}\{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda_j;\boldsymbol{\phi})\right]/\partial\boldsymbol{\phi}\}\mathbf{M}(\lambda_j;\boldsymbol{\phi})\mathbf{m}(\lambda_j;\boldsymbol{\phi}), &(13)\\
\mathbf{m}(\lambda_j;\boldsymbol{\phi}) &= vec[2\pi\mathbf{z}_j^{\mathbf{y}c}\mathbf{z}_j^{\mathbf{y}\prime} - \mathbf{G}'_{\mathbf{yy}}(\lambda_j;\boldsymbol{\phi})], &(14)\\
\mathbf{M}(\lambda_j;\boldsymbol{\phi}) &= \mathbf{G}_{\mathbf{yy}}^{-1}(\lambda_j;\boldsymbol{\phi})\otimes\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda_j;\boldsymbol{\phi}), &(15)
\end{aligned}$$

[5] Otherwise, a linear combination of the components of the $\mathbf{y}_t's$ at frequency λ would be identically 0.

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

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

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

$$\mathcal{I}(\boldsymbol{\phi}) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda_j;\boldsymbol{\phi})\right]/\partial\boldsymbol{\phi}\}\mathbf{M}(\lambda;\boldsymbol{\phi})\{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda_j;\boldsymbol{\phi})\right]/\partial\boldsymbol{\phi}\}^* d\lambda, \tag{16}$$

a consistent estimator of which will be provided by either by the outer product of $\mathbf{s}_{\boldsymbol{\phi}j}(\boldsymbol{\phi})$ or by

$$\boldsymbol{\Phi}(\boldsymbol{\phi}) = \frac{1}{2T}\sum_{j=0}^{T-1}\{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda_j;\boldsymbol{\phi})\right]/\partial\boldsymbol{\phi}\}\mathbf{M}(\lambda_j;\boldsymbol{\phi})\{\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda_j;\boldsymbol{\phi})\right]/\partial\boldsymbol{\phi}\}^*. \tag{17}$$

In fact, by selecting an artificially large value for T in (17), one can approximate (16) to any desired degree of accuracy.

Formal results showing the consistency and asymptotic normality of the resulting ML estimators in identified dynamic latent variable models under suitable regularity conditions were provided by Dunsmuir (1979) and Dzhaparidze (1986) among others, who generalised earlier results for VARMA models by Hannan (1970) and Dunsmuir and Hannan (1976). Those authors also showed the asymptotic equivalence between time and frequency domain ML estimators and the validity of the trinity of classical hypothesis tests in this context. In addition, they explicitly acknowledged the possibility that the normality assumption does not hold, in which case the criterion function (12) must be understood as a pseudo log-likelihood. Appendix D provides a precise statement of Dunsmuir's (1979) regularity conditions for the dynamic factor model in (1), and derives the asymptotic covariance matrix of the Gaussian estimators. Importantly, the models we consider in the Monte Carlo experiments in section 4 satisfy stronger versions of those conditions.

To increase the speed and accuracy of the estimators and their standard errors, we can make use of the numerically reliable and fast to compute expressions for the Jacobian of $vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda)\right]$ and the spectral scores $\mathbf{s}_{\boldsymbol{\phi}j}(\boldsymbol{\phi})$ in appendix C of Fiorentini et al (2018), whose appendix E includes analogous formulae for the information matrix (16). Those expressions make extensive use of the complex version of the Woodbury formula in (10).

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

In any given realisation of the vector process $\{\mathbf{y}_t\}$, the values of $\{x_t\}$ could be regarded as a set of T parameters. With this interpretation in mind, we can define $x_{t|\infty}^G$ as the spectral GLS estimator of x_t through the transformation

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

Similarly, we can define $\mathbf{u}_{t|\infty}^G$ though

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

It is then easy to see that the joint spectral density of $x^G_{t|\infty}$ and $\mathbf{u}^G_{t|\infty}$ will be

$$\begin{bmatrix} G_{xx}(\lambda)+[\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1} & \mathbf{0}' \\ \mathbf{0} & \mathbf{G}_{\mathbf{yy}}(\lambda)-\mathbf{c}(e^{-i\lambda})[\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}\mathbf{c}'(e^{i\lambda}) \end{bmatrix}, \quad (18)$$

with the second block being of rank $N-1$. Since the Jacobian of this orthogonalisation is 1, we can factorise the spectral log-likelihood function of $\mathbf{y}_t$ as the sum of the log-likelihood function of $x^G_{t|\infty}$, which is univariate, and the log-likelihood function of $\mathbf{u}^G_{t|\infty}$. Importantly, the parameters characterising $G_{xx}(\lambda)$ only enter through the first component. In contrast, the remaining parameters affect both components. Moreover, we can easily show that

1. $x^G_{t|\infty} = x_t + \zeta^G_{t|\infty}$, with x_t and $\zeta^G_{t|\infty}$ orthogonal at all leads and lags
2. The smoothed estimator of x_t obtained by applying the Wiener-Kolmogorov filter to $x^G_{t|\infty}$ coincides with $x^K_{t|\infty}$.

This confirms that $x^G_{t|\infty}$ constitute minimal sufficient statistics for x_t, thereby generalising earlier results by Jungbacker and Koopman (2015), who considered models in which $\mathbf{c}(e^{-i\lambda}) = \mathbf{c}$ for all λ, and Fiorentini et al (2004), who looked at the related class of factor models with time-varying volatility (see also Gouriéroux et al (1991)). In addition, the degree of unobservability of x_t depends exclusively on the size of $[\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}$ relative to $G_{xx}(\lambda)$ (see Sentana (2004) for a closely related discussion).

3.4 Neglected serial correlation in the common factor

We would like to test the null hypothesis $H_0 : \psi_{x1} = 0$ in the alternative model

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{c}(L)x_t + \mathbf{u}_t, \quad (1-\psi_{x1}L)\alpha_x(L)x_t = \beta_x(L)f_t, \quad \mathbf{A}(L)\mathbf{u}_t = \mathbf{B}(L)\mathbf{v}_t.$$

Given the spectral density of the dynamic GLS estimator of the common factor in (18),

$$\partial G_{x^G x^G}(\lambda)/\partial\psi_{x1} = \partial G_{xx}(\lambda)/\partial\psi_{x1}.$$

Since ψ_{x1} only enters through the marginal log-likelihood of $x^G_{t|\infty}$, its score will be

$$\frac{1}{2}\sum\nolimits_{j=0}^{T-1}[\partial G_{xx}(\lambda)/\partial\psi_{x1}]G^{-1}_{x^G x^G}(\lambda_j)[2\pi I_{x^G x^G}(\lambda_j) - G_{x^G x^G}(\lambda_j)].$$

But $\partial G_{xx}(\lambda)/\partial\psi_{x1} = 2\cos\lambda G_{xx}(\lambda)$ when $\psi_{x1} = 0$, so after some straightforward algebraic manipulations, we can show that this score can be written under the null as

$$\sum\nolimits_{j=0}^{T-1}\cos\lambda_j G^{-1}_{xx}(\lambda_j)[2\pi I_{x^K x^K}(\lambda_j) - G_{x^K x^K}(\lambda_j)] = \sum\nolimits_{j=0}^{T-1}\cos\lambda_j[2\pi I_{f^K f^K}(\lambda_j) - G_{f^K f^K}(\lambda_j)].$$

Hence, the time domain counterpart to the spectral score with respect to ψ_{x1} is (asymptotically) proportional to the difference between the first sample (circulant) autocovariance of $f^K_{t|\infty}$ and its theoretical counterpart under H_0. Thus, the only difference with a situation in which x_t is

observable is that the autocovariance of $f^K_{t|\infty}$ is no longer 0 when $\psi_{x1} = 0$, although it approaches 0 as the signal to noise ratio increases, in which case our proposed test would converge to the usual Breusch (1978) - Godfrey (1978) LM test for neglected serial correlation.

Let us illustrate our procedure in a simple example. Imagine that the model under the alternative is a second order version of (6). The results in section 2.6 of Fiorentini and Sentana (2013) imply that $x^K_{t|\infty}$ will have the autocorrelation structure of an AR(2) when $\psi_{x1} = 0$, while $f^K_{t|\infty}$ will follow an AR(1) with first order autocovariance $(\mathbf{c}'\mathbf{\Gamma}^{-1}\mathbf{c})\alpha_{x1}/(1-\alpha^2_{fK})$, where

$$\alpha_{fK} = \left\{1 + \alpha^2_{x1} + (\mathbf{c}'\mathbf{\Gamma}^{-1}\mathbf{c}) - \sqrt{[(1+\alpha_{x1})^2 + (\mathbf{c}'\mathbf{\Gamma}^{-1}\mathbf{c})][(1-\alpha_{x1})^2 + (\mathbf{c}'\mathbf{\Gamma}^{-1}\mathbf{c})]}\right\}/(2\alpha_{x1}).$$

Therefore, the larger $(\mathbf{c}'\mathbf{\Gamma}^{-1}\mathbf{c})$ is, the closer $f^K_{t|\infty}$ will be to white noise. In general, the LM test of $H_0 : \psi_{x1} = 0$ will simply compare the first sample autocovariance of $f^K_{t|\infty}$ to its theoretical value above. This interpretation is in line with Maravall's (1987) suggestion that large discrepancies between the theoretical and empirical autocovariance functions of the estimators of the unobserved components provide an indication of model misspecification in UCARIMA models. However, our proposed LM statistics carry out this comparison as formal statistical tests. In addition, an important advantage of our frequency domain approach is that we implicitly compute the required autocovariances without explicitly obtaining the time processes for the smoothed estimates of the unobserved components through the Riccati equation.

Unfortunately, the approach that we have used to obtain the score for neglected autocorrelation in the common factor cannot be generally applied to the specific factors because the parameters in $\mathbf{G_{uu}}(\lambda)$ affect both components of the orthogonalised spectral log-likelihood function. Nevertheless, we can start from first principles by exploiting the fact that

$$\partial vec[\mathbf{G_{yy}}(\lambda)]/\partial\psi_{x1} = [\mathbf{c}(e^{i\lambda}) \otimes \mathbf{c}(e^{-i\lambda})]\partial G_{xx}(\lambda)/\partial\psi_{x1}.$$

Not surprisingly, if we introduce these derivatives in the formula for the spectral score with respect to ψ_{x1}, we end up with exactly the same frequency-domain and time-domain expressions.

Empirical researchers often assume that the common factors are white noise for identification purposes, so that $G_{xx}(\lambda) = 1$ under the null. Since we make no assumptions on p_x and q_x, our tests trivially apply in that situation too. Similarly, generalisations to test ARMA(p,q) vs ARMA(p+k,q) in the common factor are straightforward, as they only involve higher order autocovariances of $f^K_{t|\infty}$. Moreover, it is easy to show that ARMA(p+k,q) and ARMA(p,q+k) multiplicative alternatives are locally asymptotically equivalent, as in the case of univariate tests for serial correlation in observable time series (see e.g. Godfrey (1988)). Finally, we could also consider (multiplicative) seasonal alternatives.

3.5 Neglected serial correlation in specific factors

Next, we focus on the null hypothesis $H_0 : \boldsymbol{\psi}_{\mathbf{u}1} = \mathbf{0}$ in the alternative model

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{c}(L)x_t + \mathbf{u}_t, \quad \alpha_x(L)x_t = \beta_x(L)f_t, \quad [\mathbf{I} - diag(\boldsymbol{\psi}_{\mathbf{u}1})L]\mathbf{A}(L)\mathbf{u}_t = \mathbf{B}(L)\mathbf{v}_t,$$

where $\boldsymbol{\psi}'_{\mathbf{u}1} = (\psi_{u_1 1}, \ldots, \psi_{\mathbf{u}_N 1})$. In this case, we have that

$$\partial vec[\mathbf{G}_{\mathbf{yy}}(\lambda)]/\partial \boldsymbol{\psi}'_{\mathbf{u}_1} = \mathbf{E}_N \cdot \partial vecd[\mathbf{G}_{\mathbf{uu}}(\lambda)]/\partial \boldsymbol{\psi}'_{\mathbf{u}_1},$$

where $\mathbf{E}'_N = (\mathbf{e}_1\mathbf{e}'_1|\ldots|\mathbf{e}_N\mathbf{e}'_N)$, with $(\mathbf{e}_1|\ldots|\mathbf{e}_N) = \mathbf{I}_N$, is the unique $N^2 \times N$ "diagonalisation" matrix that transforms $vec(\mathbf{A})$ into $vecd(\mathbf{A})$ as $vecd(\mathbf{A}) = \mathbf{E}'_N vec(\mathbf{A})$ (see Magnus (1988)). Straightforward algebra implies that the score with respect to $\psi_{u_i 1}$ under the null will be

$$\sum_{j=0}^{T-1} \cos\lambda_j G^{-1}_{u_i u_i}(\lambda_j)[2\pi I_{u_i^K u_i^K}(\lambda_j) - G_{u_i^K u_i^K}(\lambda_j)] = \sum_{j=0}^{T-1} \cos\lambda_j [2\pi I_{v_i^K v_i^K}(\lambda_j) - G_{v_i^K v_i^K}(\lambda_j)].$$

Thus, the time domain counterpart to the spectral score with respect to $\psi_{u_i 1}$ will be proportional to the difference between the first sample autocovariance of v_{it}^K and its theoretical value under H_0. Joint tests that look at several idiosyncratic terms together, as well as the common factor, can be easily obtained by combining the different scores involved. As we shall see in sections 4.2 and 5, though, the individual tests are rather good at identifying the source of the rejection.

3.6 Additional lags in the dynamic factor loadings

As we mentioned in section 2.1, the dynamic nature of model (1) is due to three characteristics: 1) the serial correlation of the common factor x_t; 2) the serial correlation of the idiosyncratic factors $\mathbf{u}_t$ and 3) the heterogeneous dynamic impact of the common factor on each of the observed variables through the cross-sectionally heterogeneous dynamic loadings $c_i(L)$.

We have already discussed dynamic specification tests for the first two characteristics in sections 3.4 and 3.5, respectively, so in this section we concentrate in the last one. For the sake of brevity, we focus on multiplicative alternatives involving a single additional lag, although it is straightforward to consider additive alternatives, multiple lags or indeed combinations of leads and lags. Specifically, we look at the null hypothesis $H_0 : \boldsymbol{\psi}_c = \mathbf{0}$ in the alternative model

$$\mathbf{y}_t = \boldsymbol{\mu} + [(\boldsymbol{\iota}_N - \boldsymbol{\psi}_c L) \odot \mathbf{c}(L)]x_t + \mathbf{u}_t, \quad \alpha_x(L)x_t = \beta_x(L)f_t, \quad \mathbf{A}(L)\mathbf{u}_t = \mathbf{B}(L)\mathbf{v}_t, \tag{19}$$

with $\odot$ denoting Hadamard products and

$$(\boldsymbol{\iota}_N - \boldsymbol{\psi}_c L) = (1 - \psi_{c_1} L, \ldots, 1 - \psi_{c_N} L)'.$$

Given that the dynamic loadings become

$$(\boldsymbol{\iota}_N - \boldsymbol{\psi}_c L) \odot \mathbf{c}(L) = \sum_{k=-m}^{n} \mathbf{c}_k L^k - \sum_{k=-m+1}^{n+1} (\mathbf{c}_{k-1} \odot \boldsymbol{\psi}_c) L^k$$

under the alternative, we will have that the score corresponding to $\boldsymbol{\psi}_c$ will be given by the sum across frequencies of terms of the form

$$-\sum_{k=-m+1}^{n+1} \mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}_{k-1}\{e^{-ik\lambda}[2\pi\mathbf{I}'_{x^K\mathbf{u}^K}(\lambda) - \mathbf{G}'_{x^K\mathbf{u}^K}(\lambda)] + e^{-ik\lambda}[2\pi\mathbf{I}_{\mathbf{u}^K x^K}(\lambda) - \mathbf{G}_{\mathbf{u}^K x^K}(\lambda)]\}.$$

The time domain analogue to this expression is easiest to understand with white noise idiosyncratic components. Then, the score with respect to ψ_{c_i} evaluated under H_0 would be exactly proportional to the difference between the sample and population (circulant) covariance of u_{it} with the distributed lag $\sum_{k=-m+1}^{n+1} c_{i,k-1}x_{t-k}^K$. More generally, it will contain the difference between the sample and population covariance of v_{it} with the distributed lag $\sum_{k=-m+1}^{n+1} c_{i,k-1}x_{i,t-k}^{*K}$, where $x_{i,t-n-1}^{*K}$ is the smoothed value of the $(n+1)^{th}$ lag of the GLS-transformed regressor

$$x_{i,t}^* = \alpha_{u_i}(L)\beta_{u_i}^{-1}(L)x_t.$$

Importantly, we obtain the same score (with an opposite sign) if we consider the alternative

$$(1-\psi_{c_i}L)^{-1}c_i(L) = c_i(L)\sum_{s=0}^{\infty}\psi_{c_i}^s L^s, \tag{20}$$

for which there is no longer a static factor model representation with less than N factors.

An interesting special case arises if we consider the following restricted version of (20):

$$\mathbf{y}_t = \boldsymbol{\mu} + [(1-\psi_c L)^{-1}\mathbf{c}(L)]x_t + \mathbf{u}_t,$$

so that the null hypothesis is $H_0 : \psi_c = 0$. Given that we can write this alternative model as

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{c}(L)x_t^* + \mathbf{u}_t, \quad (1-\psi_c L)\alpha_x(L)x_t^* = \beta_x(L)f_t,$$

the spectral score with respect to a common value of ψ_c is numerically identical to the score of an additional autoregressive term ψ_x in the process for x_t^*. This result confirms the partial substitutability of dynamics in the common factor by dynamics in the factor loadings that we mentioned in section 2.1. It also implies that we cannot simultaneously test the null hypotheses $H_0 : \boldsymbol{\psi}_c = \mathbf{0}$ and $H_0 : \psi_x = 0$ because the latter is implied by the former.[7]

3.7 Parameter uncertainty

So far we have implicitly assumed the true values of the parameters of model (1) are known. In practice, those parameters will have to be estimated under the null. Maximum likelihood estimation of the dynamic factor model parameters can be done either in the time domain using the Kalman filter or in the frequency domain, as explained in section 3.2. The sampling

[7] This problem would not arise if x_t were observed because ψ_c enters through the conditional model of $\mathbf{y}_t$ given x_t while ψ_x enters through the marginal model for x_t. In that case, the score with respect to ψ_c would simply be the sum across the N series of the scores for the different $\psi'_{c_i}s$.

uncertainty surrounding the sample mean $\boldsymbol{\mu}$ is asymptotically inconsequential because the information matrix is block diagonal. The sampling uncertainty surrounding the other parameters is not necessarily so. Let us partition the d vector of model parameters $\boldsymbol{\phi}$ as $(\boldsymbol{\theta}', \boldsymbol{\psi}')'$, where $\boldsymbol{\theta}$ contains the d_1 parameters of the model under the null and $\boldsymbol{\psi}$ and the d_2 parameters that are tested under the alternative. A block diagonal information matrix for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ is only obtained in some special cases. One example arises when $\mathbf{c}(e^{-i\lambda}) = \mathbf{c}$ and both common and idiosyncratic factors follow AR(1) processes with a common autoregressive coefficient. Another important example are the static factor models considered by Fiorentini and Sentana (2015). In that situation, all final prediction errors are white noise under the null, and one can safely ignore the estimation error in $\boldsymbol{\theta}$.

More generally, we need to take into account the asymptotic dependence between $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$. The solution is the standard one: replace $\mathcal{I}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\theta}_0, \mathbf{0})$ with

$$\mathcal{I}^{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\theta}_0, \mathbf{0}) = [\mathcal{I}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\theta}_0, \mathbf{0}) - \mathcal{I}^{-1}_{\boldsymbol{\psi}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathbf{0})\mathcal{I}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathbf{0})\mathcal{I}^{-1\prime}_{\boldsymbol{\psi}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathbf{0})]^{-1}, \tag{21}$$

which is the $(\boldsymbol{\psi}, \boldsymbol{\psi})$ block of the inverse information matrix, so that the test statistic becomes

$$LM_T = T \cdot \bar{\mathbf{s}}'_{\boldsymbol{\psi}T}(\tilde{\boldsymbol{\theta}}_T, \mathbf{0})\mathcal{I}^{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\theta}_0, \mathbf{0})\bar{\mathbf{s}}'_{\boldsymbol{\psi}T}(\tilde{\boldsymbol{\theta}}_T, \mathbf{0}), \tag{22}$$

where $\tilde{\boldsymbol{\theta}}_T$ is the MLE obtained under the null. The analytical expressions for the information matrix in appendix E of Fiorentini et al (2018) provide a computationally efficient method for (21). Importantly, the dual nature of our proposed tests implies that they can be applied regardless of the model having been estimated in the time or frequency domains.

3.8 Reinterpreting reduced form tests for neglected serial correlation

In the context of univariate time series models written in state space form, Harvey (1989), Harvey and Koopman (1992) and Durbin and Koopman (2012) suggest the calculation of neglected serial correlation tests for the reduced form residuals, which should be white noise under the null of correct dynamic specification. The analogue procedure in the context of the dynamic factor model (1) involves testing for neglected serial correlation in the multivariate vector of Wold innovations $\mathbf{w}_t$ in (5). In the first order case, in particular, one would test the null hypothesis $H_0 : \boldsymbol{\Psi}_w = \mathbf{0}$ in the VAR(1) model $\mathbf{w}_t = \boldsymbol{\Psi}_w \mathbf{w}_{t-1} + \boldsymbol{\eta}_t$, as in Hendry (1971), Guilkey (1974) and Harvey (1982) (see also Hosking (1981)).

Such a test may seem to offer a substantial computational edge over ours because apparently it only require the OLS regression of the Kalman filter one-period ahead prediction errors on their first lag. However, we know from Durbin (1970) that the asymptotic size of serial correlation tests for observed variables applied to estimated residuals in dynamic models is wrong. In fact,

their correct computation requires purging the scores corresponding to the elements of $\mathbf{\Psi}_w$ of the sampling variability in all the model parameters estimated under the null.

In turns out that one can express the parametric restrictions assessed by reduced form tests in terms of the structural parameters of a dynamic factor model. For pedagogical reasons, in this section we do so for the simpler null hypothesis $H_0 : \boldsymbol{\psi}_w = vecd(\mathbf{\Psi}_w) = \mathbf{0}$, postponing the discussion of the general case to appendix C.[8] In particular,

Proposition 1 *After correcting for parameter uncertainty, the LM test of* $H_0 : \boldsymbol{\psi}_w = \mathbf{0}$ *in the model* $[\mathbf{I} - diag(\boldsymbol{\psi}_w)L]\mathbf{w}_t = \boldsymbol{\eta}_t$ *for the reduced form residuals coincides with the LM test of the same null hypothesis in the dynamic factor model:*

$$\mathbf{y}_t = \boldsymbol{\mu} + [(\boldsymbol{\iota}_N - \boldsymbol{\psi}_w L) \odot \mathbf{c}(L)]x_t + \mathbf{u}_t,\ \alpha_x(L)x_t = \beta_x(L)f_t,\ \mathbf{A}(L)\mathbf{u}_t = [\mathbf{I} - diag(\boldsymbol{\psi}_w)L]\mathbf{B}(L)\mathbf{v}_t. \quad (23)$$

In other words, testing for univariate serial correlation in the reduced form of one of the observed series, say y_{it}, is equivalent to simultaneously testing against an alternative whose factor loading is $c_i(L)(1 - \psi_{c_i}L)$, as in section 3.6, and the MA part of the process for the idiosyncratic component u_{it} contains the neglected multiplicative term $(1 - \psi_{u_i}L)$, as in section 3.5, under the maintained assumption that $\psi_{c_i} = \psi_{u_i} = \psi_{w_i}$. In contrast, a test for neglected serial correlation in the loadings only focuses on ψ_{c_i}, while a test for neglected serial correlation in the idiosyncratic component concentrates on ψ_{u_i}. As a result, the relative power of those three tests will depend on the nature of the true model under the alternative. In particular, if we represent ψ_{c_i} on the horizontal axis and ψ_{u_i} on the vertical axis, the reduced form test will have maximum power for alternatives along the 45° degree line $\psi_{c_i} = \psi_{u_i}$, while the structural form tests of the null hypotheses $H_0 : \psi_{c_i} = 0$ and $H_0 : \psi_{u_i} = 0$ will have maximum power along their respective axis. Finally, it is also possible to compare the power of those three tests to the power of the joint test of $H_0 : \psi_{c_i} = \psi_{u_i} = 0$, which has twice as many degrees of freedom (see Demos and Sentana (1998) for a related discussion in the context of ARCH tests).

None of those procedures, though, is likely to have much power against neglected serial correlation in the common factor, as this misspecification will manifest itself mostly in the dynamic cross-correlations, especially if the signal to noise ratio is low. For that reason, we explicitly compare the power of all the different tests in the Monte Carlo experiments.

Importantly, given that the reduced form will be dynamically misspecified when any of three dynamic characteristics of model (1) is misspecified, the validity of the null distribution of all the tests that we consider, including those based on the reduced form, requires the correct specification of the structural model (1) under H_0.

[8] Nevertheless, we take into account that the elements of $\mathbf{w}_t$ will be contemporaneously correlated even when they are serially uncorrelated. Thus, our test differs from Test 2 in Harvey (1982), which looks at the null hypothesis $H_0 : \boldsymbol{\psi}_w = \mathbf{0}$ when $\mathbf{w}_t$ is observed under the maintained assumption that $V(\boldsymbol{\eta}_t)$ is diagonal.

3.9 Robustness of the Gaussian tests

We have so far treated the Gaussian assumption made for estimation as an integral part of the model. But for the purposes of testing the validity of the dynamic specification, it should be regarded as a maintained assumption instead. A valid concern, therefore, is whether our tests will be affected if the latent variables are non-Gaussian but we still use $\mathcal{I}^{\psi\psi}(\boldsymbol{\theta}_0, \mathbf{0})$ in (22).

Given that under the regularity conditions stated in appendix D, which are satisfied by the models considered in section 4, the asymptotic distribution of the Gaussian pseudo maximum likelihood estimator of $\boldsymbol{\phi}$ (appropriately centred and scaled) will be normal with zero mean and covariance matrix given by the usual sandwich formula $\mathbf{C}_0 = \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}$, where $\mathbf{A}$ is the plim of the (minus) expected Hessian and $\mathbf{B}$ the asymptotic variance of the Gaussian pseudo log-likelihood score $\bar{\mathbf{s}}_{\boldsymbol{\phi}T}(\boldsymbol{\phi})$, we could always resort to the robust Gaussian pseudo score test

$$T \cdot \bar{\mathbf{s}}'_{\boldsymbol{\psi}T}(\tilde{\boldsymbol{\theta}}_T, \mathbf{0})\mathbf{A}^{\psi\psi}(\boldsymbol{\theta}_0, \mathbf{0})\,\mathbf{C}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\theta}_0, \mathbf{0})\,\mathbf{A}^{\psi\psi}(\boldsymbol{\theta}_0, \mathbf{0})\,\bar{\mathbf{s}}_{\boldsymbol{\psi}T}(\tilde{\boldsymbol{\theta}}_T, \mathbf{0}), \tag{24}$$

where $\mathbf{C}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ and $\mathbf{A}^{\psi\psi}$ are the relevant blocks of $\mathbf{C}$ and the inverse of $\mathbf{A}$, respectively (see e.g. Engle (1984)). Under normality, of course, the information matrix equality $\mathbf{B} = \mathbf{A} = \mathcal{I}$ holds, so that (24) and (22) coincide, but in general $\mathbf{B}$ will differ from $\mathcal{I}$ under non-normality. Consistently estimating $\mathbf{A}$ is straightforward because $\mathbf{A} = \mathcal{I}$ irrespective of the Gaussian log-likelihood being the true one for many models, including the dynamic factor model in (1). In the time domain, it is also easy to consistently estimate $\mathbf{B}$ by means of the outer product of the score. In the frequency domain, in contrast, the sample variance of the spectral scores converges in probability to $\mathcal{I}$ instead regardless of the assumption of normality. In principle, we could try to estimate $\mathbf{B}$ by using the analytical expressions we develop in appendix D. Nevertheless, it turns out that those calculations are unnecessary. To argue our claim, we proceed in two steps. First, Proposition 2 below provides a necessary and sufficient condition for $\mathbf{C}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ to coincide with $\mathcal{I}^{\psi\psi}(\boldsymbol{\theta}_0, \mathbf{0})$. Second, we verify that it is satisfied under certain assumptions by the alternatives that we have considered in sections 3.4, 3.5 and 3.6 when the null hypothesis $H_0 : \boldsymbol{\psi} = \mathbf{0}$ holds.

Proposition 2 *Let* $\bar{\mathbf{s}}_{\boldsymbol{\psi}|\boldsymbol{\theta}T}(\boldsymbol{\phi}) = \bar{\mathbf{s}}_{\boldsymbol{\psi}T}(\boldsymbol{\phi}) - \mathbf{A}_{\boldsymbol{\psi}\boldsymbol{\theta}}\mathbf{A}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\boldsymbol{\phi})$. *Under the maintained assumption that* $\mathbf{A} = \mathcal{I}$ *holds,*

$$\mathbf{C}_{\boldsymbol{\psi}\boldsymbol{\psi}} = \mathcal{I}^{\psi\psi} \textit{ if and only if } \lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\psi}|\boldsymbol{\theta}T}(\boldsymbol{\phi}_0)] = -\lim_{T\to\infty} E[\partial\bar{\mathbf{s}}_{\boldsymbol{\psi}|\boldsymbol{\theta}T}(\boldsymbol{\phi}_0)/\partial\boldsymbol{\psi}']. \tag{25}$$

Importantly, this proposition remains valid regardless of the true value of $\boldsymbol{\psi}$ being $\mathbf{0}$ and applies to many other situations beyond dynamic factor models. To interpret it in contexts other than $\mathbf{A}_{\boldsymbol{\psi}\boldsymbol{\theta}} = \mathbf{0}$ and $\mathbf{B}_{\boldsymbol{\psi}\boldsymbol{\psi}} = \mathbf{A}_{\boldsymbol{\psi}\boldsymbol{\psi}}$, it is convenient to recall that the generalised information matrix equality (see Newey and McFadden (1994)) implies that

$$-\lim_{T\to\infty} E[\partial\bar{\mathbf{s}}_{\boldsymbol{\psi}|\boldsymbol{\theta}T}(\boldsymbol{\phi}_0)/\partial\boldsymbol{\theta}] = \lim_{T\to\infty} cov[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\psi}|\boldsymbol{\theta}T}(\boldsymbol{\phi}_0), \sqrt{T}\bar{\mathbf{r}}_{\boldsymbol{\theta}T}(\boldsymbol{\phi}_0)] = \mathbf{A}_{\boldsymbol{\psi}\boldsymbol{\theta}} - \mathbf{A}_{\boldsymbol{\psi}\boldsymbol{\theta}}\mathbf{A}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \mathbf{0},$$

where $\bar{\mathbf{r}}_{\boldsymbol{\psi}T}(\boldsymbol{\phi}_0)$ is the true but unknown log-likelihood (average) score. Therefore, we can understand $\bar{\mathbf{s}}_{\boldsymbol{\psi}|\boldsymbol{\theta}T}(\boldsymbol{\phi})$ as the Gaussian pseudo log-likelihood average score of $\boldsymbol{\psi}$ purged of the sampling variability in estimating $\boldsymbol{\theta}$, so that the asymptotic distribution of the estimator of $\boldsymbol{\psi}$ based on it is the same regardless of $\boldsymbol{\theta}_0$ being known or estimated. In this context, the necessary and sufficient condition in (25) coincides with the efficiency condition for sequential estimators in Newey and Powell (1998), who showed that this condition guarantees that there is no efficiency loss in sequentially estimating $\boldsymbol{\psi}$ keeping $\boldsymbol{\theta}$ fixed at some initial consistent estimator (see also Amengual et al (2013), who applied condition (25) to find the optimal sequential GMM estimator of the shape parameters in multivariate dynamic location scale models).

For practical purposes, it is convenient to express condition (25) as

$$\left(\begin{array}{cc} -\mathbf{A}_{\boldsymbol{\psi}\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} & \mathbf{I}_{d_2} \end{array} \right)(\mathbf{B}-\mathbf{A}) = \mathbf{0},$$

which is very easy to check.

Fiorentini and Sentana (2016) state that one could use Dunsmuir's (1979) results for a univariate $\text{AR}(p)$ process cloaked in white noise in which the innovations to the signal component are independent of the noise to show that Gaussian tests for neglected serial correlation in the signal or the noise continue to have asymptotically valid sizes under non-normality.

It turns out that a similar result holds far more generally. Specifically, we can make use of the analytical expressions for the expected Hessian and the variance of the score in appendix D under the maintained assumption that the innovations $(f_t, v_{1,t}, \ldots, v_{N,t})$ are stochastically independent to verify that condition (25) is satisfied by the alternative models that we have considered in sections 3.4, 3.5 and 3.6 when the null hypothesis $H_0 : \boldsymbol{\psi} = \mathbf{0}$ holds.

3.10 Block-diagonal idiosyncratic autocovariance matrices

It is straightforward to extend the testing procedures we have developed in previous sections to models with multiple common factors. Although this would be intensive in notation, the only additional question would be dealing with identification issues before estimating the model. But sometimes researchers feel compelled to add more common factors to adequately capture the off-diagonal elements of the autocovariance matrices even though there seems to be a single pervasive source of variation. When the cross-sectional dimension, N, is commensurate with the time series dimension, T, one possibility is to rely on the approximate factor structures originally introduced by Chamberlain and Rothschild (1983) in the static case, which allow for some mild contemporaneous and dynamic correlation between idiosyncratic terms. An expanding, influential body of literature has shown that one may accurately recover the unobserved series by using the frequency domain version of principal components put forward by Brillinger (1981, ch. 9) and further extended by Forni et al (2000), which is based on a nonparametric estimate

of the spectral density matrix of the observed series (see Forni et al (2015) for more recent developments). In fact, it might even be possible to use static principal components if the model has a static representation (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 such as the one discussed in section 5 in which N is very small relative to T.

In those situations in which it is natural to group the N series in $\mathbf{y}_t$ into R homogeneous blocks, $\mathbf{y}_{1t}, \ldots, \mathbf{y}_{rt}, \ldots, \mathbf{y}_{Rt}$ of dimension $N_1, \ldots, N_r, \ldots N_R$, with $N_1 + \ldots + N_r + \ldots + N_R = N$, an attractive solution are dynamic bifactor models with two types of factors:

1. Pervasive common factors that affect all N series
2. Block factors that only affect those series from the same block.

Specifically, a model with a single global factor and a single factor per block is defined in the time domain by the system of dynamic stochastic difference equations

$$\left.\begin{array}{c} \mathbf{y}_{rt} = \boldsymbol{\mu}_r + \mathbf{c}_{rg}(L)x_{gt} + \mathbf{c}_{rr}(L)x_{rt} + \mathbf{u}_{rt}, \quad r = 1, \ldots, R \\ \alpha_{x_g}(L)x_{gt} = \beta_{x_g}(L)f_{gt}, \\ \alpha_{x_r}(L)x_{rt} = \beta_{x_r}(L)f_{rt}, \quad r = 1, \ldots, R \\ \alpha_{u_i}(L)u_{i,t} = \beta_{u_i}(L)v_{i,t}, \quad i = 1, \ldots, N, \\ (f_{gt}, f_{1t}, \ldots, f_{Rt}, v_{1t}, \ldots, v_{Nt})|I_{t-1}; \boldsymbol{\mu}, \boldsymbol{\phi} \sim N[0, diag(1, 1, , \ldots, 1, \gamma_{v_1}, \ldots, \gamma_{v_N})], \end{array}\right\} \tag{26}$$

where x_{gt} is the global factor, x_{rt} the r^{th} block factor, $\mathbf{u}_t = (\mathbf{u}'_{1t}, \ldots, \mathbf{u}'_{rt}, \ldots, \mathbf{u}'_{Rt})'$ the N specific factors, $\mathbf{c}_{rg}(L) = \sum_{k=-m_g}^{n_g} \mathbf{c}_{rgk}L^k$ and $\mathbf{c}_{rr}(L) = \sum_{l=-m_r}^{n_r} \mathbf{c}_{rrl}L^k$ are $N_R \times 1$ vectors of possibly two-sided polynomials in the lag operator, $\alpha_{x_g}(L)$, $\alpha_{x_r}(L)$ 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 $\boldsymbol{\phi}$ refers to all the remaining parameters. It is easy to see that the spectral density matrix of $\mathbf{y}_t$ corresponds to a dynamic single factor model with a block-diagonal idiosyncratic autocovariance structure.

Fiorentini et al (2016) explain how to efficiently exploit the sparsity of the factor loading matrix of models such as (26) to successfully estimate them by maximising the spectral Gaussian log-likelihood function with a large number of series from multiple blocks. More importantly for our purposes, the spectral scores for pervasive factors, block factors and idiosyncratic factors, as well as for the corresponding dynamic loadings that they provide in their appendix A are entirely analogous to the ones we derive in sections 3.4, 3.5 and 3.6, respectively. As a result, it is straightforward to compute LM tests for dynamic misspecification in any of those components

in models with block-diagonal idiosyncratic autocovariance matrices. As expected, those tests have exactly the same time domain interpretation as moment tests that compare the sample autocovariances and cross-covariances of the different latent variables with their theoretical values under the null of correct specification.

The first step in the specification of bifactor models is the assignment of the N series to the R blocks. Inspecting the off-diagonal elements of a preliminary, consistent estimator of the "idiosyncratic" spectral density matrix of the N-variate process $\mathbf{y}_t$, such as the difference between nonparametric estimators of $\mathbf{G_{yy}}(\lambda)$ and the spectral density matrix of the common components suggested by Forni et al (2000)), may provide a very good starting point for a clustering algorithm that assigns individual series to blocks in large N models.[9]

4 Monte Carlo simulation

4.1 Size experiment

We assess the finite sample size of the different tests that we have discussed by generating 10,000 samples of length 500 (roughly 4 decades of monthly data), plus 50 for initialization, of a trivariate dynamic factor model. The main reason for looking at such a small cross-sectional dimension is to handicap our proposed tests relative to the general first order version of the reduced form test, which involves N^2 moment conditions. The first model that we simulate and estimate under the null is

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} 0.7 \\ 0.5 \\ 0.4 \end{bmatrix} x_t + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix},$$

$$(1 - .4L - .2L^2)x_t = f_t, \quad (1 + .4L)u_{1,t} = v_{1,t}, \quad (1 - .6L)u_{2,t} = v_{2,t}, \quad (1 - .2L)u_{3,t} = v_{3,t},$$

$$V(f_t) = 1, \qquad vecd'[V(\mathbf{v}_t)] = (0.4, 0.3, 0.8).$$

with $(f_t, v_{1t}, v_{2t}, v_{3t})$ being contemporaneously independent Gaussian white noises. Thus, under the null the common factor follows an AR(2), its loadings are static and the idiosyncratic terms follow AR(1) processes. Since we work with demeaned variables, the true value of $\boldsymbol{\mu}$ is numerically inconsequential, so we fix it to 0.

We compute LM tests against first order versions of:

1. Neglected residual autocorrelation in the common factor (χ^2_1)
2. Neglected residual autocorrelation in all the specific factors (χ^2_3)
3. Neglected multiplicative polynomials in the dynamic loadings (χ^2_3)
4. A combination of 1. and 2. (χ^2_4)
5. A combination of 2. and 3. (χ^2_6)

[9] We are grateful to the referee for this suggestion.

6. Neglected serial correlation in the reduced form residuals (χ^2_9)
7. Diagonal version of the reduced form test in 6. (χ^2_3)

Importantly, all our tests are numerically invariant to whether in estimating the model we normalise the variance of the common factor x_t or its innovation f_t to 1 because of the way we compute the information matrix (see Dufour and Dagenais (1991)). Further, additive and multiplicative versions of the dynamic loadings tests are also numerically identical once we correct for sampling uncertainty under the null.

The left panel of Table 1 shows that all tests have remarkably small size distortions.

Then, we conduct a second experiment with a virtually identical design, except that the Gaussian white noises are replaced by Student $t's$ with ten degrees of freedom but the same variances. The results reported in the right panel of Table 1 confirm the robustness of the Gaussian test we explained in section 3.9.

Table 1: Size of dynamic misspecification tests

	Empirical rejection rates (%)					
	Gaussian			Student t		
Nominal size	10%	5%	1%	10%	5%	1%
Common Factor	9.64	4.86	0.86	9.87	4.86	0.97
Specific Factors	10.46	5.16	1.07	10.24	5.01	1.03
Loadings	9.85	4.8	0.97	9.88	5.10	0.99
All Factors	10.08	5.04	0.88	10.08	4.75	0.85
Loadings+Specific	10.07	5.18	0.99	10.04	4.78	1.88
Reduced form	10.53	5.41	1.02	10.09	5.04	1.06
Diagonal reduced form	10.86	5.19	0.99	9.71	4.89	0.98

4.2 Power experiments

Next, we carry out four additional simulation experiments to assess the relative power of our proposed tests and the reduced form tests. In the first experiment, we simulate and estimate another 10,000 samples of length 500 in which the DGP for the common factors has $\psi_x(L) = (1 - .5L)^{-1}$ but the same first and second-order autocorrelation as under the null, so that

$$x_t = 0.874x_{t-1} - 0.037x_{t-2} + f_t - 0.5f_{t-1}. \tag{27}$$

We also re-scale the loadings so as to maintain the same unconditional covariance matrix as in section 4.1 in order to achieve the same average signal to noise ratio (see the discussion at the end of section 3.3). Anything else is left unchanged. The raw empirical rejection rates at the 5% nominal level are reported in the first column of Table 2 under the heading "alternative 1". As can be seen, the test that focuses on the common factor has the largest power by far, followed by the dynamic loadings test. In contrast, the reduced form tests have much less power, as we had anticipated in section 3.8 for alternatives such as (27) (see Fiorentini and Sentana (2015) for a related discussion for static factor models). Not surprisingly, the least powerful test looks

at the specific factors, which nevertheless retains some small power because their estimators are affected by the neglected serial correlation in the common factor.

We also simulate and estimate 10,000 samples of the same length as above in which the DGP for the specific factors has $\psi_{u_i}(L) = (1 + .2L)$, for $i = 1, 2, 3$, but the same first-order autocorrelation as under the null, so that

$$\left.\begin{array}{rcl} u_{1,t} &=& -0.418u_{1,t-1} - 0.044u_{1,t-2} + v_{1,t}, \\ u_{2,t} &=& 0.514u_{2,t-1} + 0.143u_{2,t-2} + v_{2,t}, \\ u_{3,t} &=& 0.185u_{3,t-1} + 0.077u_{3,t-2} + v_{3,t}. \end{array}\right\} \tag{28}$$

Again, we re-scale $V(\mathbf{v}_t)$ in order to exactly match the unconditional covariance matrix under the null, leaving everything else unchanged. The results, reported in the second column of Table 2 under the heading "alternative 2", clearly show that the test that focuses on the idiosyncratic factors has the largest power, followed by the joint test. In this case, though, the reduced form tests have reasonable power, as we anticipated in section 3.8 for alternatives such as (28). Finally, the test that focuses on the common factor has power essentially equal to nominal size.

We consider a third design in which the common factor follows the same AR(2) process as under the null, but the static factor loadings are multiplied by heterogeneous MA(1) polynomials $(1 - \psi_{c_i}L)$. Since the unconditional covariance matrix of the resulting model has a two factor structure, it is impossible to adjust the remaining model parameters to achieve the single factor structure of the null model in section 4.1. For that reason, after setting $\psi_{c_1} = .42$, $\psi_{c_2} = .5$ and $\psi_{c_3} = .58$ so that their cross-sectional mean coincides with alternative 1, we scale the loadings as in that alternative to minimise the difference between the two unconditional covariance matrices. The third column of Table 2 displays the results, which indicate that the dynamic loading tests is noticeably more powerful than the rest, while the specific factor test is the least powerful.

Finally, we look at a fourth design which combines features of the previous two. Specifically, we multiply both the static loading and the idiosyncratic term for each series by the heterogenous AR(1) polynomial $(1 - \psi_i L)^{-1}$, with $\psi_1 = .1$, $\psi_2 = .2$ and $\psi_3 = .3$, but keep the common factor as under the null. As we mentioned in section 3.6, this model no longer has a static factor representation with fewer than 3 factors, so it is not possible to replicate the unconditional covariance structure of the null design either. The results in the fourth column of Table 2 show that the diagonal reduced form test is the most powerful, as expected. Nevertheless, the joint test that simultaneously looks at factor loadings and idiosyncratic factors but without imposing that $\psi_{c_i} = \psi_{u_i}$ $\forall i$ also does very well. In contrast, the dynamic specification test for the common factor has very little power.

In summary, our Monte Carlo results clearly indicate that the main advantage of our proposed

Table 2: Power of dynamic misspecification tests
Empirical rejection rates (%) at 5% sigificance level

Alternative	1	2	3	4
Common Factor	47.67	4.78	38.26	6.00
Specific Factors	5.7	48.23	10.07	51.06
Loadings	32.67	10.02	64.68	36.60
All Factors	28.90	44.07	29.42	48.27
Loadings+Specific	24.86	38.1	53.14	60.17
Reduced form	9.51	27.89	38.47	51.10
Diagonal reduced form	8.56	38.08	43.48	67.36

LM tests is that rejections provide a very strong indication of the directions along which the efforts to improve the specification of the model should focus.

5 Empirical illustration

We initially replicate the results in Camacho et al (2015), who construct a monthly US coincident index by combining the indicators of economic activity previously analysed by Stock and Watson (1991), Chauvet (1998) and Chauvet and Pigier (2008). Specifically, they use the industrial production index (IPI), nonfarm payroll employment (EMP), personal income less transfer payments (INC) and real manufacturing and trade sales (SAL). The sample covers the period January 1967 to November 2010 for a total effective sample length of 526 observations. As usual, the seasonally adjusted series are log-transformed and differenced to achieve stationarity. Their basic specification, which naturally contains a single factor, is

$$\begin{bmatrix} IPI_t \\ EMP_t \\ INC_t \\ SAL_t \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} x_t + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \\ u_{4,t} \end{bmatrix},$$

$$x_t = \phi_{x,1}x_{t-1} + \phi_{x,2}x_{t-2} + f_t, \quad u_{i,t} = \phi_{i,1}u_{i,t-1} + \phi_{i,2}u_{i,t-2} + v_{i,t},\ i = 1, \ldots, 4.$$

Each variable is individually standardised, the first two observations are discarded and the scale indeterminacy is eliminated by setting $Var(f_t) = 1$. We report the spectral maximum likelihood estimates of the parameters in Table 3, which are very close to the estimates obtained on the basis of the usual time domain log-likelihood.

Table 3: Spectral maximum likelihood estimates

	x	IPI	EMP	INC	SAL
b_i	-	0.68	0.50	0.28	0.45
ϕ_1	0.43	-0.25	0.24	-0.20	-0.36
ϕ_2	0.22	-0.21	0.52	-0.05	-0.16
σ^2	1	0.27	0.25	0.85	0.59

Camacho et al (2015) argue that many features of the business cycle are better represented by a Markov switching model than by a linear model. In this regard, we proved in appendix C of Fiorentini and Sentana (2013) that a simple two-regime Markov model for the mean of the common factor would generate the autocorrelation structure of an ARMA(1,1) process for x_t, which suggests that their AR(2) specification should be rejected. And indeed it is. Our spectral LM test against first order neglected residual serial correlation in the common factor takes the value of 4.28 with a p-value of 3.9%. The same specification test for all four idiosyncratic factors is 34.01, a large fraction of which comes from the nonfarm payroll employment and personal income series. In turn, a joint test for a first-order lag in the factor loadings yields a highly significant 29.78, mostly due to the real manufacturing and trade sales series. Not surprisingly, the test for first order multivariate serial correlation in the reduced form rejects the model massively. In contrast, the only rejection of the individual first order univariate reduced form test corresponds to the fourth residual.

We then decided to estimate a model with an ARMA(2,1) process for the common factor and MA(1)-type dynamic factor loadings for all series, which led to a very substantial improvement in fit of 62 log-likelihood units with only five additional parameters. Although the common factor test no longer rejects, the test for additional lags in the dynamic factor loadings still does (again, mostly due to the sales series) and the same is true for the idiosyncratic factors test. On this basis, we ended up adding two further lags to the dynamic loadings of real manufacturing and trade sales, one MA root to the employment idiosyncratic factor and another AR root to the income one. The resulting model achieved a further 19 points increase in the log-likelihood and more importantly, it successfully passed all the different dynamic specification tests.

Therefore, our results suggest that Camacho et al (2015) should probably consider a more general Markov switching model, allowing for more flexible dynamics not only in common and idiosyncratic factors but also in the dynamic impact of the common factor on the observed series.

6 Conclusions and extensions

We derive computationally simple expressions for score tests of neglected serial correlation in common and idiosyncratic factors, as well as dynamic misspecification of the factor loadings in dynamic confirmatory factor models using frequency domain methods. Our tests can assess those dynamic characteristics either individually or jointly. Importantly, we interpret the specification tests that we propose in terms of simple to understand moment tests which assess whether certain covariances involving the smoothed values of the latent variables are in line with their theoretical values under the null. We show that the implicit orthogonality conditions are analogous to the conditions obtained by treating the Wiener-Kolmogorov-Kalman smoothed estimators of the

innovations in common and idiosyncratic factors as if they were observed, but they account for their final estimation errors. And although we initially focus on Gaussian factor models with a diagonal idiosyncratic dynamic covariance structure for pedagogical reasons, we relax both these assumptions later on. In particular, we exploit the results in Dunsmuir (1979) to show that our Gaussian tests are robust to nonnormality when the innovations are independent.

We also explicitly relate our proposals to alternative tests based on one-period-ahead prediction errors, which should be white noise under correct dynamic specification. In particular, we express those reduced form tests in terms of homogeneous restrictions on the dynamic factor loadings and idiosyncratic components, explain how to make them robust to parameter uncertainty and study their relative power.

Our simulation results suggest that all the tests that we consider have rather accurate sizes in finite samples both when the innovations in the latent variables are Gaussian, and when they follow Student $t's$, thereby confirming our theoretical results. They also indicate that our proposed model validation tools have power to detect dynamic misspecification, and that they are systematically able to correctly identify the source of the rejection.

Finally, we evaluate the empirical usefulness of our tests by assessing the dynamic factor model used by Camacho et al (2015) to construct a coincident indicator for the US. Once again, our proposals prove very informative for improving the original specification. In particular, our results suggest that adding additional lags to common and specific factors is not enough, being necessary to allow the common factor to dynamically impact the observed series.

Our paper is a reminder that spectral methods for time series are very powerful, and can still be successfully applied to tackle important issues of practical interest. For example, we could exploit the asymptotic orthogonality of the frequency components of the Whittle likelihood to devise suitable bootstrap procedures (see Dahlhaus and Janas (1996) or Kirch and Politis (2011)). A more thorough analysis of the power of our tests using both local power calculations and a more extensive set of Monte Carlo exercises would also be worthwhile.

The extension of our methods to models in which N/T is non-negligible would also constitute a very valuable addition with potentially interesting empirical applications. Doz et al (2012) proved the consistency of the common factor estimators for their true underlying values, while Bai and Li (2016) have also obtained rates of convergence and asymptotic variances under some restrictions.

The relationship between small N models and large N models that nest them is also worth studying. Take for example the model in section 5 and a much larger model that augments it with the so-called Stock and Watson dataset (see e.g. Stock and Watson (2006)). The spectral

density matrix of the quadrivariate model and the 4×4 block of the spectral density matrix of the large N model should be identical if the two models are consistent. Therefore, their differences with respect to the periodogram of those four series should have zero mean under correct specification, the reason being that the periodogram is a rather inneficient but unbiased nonparametric estimator of the true spectral density. As a result, both models can be subject to the spectral moment tests that we have put forward in our paper.

However, their comparison presents non-trivial challenges too. Specifically, although the number of dynamic factors in the high-dimensional model can be consistently estimated using for example the procedure in Amengual and Watson (2007), it cannot necessarily be used in the small N model. The problem is one of identification. The so-called Ledermann bound[10] for static factor models applied on a frequency by frequency basis implies that even if we assume a diagonal idiosyncratic spectral density matrix at all frequencies, we can nonparametrically identify a model with a single common factor at most in a model with four series.

Another challenge is the following. The objective of the Camacho, Pérez-Quirós and Poncela (2015) paper was to come up with a real activity indicator using four carefully selected series observed at the monthly frequency. Assuming one could estimate a model for four series with the number of dynamic factors determined in the large-N model by making rather restrictive parametric assumptions on the dynamics of those common factors and their loadings, as well as the dynamics of the idiosyncratic terms, a sensible way of combining those factors into a single index of real activity would be necessary, possibly along the lines of Forni et (2000) or Altissimo et al (2010).

Finally, it is worth mentioning that although we have exploited some specificities of dynamic factor models, our procedures can be easily extended to most unobserved components time series processes in which a finite dimensional vector of N observed series, $\mathbf{y}_t$, can be recursively defined in the time domain by the system of equations

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{C}(\boldsymbol{\phi})\mathbf{x}_t, \quad \mathbf{x}_t = \mathbf{A}(\boldsymbol{\phi})\mathbf{x}_{t-1} + \mathbf{B}(\boldsymbol{\phi})\mathbf{u}_t, \quad \mathbf{u}_t | I_{t-1}; \boldsymbol{\mu}, \boldsymbol{\phi} \sim N[\mathbf{0}, \boldsymbol{\Omega}(\boldsymbol{\phi})].$$

Such models are the subject of the monographs by Harvey (1989) and Durbin and Koopman (2012), among others, and the list of empirical studies that make use of them is vast. We are currently pursuing some of these research avenues.

[10] As explained in Bekker and ten Berge (1997), the Ledermann bound - $N + k \leq (N - k)^2$ - yields the largest possible number of factors for which a unique decomposition of the covariance matrix is possible.

References

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M. and Veronese, G. (2010), "New eurocoin: tracking economic growth in real time", *Review of Economics and Statistics* 92, 1024-1034.

Amengual, D., Fiorentini, G. and Sentana, E. (2013): "Sequential estimators of shape parameters in multivariate dynamic models", *Journal of Econometrics* 177, 233-249.

Amengual, D. and Watson, M.W. (2007): "Consistent estimation of the number of dynamic factors in a large N and T panel," *Journal of Business and Economic Statistics* 25, 91-96.

Aruoba, B., Diebold, F. and Scotti, C. (2009): "Real-time measurement of business conditions", *Journal of Business and Economic Statistics* 27, 417-427.

Bai, J.. and Li, K. (2016): "Maximum likelihood estimation and inference for approximate factor models of high dimension", *Review of Economics and Statistics* 98, 298-309.

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

Bai, J.. and Wang, P. (2016): "Identification theory for high dimensional static and dynamic factor models", *Journal of Econometrics* 178, 794-804.

Bańbura, M. and Rünstler, G. (2011): "A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP", *International Journal of Forecasting* 27, 333-346.

Bekker, P.A. and ten Berge, J.M.F. (1997): "Generic global indentification in factor analysis", *Linear Algebra and its Applications* 264, 255-263.

Breusch, T. S. (1978): "Testing for autocorrelation in dynamic linear models", *Australian Economic Papers* 17, 334–355.

Brillinger, D. R. (1981): *Time series: data analysis and theory*, Holt, Rinehart and Winston.

Camacho, M., Pérez-Quirós, G. and Poncela, P. (2015): "Extracting nonlinear signals from several economic indicators", *Journal of Applied Econometrics* 30, 1073-1089.

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

Chauvet, M. (1998): "An econometric characterization of business cycle dynamics with factor structure and regime switches", *International Economic Review* 39, 969-96.

Chauvet, M., and Piger, J. (2008): "A comparison of the real-time performance of business cycle dating methods", *Journal of Business and Economic Statistics* 26, 42-49.

Dahlhaus, R. and Janas, D. (1996): "A frequency domain bootstrap for ratio statistics in time series analysis", *Annals of Statistics* 24, 1934-1963.

Del Negro, M. and Otrok, C. (2007): "99 luftballons: monetary policy and the house price boom across U.S. states", *Journal of Monetary Economics* 54, 1962–1985.

Demos, A. and Sentana, E. (1998): "Testing for GARCH effects: a one-sided approach",

Journal of Econometrics 86, 97-127.

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

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

Doz, C., Giannone, D. and Reichlin, L. (2012): "A quasi maximum likelihood approach for large approximate dynamic factor models", *Review of Economics and Statistics* 94, 1014-1024.

Dufour, J.M. and Dagenais, M.G. (1991): "Invariance, nonlinear models and asymptotic tests", *Econometrica* 59, 1601-1615.

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

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

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

Durbin, J. (1970): "Testing for serial correlation in least-squares regressions when some of the regressors are lagged dependent variables", *Econometrica* 38, 410-421.

Durbin, J. and Koopman, S.J. (2012): *Time series analysis by state space methods*, 2nd ed., Oxford University Press.

Dzhaparidze, K. (1986): *Parameter estimation and hypothesis testing in spectral analysis of stationary time series*, Springer-Verlag.

Engle, R.F. (1984): "Wald, likelihood ratio and Lagrange multiplier tests in econometrics", in *Handbook of Econometrics* vol II, ed. Griliches and Intrilligator, 775-826, North Holland.

Engle, R.F. and Watson, M.W. (1980): "Formulation générale et estimation de models multidimensionnels temporels a facteurs explicatifs non observables", *Cahiers du Séminaire d'Économétrie* 22, 109-125.

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

Fernández, F.J. (1990): "Estimation and testing of a multivariate exponential smoothing model", *Journal of Time Series Analysis* 11, 89–105.

Fiorentini, G., Galesi, A. and Sentana, E. (2016): "Fast ML estimation of dynamic bifactor models: an application to European inflation", in S.J. Koopman and E.T. Hillebrand (eds.) *Dynamic factor models*, *Advances in Econometrics* 35, 215-282, Emerald.

Fiorentini, G., Galesi, A. and Sentana, E. (2018): "A spectral EM algorithm for dynamic factor models", *Journal of Econometrics* 205, 249-279.

Fiorentini, G. and Planas, C. (1998): "From autocovariances to moving average: an algorithm comparison", *Computational Statistics* 13, 477-484.

Fiorentini, G. and Sentana, E. (2009): "Dynamic specification tests for static factor models", CEMFI Working Paper 0912.

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

Fiorentini, G. and Sentana, E. (2015): "Tests for serial dependence in static, non-Gaussian factor models", in S.J. Koopman and N. Shephard (eds.) *Unobserved components and time series econometrics*, 118-189, Oxford University Press.

Fiorentini, G. and Sentana, E. (2016): "Neglected serial correlation tests in UCARIMA models", *SERIES* 7, 121-178.

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

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

Forni, M., Hallin, M., Lippi, M. and Zaffaroni, P. (2015): "Dynamic factor models with infinite-dimensional factor spaces: one-sided representations", *Journal of Econometrics* 185, 359-371.

Forni, M. and Lippi, M. (2001): "The generalized dynamic factor model: representation theory", *Econometric Theory* 17, 1113-1141.

Forni, M. and Lippi, M. (2011): "The general dynamic factor model: one-sided representation results", *Journal of Econometrics* 163, 23-28.

French, D. and O'Hare, C. (2013): "A dynamic factor approach to mortality modeling", *Journal of Forecasting* 32, 587–599.

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

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

Godfrey, L. G. (1978): "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables", *Econometrica* 46, 1293–1301.

Godfrey, L.G. (1988): *Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches*, Econometric Society Monographs.

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

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

Guilkey, D.K. (1974): "Alternative tests for a first-order vector autoregressive error specification", *Journal of Econometrics* 4, 95-104.

Hannan, E. J. (1970): *Multiple time series*, Wiley.

Harvey, A.C. (1981): *Time series models*, Philip Allan.

Harvey, A.C. (1982): "A test of misspecification for systems of equations", LSE Econometrics Programme Discussion Paper A31.

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

Harvey, A.C. and Koopman, S.J. (1992): "Diagnostic checking of unobserved components time series models", *Journal of Business & Economic Statistics* 10, 377-389.

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

Hendry, D.F. (1971): "Maximum likelihood estimation of systems of simultaneous regression equations with errors generated by a vector autoregressive process", *International Economic Review* 12, 257-272.

Hosking, J. R. M. (1981): "Lagrange-multiplier tests of multivariate time series models", *Journal of the Royal Statistical Society B* 43. 219-230.

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

Jungbacker, B. and Koopman, S.J. (2015): "Likelihood-based dynamic factor analysis for measurement and forecasting", *Econometrics Journal* 17, 1-21.

Kirch, C. and Politis, D.N. (2011): "TFT-Bootstrap: resampling time series in the frequency domain to obtain replicates in the time domain", *Annals of Statistics* 39, 1427-1470.

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

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

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

Maravall, A. (1999): "Unobserved components in economic time series" in M.H. Pesaran and M.R. Wickens (eds.), *Handbook of Applied Econometrics. Volume I: Macroeconomics*, Blackwell.

Mariano, R.S. and Murasawa, Y. (2003): "A new coincident index of business cycles based on monthly and quarterly series", *Journal of Applied Econometrics* 18, 427-443.

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

Newey, W.K. and McFadden, D.L. (1994): "Large sample estimation and hypothesis testing", in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* vol. IV, 2111-2245, Elsevier.

Newey, W.K. and Powell, J.L. (1998): "Two-step estimation, optimal moment conditions, and sample selection models", mimeo, MIT.

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

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

Sargent, T.J. (1989): "Two models of measurements and the investment accelerator", *Journal of Political Economy* 97, 251–287.

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.

Scherrer, W. and Deistler, M. (1998): "A structure theory for linear dynamic errors-in-variables models", *SIAM Journal on Control and Optimization* 36, 2148-2175.

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

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

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

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

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

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

Stock, J.H. and Watson, M.W. (2006): "Forecasting with many predictors", in G. Elliott, C.W.J. Granger and A. Timmermann (eds.), *Handbook of Forecasting*, vol. 1, Elsevier.

Stock, J.H. and Watson, M.W. (2011): "Dynamic factor models", in M.P. Clements and D.F. Hendry (eds.), *Oxford Handbook of Economic Forecasting*, Oxford University Press.

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

Appendix

A Proofs

A.1 Proposition 1

Given that the reduced form of the diagonal VMA augmented structural model (23) will be

$$\alpha_x(L)\mathbf{A}(L)(\mathbf{y}_t-\boldsymbol{\mu})=[\mathbf{I}-diag(\boldsymbol{\psi}_w)L][\mathbf{A}(L)\mathbf{c}(L)\beta_x(L)f_t+\alpha_x(L)\mathbf{B}(L)\mathbf{v}_t],$$

the spectral scores with respect to $\boldsymbol{\psi}_w$ will be given by the sum of the spectral scores with respect to $\boldsymbol{\psi}_u$ and and $\boldsymbol{\psi}_c$ derived in sections 3.5 and 3.6, respectively, evaluated at $\boldsymbol{\psi}_u=\boldsymbol{\psi}_c=\boldsymbol{\psi}_w$. But those scores are numerically identical to the scores of the alternative diagonal VAR augmented structural model

$$\mathbf{y}_t=\boldsymbol{\mu}+[\mathbf{I}-diag(\boldsymbol{\psi}_w)L]^{-1}\mathbf{c}(L)x_t+\mathbf{u}_t,\quad \alpha_x(L)x_t=\beta_x(L)f_t,\quad [\mathbf{I}-diag(\boldsymbol{\psi}_w)L]\mathbf{A}(L)\mathbf{u}_t=\mathbf{B}(L)\mathbf{v}_t,$$

whose reduced form will be precisely

$$[\mathbf{I}-diag(\boldsymbol{\psi}_w)L]\alpha_x(L)\mathbf{A}(L)(\mathbf{y}_t-\boldsymbol{\mu})=\mathbf{A}(L)\mathbf{c}(L)\beta_x(L)f_t+\alpha_x(L)\mathbf{B}(L)\mathbf{v}_t.$$

Given that the null model is also the same, the tests that correct for parameter uncertainty will coincide. □

A.2 Proposition 2

First of all, it is easy to see that

$$\begin{aligned}\lim_{T\to\infty}V[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\psi}|\boldsymbol{\theta}T}(\boldsymbol{\phi}_0)]&=\lim_{T\to\infty}V[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\psi}T}(\boldsymbol{\phi}_0)-\mathbf{A}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\theta}T}(\boldsymbol{\phi}_0)]\\&=\mathbf{B}_{\boldsymbol{\psi\psi}}+\mathbf{A}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{B}_{\boldsymbol{\theta\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{A}_{\boldsymbol{\psi\theta}}'-\mathbf{B}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{A}_{\boldsymbol{\psi\theta}}'-\mathbf{A}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{B}_{\boldsymbol{\psi\theta}}'.\end{aligned}$$

In turn, the generalised information matrix equality implies that

$$-\lim_{T\to\infty}E[\partial\bar{\mathbf{s}}_{\boldsymbol{\psi}|\boldsymbol{\theta}T}(\boldsymbol{\phi}_0)/\partial\boldsymbol{\psi}]=\lim_{T\to\infty}cov[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\psi}|\boldsymbol{\theta}T}(\boldsymbol{\phi}_0),\sqrt{T}\bar{\mathbf{r}}_{\boldsymbol{\psi}T}(\boldsymbol{\phi}_0)]=\mathbf{A}_{\boldsymbol{\psi\psi}}-\mathbf{A}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{A}_{\boldsymbol{\psi\theta}}',$$

where $\bar{\mathbf{r}}_{\boldsymbol{\psi}T}(\boldsymbol{\phi}_0)$ is the true log-likelihood (average) score. Given that the partitioned inverse formula

$$\begin{aligned}\mathbf{A}^{-1}&=\begin{pmatrix}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}+\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{A}_{\boldsymbol{\psi\theta}}'\mathbf{A}^{\boldsymbol{\psi\psi}}\mathbf{A}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1} & -\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{A}_{\boldsymbol{\psi\theta}}'\mathbf{A}^{\boldsymbol{\psi\psi}}\\ -\mathbf{A}^{\boldsymbol{\psi\psi}}\mathbf{A}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1} & \mathbf{A}^{\boldsymbol{\psi\psi}}\end{pmatrix},\\ \mathbf{A}^{\boldsymbol{\psi\psi}}&=(\mathbf{A}_{\boldsymbol{\psi\psi}}-\mathbf{A}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{A}_{\boldsymbol{\psi\theta}}')^{-1},\end{aligned}$$

implies that

$$\begin{aligned}\mathbf{C}_{\boldsymbol{\psi\psi}}&=\mathbf{A}^{\boldsymbol{\psi\psi}}\mathbf{B}_{\boldsymbol{\psi\psi}}\mathbf{A}^{\boldsymbol{\psi\psi}}+\mathbf{A}^{\boldsymbol{\psi\psi}}\mathbf{A}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{B}_{\boldsymbol{\theta\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{A}_{\boldsymbol{\psi\theta}}'\mathbf{A}^{\boldsymbol{\psi\psi}}\\ -\mathbf{A}^{\boldsymbol{\psi\psi}}\mathbf{B}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{A}_{\boldsymbol{\psi\theta}}'\mathbf{A}^{\boldsymbol{\psi\psi}}&-\mathbf{A}^{\boldsymbol{\psi\psi}}\mathbf{A}_{\boldsymbol{\psi\theta}}\mathbf{A}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{B}_{\boldsymbol{\psi\theta}}'\mathbf{A}^{\boldsymbol{\psi\psi}}=\mathbf{A}^{\boldsymbol{\psi\psi}}\lim_{T\to\infty}V[\sqrt{T}\bar{\mathbf{s}}_{\boldsymbol{\psi}|\boldsymbol{\theta}T}(\boldsymbol{\phi}_0)]\mathbf{A}^{\boldsymbol{\psi\psi}},\end{aligned}$$

the result immediately follows from the maintained assumption that $\mathbf{A}=\mathcal{I}$. □

Supplemental Appendices for

Dynamic specification tests for dynamic factor models

Gabriele Fiorentini

Università di Firenze and RCEA, Viale Morgagni 59, I-50134 Firenze, Italy

<fiorentini@disia.unifi.it>

Enrique Sentana

CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain

<sentana@cemfi.es>

October 2018

B Reduced form of a multivariate AR(p) plus noise

The purpose of this appendix is to find the Wold representation (5) of the VMA(p) component

$$\mathbf{m}_t = \mathbf{c}f_t + (1 - \alpha_1 L - ... - \alpha_p L^p)\mathbf{u}_t,$$

of process (6). Since we can trivially recover x_t from $\mathbf{y}_t$ without error when $|\mathbf{\Gamma}_u| = 0$, we rule this possibility out henceforth. As a result, we can work with the transformed system

$$\mathbf{y}_t^* = \mathbf{\Gamma}_u^{-1/2}\mathbf{y}_t = \mathbf{\Gamma}_u^{-1/2}\mathbf{c}x_t + \mathbf{\Gamma}_u^{-1/2}\mathbf{u}_t = \mathbf{c}^* x_t + \mathbf{u}_t^*,$$

which has the advantage that the covariance matrix of $\mathbf{u}_t^*$ is the identity matrix. Importantly, the diagonality of $\mathbf{\Gamma}_u$ plays no role in this transformation. Similarly, given that we are focusing on second order properties of the observable process, normality will play no role either.

In this context, our goal is to obtain the invertible reduced form representation

$$\mathbf{m}_t^* = [\mathbf{\Gamma}_u^{-1/2}\mathbf{D}(L)\mathbf{\Gamma}_u^{1/2}]\mathbf{\Gamma}_u^{-1/2}\mathbf{w}_t = \mathbf{D}^*(L)\mathbf{w}_t^* = (\mathbf{I}_N + \mathbf{D}_1^* L + \ldots + \mathbf{D}_p^* L^p)\mathbf{w}_t^*,$$

with $\mathbf{w}_t^*|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots \sim N(\mathbf{0}, \mathbf{\Sigma}^*)$. Having done so, we can easily recover $\mathbf{\Sigma} = \mathbf{\Gamma}_u^{1/2}\mathbf{\Sigma}^*\mathbf{\Gamma}_u^{1/2}$ and $\mathbf{D}_i = \mathbf{\Gamma}_u^{1/2}\mathbf{D}_i^*\mathbf{\Gamma}_u^{-1/2}$ for $i = 1, \ldots, p$.

As we mentioned in section 2.2, $\mathbf{m}_t^*$ also has a dynamic factor structure, although in this case the common factor is white noise while the specific factors follow a VMA(p) process with scalar polynomial $(1 - \alpha_1 L - ... - \alpha_p L^p)\mathbf{I}_N$. Thus, the autocovariance matrices of $\mathbf{m}_t^*$ will be:

$$V(\mathbf{m}_t^*) = \gamma_f \mathbf{c}^*\mathbf{c}^{*\prime} + \gamma(0)\mathbf{I}_N;\ cov(\mathbf{m}_t^*, \mathbf{m}_{t-j}^*) = \gamma(j)\mathbf{I}_N,\ j = 1, \ldots, p;\ cov(\mathbf{m}_t^*, \mathbf{m}_{t-j}^*) = 0 \quad j > p,$$

where $\gamma(0), \gamma(1), \ldots, \gamma(p)$ are the autocovariances of a univariate MA(p) process with polynomial $(1 - \alpha_1 L - ... - \alpha_p L^p)$ and standardised innovations. But we also know that

$$\begin{aligned} V(\mathbf{m}_t^*) &= \mathbf{\Sigma}^* + \mathbf{D}_1^*\mathbf{\Sigma}^*\mathbf{D}_1^{*\prime} + \ldots + \mathbf{D}_p^*\mathbf{\Sigma}^*\mathbf{D}_p^{*\prime}, \\ cov(\mathbf{m}_t^*, \mathbf{m}_{t-1}^*) &= \mathbf{D}_1^*\mathbf{\Sigma}^* + \ldots + \mathbf{D}_p^*\mathbf{\Sigma}^*\mathbf{D}_{p-1}^{*\prime}, \ldots, cov(\mathbf{m}_t^*, \mathbf{m}_{t-p}^*) = \mathbf{D}_p^*\mathbf{\Sigma}^*, \end{aligned}$$

so we can obtain the required reduced form coefficients by matching the structural and reduced form expressions for the autocovariance matrices of $\mathbf{m}_t^*$. There are several well-known methods for solving the resulting equations in the univariate case (see Fiorentini and Planas (1998) for a comparison), but the task is far more daunting in the multivariate context. Nevertheless, the dynamic factor structure imposes many restrictions that we can successfully exploit.

First of all, the one-period ahead forecast errors of $\mathbf{m}_t^*$ based on its past values alone coincide with the one-period ahead forecast errors in $\mathbf{y}_t^*$ given the past of the observed series. In turn, it is easy to see that the state space representation of $\mathbf{y}_t$ implies that the covariance matrix of the one-period ahead forecasting errors of $\mathbf{y}_t^*$ based on its entire past history will have a restricted single factor structure regardless of p (see appendix A in Fiorentini and Sentana (2013) for $p = 2$). Therefore, we can safely conclude that

$$\mathbf{\Sigma}^* = a_0\mathbf{c}^*\mathbf{c}^{*\prime} + b_0\mathbf{I}_N.$$

On this basis, we begin by conjecturing that

$$\mathbf{D}_1^* = a_1\mathbf{c}^*\mathbf{c}^{*\prime} + b_1\mathbf{I}_N, \ldots, \mathbf{D}_p^* = a_p\mathbf{c}^*\mathbf{c}^{*\prime} + b_p\mathbf{I}_N,$$

where $a_0, b_0, \ldots a_p, b_p$ are unknown scalars to be determined, and then verify our conjecture. As an illustration, suppose $p = 2$, in which case the system of equations becomes

$$\begin{aligned}\boldsymbol{\Sigma}^*+\mathbf{D}_1^*\boldsymbol{\Sigma}^*\mathbf{D}_1^{*\prime}+\mathbf{D}_2^*\boldsymbol{\Sigma}^*\mathbf{D}_2^{*\prime} &= \gamma_f\mathbf{c}^*\mathbf{c}^{*\prime}+\gamma(0)\mathbf{I}_N;\ \mathbf{D}_1^*\boldsymbol{\Sigma}^*+\mathbf{D}_2^*\boldsymbol{\Sigma}^*\mathbf{D}_1^{*\prime}=\gamma(1)\mathbf{I}_N;\ \mathbf{D}_2^*\boldsymbol{\Sigma}^*=\gamma(2)\mathbf{I}_N,\\ \gamma(0) &= 1+\alpha_1^2+\alpha_2^2;\ \gamma(1)=-\alpha_1(1-\alpha_2);\ \gamma(2)=-\alpha_2.\end{aligned}$$

The last matrix equation immediately implies that

$$\mathbf{D}_2^* = \gamma(2)\boldsymbol{\Sigma}^{*-1} = \frac{\gamma(2)}{b_0}\mathbf{I}_N - \frac{\gamma(2)}{b_0^2(a_0^{-1}+b_0^{-1}|\mathbf{c}^*|^2)}\mathbf{c}^*\mathbf{c}^{*\prime},$$

where $|\mathbf{c}^*|^2 = \mathbf{c}^{*\prime}\mathbf{c}^* = \mathbf{c}'\boldsymbol{\Gamma}_u^{-1}\mathbf{c}$, which means that

$$a_2 = -\frac{\gamma(2)}{b_0^2(a_0^{-1}+b_0^{-1}|\mathbf{c}^*|^2)} = b_2\frac{\gamma(2)}{b_0a_0^{-1}+|\mathbf{c}^*|^2}, \quad b_2 = \frac{\gamma(2)}{b_0}.$$

If we then replace $\mathbf{D}_2^*$ by this expression in the equation for $cov(\mathbf{m}_t^*, \mathbf{m}_{t-1}^*)$, we end up with

$$\mathbf{D}_1^*\boldsymbol{\Sigma}^* + \gamma(2)\mathbf{D}_1^{*\prime} = \gamma(1)\mathbf{I}_N.$$

But $\mathbf{D}_1^*\boldsymbol{\Sigma}^* = (a_1\mathbf{c}^*\mathbf{c}^{*\prime} + b_1\mathbf{I}_N)(a_0\mathbf{c}^*\mathbf{c}^{*\prime} + b_0\mathbf{I}_N) = (a_1a_0|\mathbf{c}^*|^2 + a_1b_0 + a_0b_1)\mathbf{c}^*\mathbf{c}^{*\prime} + b_1b_0\mathbf{I}_N$, whence $(a_1a_0|\mathbf{c}^*|^2 + a_1b_0 + a_0b_1 + \gamma(2)a_1)\mathbf{c}^*\mathbf{c}^{*\prime} + b_1(b_0+\gamma(2))\mathbf{I}_N = \gamma(1)\mathbf{I}_N$, which leads to the equations

$$b_1 = \frac{\gamma(1)}{b_0+\gamma(2)}, \ a_1 = \frac{-a_0b_1}{a_0|\mathbf{c}^*|^2+b_0+\gamma(2)}.$$

Finally, the equation for the covariance matrix of $\mathbf{m}_t^*$ becomes

$$\begin{aligned}&a_0\mathbf{c}^*\mathbf{c}^{*\prime} + b_0\mathbf{I}_N + [a_1^2a_0|\mathbf{c}^*|^4 + (a_1^2b_0 + 2a_1a_0b_1 + a_1b_1b_0)]|\mathbf{c}^*|^2 + 2a_1b_1b_0 + a_0b_1^2]\mathbf{c}^*\mathbf{c}^{*\prime}\\ &+b_1^2b_0\mathbf{I}_N + \frac{\gamma^2(2)}{b_0}\mathbf{I}_N - \frac{\gamma^2(2)}{b_0^2(a_0^{-1}+b_0^{-1}|\mathbf{c}^*|^2)}\mathbf{c}^*\mathbf{c}^{*\prime},\end{aligned}$$

where we have exploited the fact that

$$\begin{aligned}\mathbf{D}_1^*\boldsymbol{\Sigma}^*\mathbf{D}_1^{*\prime} &= [(a_1a_0|\mathbf{c}^*|^2 + a_1b_0 + a_0b_1)\mathbf{c}^*\mathbf{c}^{*\prime} + b_1b_0\mathbf{I}_N](a_1\mathbf{c}^*\mathbf{c}^{*\prime} + b_1\mathbf{I}_N)\\ &= [a_1^2a_0|\mathbf{c}^*|^4 + (a_1^2b_0 + 2a_1a_0b_1 + a_1b_1b_0)]|\mathbf{c}^*|^2 + 2a_1b_1b_0 + a_0b_1^2]\mathbf{c}^*\mathbf{c}^{*\prime} + b_1^2b_0\mathbf{I}_N.\end{aligned}$$

This expression leads to the following two scalar equations

$$\begin{aligned}a_0 + a_1^2a_0|\mathbf{c}^*|^4 + (a_1^2b_0 + 2a_1a_0b_1 + a_1b_1b_0)]|\mathbf{c}^*|^2 + 2a_1b_1b_0 + a_0b_1^2 - \frac{\gamma^2(2)}{b_0^2(a_0^{-1}+b_0^{-1}|\mathbf{c}^*|^2)} &= \gamma_f,\\ b_0 + b_1^2b_0 + \frac{\gamma^2(2)}{b_0} &= \gamma(0),\end{aligned}$$

which can be used in combination with the expressions for a_1 and b_1 to find all the necessary parameters. Importantly, we must choose the solution to this system of equations that renders the reduced form process invertible, but this is easy to verify.

An extension of this procedure for higher values of p is tedious but straightforward.

C Reduced form tests

Given the local asymptotic equivalence between VAR and VMA alternatives, for simplicity, but without loss of generality, in this appendix we generalise Proposition 1 by focusing on testing the null hypothesis that the reduced form residuals $\mathbf{w}_t$ in (5) are serially uncorrelated against the alternative that they follow a VMA(1) process by considering the following structural model

$$\mathbf{y}_t = \boldsymbol{\mu}+\mathbf{A}^{-1}(L)(\mathbf{I}-\boldsymbol{\Psi}_w L)\mathbf{A}(L)\mathbf{c}(L)x_t+\mathbf{u}_t, \quad \alpha_x(L)x_t = \beta_x(L)f_t, \quad \mathbf{A}(L)\mathbf{u}_t = (\mathbf{I}-\boldsymbol{\Psi}_w L)\mathbf{B}(L)\mathbf{v}_t,$$

whose reduced form will be

$$\alpha_x(L)\mathbf{A}(L)(\mathbf{y}_t - \boldsymbol{\mu}) = (\mathbf{I} - \boldsymbol{\Psi}_w L)[\mathbf{A}(L)\mathbf{c}(L)\beta_x(L)f_t + \alpha_x(L)\mathbf{B}(L)\mathbf{v}_t].$$

Under this alternative, the spectral density matrix becomes

$$\begin{aligned}\mathbf{G}_{\mathbf{yy}}(\lambda) &= \mathbf{A}^{-1}(e^{-i\lambda})(\mathbf{I} - \boldsymbol{\Psi}_w e^{-i\lambda})\mathbf{A}(e^{-i\lambda})\mathbf{c}(e^{-i\lambda})g_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{A}(e^{i\lambda})(\mathbf{I} - \boldsymbol{\Psi}'_w e^{i\lambda})\mathbf{A}^{-1}(e^{i\lambda}) \\ &\quad +\mathbf{A}^{-1}(e^{-i\lambda})(\mathbf{I} - \boldsymbol{\Psi}_w e^{-i\lambda})\mathbf{B}(e^{-i\lambda})\boldsymbol{\Sigma}_{vv}\mathbf{B}(e^{i\lambda})(\mathbf{I} - \boldsymbol{\Psi}'_w e^{i\lambda})\mathbf{A}^{-1}(e^{i\lambda}).\end{aligned}$$

Since we already have all the other gradients under the null, we assume that all parameters except $\boldsymbol{\Psi}_w$ are known, in which case the differential of $\mathbf{G}_{\mathbf{yy}}(\lambda)$ will be given by

$$\begin{aligned}d\mathbf{G}_{\mathbf{yy}}(\lambda) = &-\mathbf{A}^{-1}(e^{-i\lambda})e^{-i\lambda}d\boldsymbol{\Psi}_w\mathbf{A}(e^{-i\lambda})\mathbf{c}(e^{-i\lambda})g_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{A}(e^{i\lambda})(\mathbf{I} - \boldsymbol{\Psi}'_w e^{i\lambda})\mathbf{A}^{-1}(e^{i\lambda}) \\ &-\mathbf{A}^{-1}(e^{-i\lambda})(\mathbf{I} - \boldsymbol{\Psi}_w e^{-i\lambda})\mathbf{A}(e^{-i\lambda})\mathbf{c}(e^{-i\lambda})g_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{A}(e^{i\lambda})d\boldsymbol{\Psi}'_w e^{i\lambda}\mathbf{A}^{-1}(e^{i\lambda}) \\ &-\mathbf{A}^{-1}(e^{-i\lambda})e^{-i\lambda}d\boldsymbol{\Psi}_w\mathbf{B}(e^{-i\lambda})\boldsymbol{\Sigma}_{vv}\mathbf{B}(e^{i\lambda})(\mathbf{I} - \boldsymbol{\Psi}'_w e^{i\lambda})\mathbf{A}^{-1}(e^{i\lambda}) \\ &-\mathbf{A}^{-1}(e^{-i\lambda})(\mathbf{I} - \boldsymbol{\Psi}_w e^{-i\lambda})\mathbf{B}(e^{-i\lambda})\boldsymbol{\Sigma}_{vv}\mathbf{B}(e^{i\lambda})d\boldsymbol{\Psi}'_w e^{i\lambda}\mathbf{A}^{-1}(e^{i\lambda}).\end{aligned}$$

Hence, we obtain that $dvec[\mathbf{G}_{\mathbf{yy}}(\lambda)]$ will be given by

$$\begin{aligned}&-\mathbf{K}_{NN}[\mathbf{A}^{-1}(e^{-i\lambda}) \otimes \mathbf{A}^{-1}(e^{i\lambda})(\mathbf{I} - \boldsymbol{\Psi}_w e^{i\lambda})\mathbf{A}(e^{i\lambda})\mathbf{c}(e^{i\lambda})g_{xx}(\lambda)\mathbf{c}'(e^{-i\lambda})\mathbf{A}(e^{-i\lambda})]e^{-i\lambda}dvec(\boldsymbol{\Psi}'_w) \\ &-[\mathbf{A}^{-1}(e^{i\lambda}) \otimes \mathbf{A}^{-1}(e^{-i\lambda})(\mathbf{I} - \boldsymbol{\Psi}_w e^{-i\lambda})\mathbf{A}(e^{-i\lambda})\mathbf{c}(e^{-i\lambda})g_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{A}(e^{i\lambda})]e^{i\lambda}dvec(\boldsymbol{\Psi}'_w) \\ &-\mathbf{K}_{NN}[\mathbf{A}^{-1}(e^{-i\lambda}) \otimes \mathbf{A}^{-1}(e^{i\lambda})(\mathbf{I} - \boldsymbol{\Psi}_w e^{i\lambda})\mathbf{B}(e^{i\lambda})\boldsymbol{\Sigma}_{vv}\mathbf{B}(e^{-i\lambda})]e^{-i\lambda}dvec(\boldsymbol{\Psi}'_w) \\ &-[\mathbf{A}^{-1}(e^{i\lambda}) \otimes \mathbf{A}^{-1}(e^{-i\lambda})(\mathbf{I} - \boldsymbol{\Psi}_w e^{-i\lambda})\mathbf{B}(e^{-i\lambda})\boldsymbol{\Sigma}_{vv}\mathbf{B}(e^{i\lambda})]e^{i\lambda}dvec(\boldsymbol{\Psi}'_w),\end{aligned}$$

where $\mathbf{K}_{NN}$ is the commutation matrix of orders (N, N) such that $vec(\boldsymbol{\Psi}_w) = \mathbf{K}_{NN}vec(\boldsymbol{\Psi}'_w)$. As a result, the Jacobian of $vec[\mathbf{G}_{\mathbf{yy}}(\lambda)]$ with respect to $vec(\boldsymbol{\Psi}'_w)$ at $\boldsymbol{\Psi}_w = \mathbf{0}$ will be

$$\frac{dvec[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{dvec'(\boldsymbol{\Psi}'_w)} = -\mathbf{K}_{NN}[\mathbf{A}^{-1}(e^{-i\lambda}) \otimes \mathbf{G}'_{\mathbf{yy}}(\lambda)\mathbf{A}(e^{-i\lambda})]e^{-i\lambda} - [\mathbf{A}^{-1}(e^{i\lambda}) \otimes \mathbf{G}_{\mathbf{yy}}(\lambda)\mathbf{A}(e^{i\lambda})]e^{i\lambda},$$

where we have used the fact that

$$\mathbf{A}^{-1}(e^{-i\lambda})\mathbf{B}(e^{-i\lambda})\boldsymbol{\Sigma}_{vv}\mathbf{B}(e^{i\lambda}) = \mathbf{G}_{\mathbf{uu}}(\lambda)\mathbf{A}(e^{i\lambda}), \quad \mathbf{A}^{-1}(e^{i\lambda})\mathbf{B}(e^{i\lambda})\boldsymbol{\Sigma}_{vv}\mathbf{B}(e^{-i\lambda}) = \mathbf{G}'_{\mathbf{uu}}(\lambda)\mathbf{A}(e^{-i\lambda}).$$

Given that $\mathbf{A}(e^{-i\lambda})$ and $\mathbf{A}^{-1}(e^{-i\lambda})$ are diagonal matrices, the required Kronecker products adopt particularly simple forms. Finally, the advantage of working with $dvec(\boldsymbol{\Psi}'_w)$ instead of $dvec(\boldsymbol{\Psi}_w)$ is that we can easily test for neglected serial correlation in a single series if desired.

D Asymptotic distribution of the spectral ML estimators

In this appendix we formally derive the asymptotic distribution of the spectral maximum likelihood estimators of the dynamic factor model parameters on the basis of the results in Dunsmuir (1979), who made the following three assumptions on the spectral matrix

1. $\mathbf{G_{yy}}(\lambda;\boldsymbol{\phi})$ is positive definite for all frequencies and all values of $\boldsymbol{\phi}$ in the admissible parameter space $\boldsymbol{\Phi} \subseteq \mathbb{R}^d$, a twice differentiable manifold of dimension $d < \infty$, and $\boldsymbol{\phi}_0 \in int(\boldsymbol{\Phi})$ is locally identified.
2. $\mathbf{G_{yy}}(\lambda;\boldsymbol{\phi})$ is twice continuously differentiable with respect to $\boldsymbol{\phi}$, and those second derivatives are continuous in λ.
3. The elements of $\mathbf{G_{yy}}(\lambda;\boldsymbol{\phi})$ belong to the Lipschitz class of order α, with $1/2 < \alpha \leq 1$.

and the following four assumptions on the vector of $N+1$ latent innovations $\boldsymbol{\xi}_t$

4.1 $E(\boldsymbol{\xi}_t|I_{t-1}) = \mathbf{0}$ a.s.

4.2 $V(\boldsymbol{\xi}_t|I_{t-1}) = \boldsymbol{\Gamma}$ a.s.

4.3 $E[vec(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') \otimes \boldsymbol{\xi}'|I_{t-1}] = \boldsymbol{\Psi}$ a.s.

4.4 $E[vec(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') \otimes vec'(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')] = (\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})(\mathbf{I}_{(N+1)^2} + \mathbf{K}_{N+1,N+1}) + vec(\boldsymbol{\Gamma})vec'(\boldsymbol{\Gamma}) + \boldsymbol{\Upsilon}$.

As long as the identification conditions discussed in section 3.1 are satisfied, the dynamic factor model in (1) will fulfill conditions 1, 2 and 3 because $\mathbf{G_{yy}}(\lambda;\boldsymbol{\phi})$ is a linear combination of the rational spectral densities of the underlying univariate ARMA models. As for assumptions 4.1-4.4, we impose them by design in the Monte Carlo experiments in section 4. Thus, we can apply the generalised version of Theorem 2.1 in Dunsmuir (1979), § 3, p. 502, to prove that

$$\begin{aligned}
\sqrt{T}\bar{\mathbf{s}}_{\phi T}(\boldsymbol{\phi}_0) &\rightarrow N(\mathbf{0},\mathbf{B}_0),\\
\sqrt{T}(\boldsymbol{\phi}_T - \boldsymbol{\phi}_0) &\rightarrow N(\mathbf{0},\mathbf{C}_0),\\
\mathbf{C}_0 &= \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1},\\
\mathbf{A}_0 &= -p\lim_{T\rightarrow\infty}\partial\bar{\mathbf{s}}_{\phi T}(\boldsymbol{\phi}_0)/\partial\boldsymbol{\phi}'.
\end{aligned}$$

Before providing detailed expressions for $\mathbf{A}$ and $\mathbf{B}$, though, let us highlight some inconsequential but potentially confusing differences in notational conventions between Dunsmuir's paper and ours. First of all, he does not divide the spectral log-likelihood function by 2, so that

$$\mathbf{A} = \frac{1}{2}\boldsymbol{\Omega}, \quad \mathbf{B} = \frac{1}{4}(2\boldsymbol{\Omega} + \boldsymbol{\Pi}) = \mathbf{A} + \frac{1}{4}\boldsymbol{\Pi}.$$

In addition, he defines the periodogram as

$$\frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_s - \boldsymbol{\mu})'e^{i(t-s)\lambda_j} = 2\pi\mathbf{z}_j'\mathbf{z}_j^c$$

and the spectral density matrix as $E(2\pi\mathbf{z}_j'\mathbf{z}_j^c)$, which means that what we call $\mathbf{G_{yy}}(\lambda;\boldsymbol{\phi})$ following e.g. Harvey (1981, p. 91), is the (simple) transpose of his spectral density and what we have called $\mathbf{I_{yy}}(\lambda_j)$ is the transpose of his periodogram. Finally, he considers frequencies in the interval $(-\pi,\pi)$ while we look at $(0,2\pi)$.

In our notation, Dunsmuir (1979) expression for the $(j,k)^{th}$ element of $\mathbf{\Omega}$ is

$$\begin{aligned}\mathbf{\Omega}_{jl} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\{\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda;\boldsymbol{\phi})[\partial\mathbf{G}_{\mathbf{yy}}'(\lambda;\boldsymbol{\phi})/\partial\phi_j]\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda;\boldsymbol{\phi})[\partial\mathbf{G}_{\mathbf{yy}}'(\lambda;\boldsymbol{\phi})/\partial\phi_k]\}d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} vec'\{[\partial\mathbf{G}_{\mathbf{yy}}'(\lambda;\boldsymbol{\phi})/\partial\phi_j]\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda;\boldsymbol{\phi})\}vec\{\left[\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda;\boldsymbol{\phi})[\partial\mathbf{G}_{\mathbf{yy}}'(\lambda;\boldsymbol{\phi})/\partial\phi_k]\right]\}\right] d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} vec'[\partial\mathbf{G}_{\mathbf{yy}}'(\lambda;\boldsymbol{\phi})/\partial\phi_k]\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda;\boldsymbol{\phi})\otimes\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda;\boldsymbol{\phi})\right]\mathbf{K}_{NN}vec[\partial\mathbf{G}_{\mathbf{yy}}'(\lambda;\boldsymbol{\phi})/\partial\phi_k]d\lambda.\end{aligned}$$

Given that $\partial vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\right]/\partial\phi_k$ is the k^{th} column of $\partial vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\right]/\partial\boldsymbol{\phi}'$, while the j^{th} row of $\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\right]/\partial\boldsymbol{\phi}$ is $\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\right]/\partial\phi_j$, we can write

$$\mathbf{\Omega} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\partial vec'\left[\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\right]/\partial\boldsymbol{\phi}\left[\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda;\boldsymbol{\phi})\otimes\mathbf{G}_{\mathbf{yy}}^{-1\prime}(\lambda;\boldsymbol{\phi})\right]\mathbf{K}_{NN}\partial vec\left[\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\right]/\partial\boldsymbol{\phi}'d\lambda.$$

The Hermitian nature of $\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})$ implies that $\mathbf{\Omega}$ coincides with $2\mathcal{I}(\boldsymbol{\phi})$ in (16).

Let us now move on to $\mathbf{\Pi}$ for the dynamic single factor model in (1), but replacing the normality assumption by conditions 4.1-4.4. To do so, it is convenient to write the observed series as in (2), so that their spectral density matrix will be

$$\begin{aligned}\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi}) = \mathbf{\Delta}(e^{-i\lambda})\mathbf{\Gamma}\mathbf{\Delta}'(e^{i\lambda}) = \mathbf{c}(e^{-i\lambda})\frac{\beta_x(e^{-i\lambda})}{\alpha_x(e^{-i\lambda})}\gamma_f\frac{\beta_x(e^{i\lambda})}{\alpha_x(e^{i\lambda})}\mathbf{c}'(e^{i\lambda}) \\ +diag\left[\frac{\beta_1(e^{-i\lambda})}{\alpha_1(e^{-i\lambda})}\gamma_{v_1}\frac{\beta_1(e^{i\lambda})}{\alpha_1(e^{i\lambda})}, \frac{\beta_2(e^{-i\lambda})}{\alpha_2(e^{-i\lambda})}\gamma_2\frac{\beta_2(e^{i\lambda})}{\alpha_2(e^{i\lambda})}, \ldots \frac{\beta_N(e^{-i\lambda})}{\alpha_N(e^{-i\lambda})}\gamma_{v_N}\frac{\beta_N(e^{i\lambda})}{\alpha_N(e^{i\lambda})}\right].\end{aligned}$$

As stated in condition 4.4, the $(1+N)^2\times(1+N)^2$ matrix of fourth-order cumulants $\mathbf{\Upsilon}$ is the difference between $E[vec(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')vec'(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')]$ and its value under normality, which is $(\mathbf{\Gamma}\otimes\mathbf{\Gamma})(\mathbf{I}_{(N+1)^2}+\mathbf{K}_{N+1,N+1})+vec(\mathbf{\Gamma})vec'(\mathbf{\Gamma})$. For example, in the case of $N=2$ the fourth-order cumulant matrix is 9×9 with typical element $\upsilon_{abcd} = E(\xi_a\xi_b\xi_c\xi_d) - E(\xi_a\xi_b)E(\xi_c\xi_d) - E(\xi_a\xi_c)E(\xi_b\xi_d) - E(\xi_a\xi_d)E(\xi_b\xi_c)$.

In addition to the multivariate Gaussian case, in which all fourth order cumulants are 0, closed-form expressions for $\mathbf{\Upsilon}$ can be obtained in some other interesting cases. Specifically, if we follow section 4 in Dunsmuir (1979) in assuming that the elements of $\boldsymbol{\xi}_t$ are stochastically independent, the only non-zero elements of $\mathbf{\Upsilon}$ are $\upsilon_{ff,ff}$, $\upsilon_{11,11}$ and $\upsilon_{22,22}$, whose values coincide with the univariate fourth-order marginal cumulants of the corresponding series.

In our notation, Dunsmuir's (1979) expression for the $(j,k)^{th}$ element of $\mathbf{\Pi}$ is

$$\mathbf{\Pi}_{jk} = \sum\nolimits_{a=1}^{1+N}\sum\nolimits_{b=1}^{1+N}\sum\nolimits_{c=1}^{1+N}\sum\nolimits_{d=1}^{1+N}\upsilon_{abcd}\mathbf{\Phi}_{ab}^{(j)}\mathbf{\Phi}_{cd}^{(k)},$$

where $\mathbf{\Phi}_{ab}^{(j)}$ denotes the $(a,b)^{th}$ element of the $(1+N)\times(1+N)$ matrix

$$\mathbf{\Phi}^{(j)} = \int_{-\pi}^{\pi}\mathbf{\Delta}'(e^{-i\lambda})[\partial\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda;\boldsymbol{\phi})/\partial\phi_j]\mathbf{\Delta}(e^{i\lambda})d\lambda.$$

Tedious algebra shows that

$$\begin{aligned}\mathbf{\Pi}_{jk} &= \sum\nolimits_{a=1}^{1+N}\sum\nolimits_{b=1}^{1+N}\sum\nolimits_{c=1}^{1+N}\sum\nolimits_{d=1}^{1+N}\upsilon_{abcd}\mathbf{\Phi}_{ab}^{(j)}\mathbf{\Phi}_{cd}^{(k)} = vec'[\mathbf{\Phi}^{(j)}]\mathbf{\Upsilon}vec[\mathbf{\Phi}^{(k)}], \\ vec[\mathbf{\Phi}^{(j)}] &= \int_{-\pi}^{\pi} vec\left[\mathbf{\Delta}'(e^{-i\lambda})[\partial\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda;\boldsymbol{\phi})/\partial\phi_j]\mathbf{\Delta}(e^{i\lambda})\right]d\lambda \\ &= \int_{-\pi}^{\pi}\left[\mathbf{\Delta}'(e^{i\lambda})\otimes\mathbf{\Delta}'(e^{-i\lambda})\right]\{\partial vec[\mathbf{G}_{\mathbf{yy}}'^{-1}(\lambda;\boldsymbol{\phi})]/\partial\phi_j\}d\lambda.\end{aligned}$$

But since $d\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi}) = -\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})d\mathbf{G}'_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})$, we can write

$$\begin{aligned} vec[\boldsymbol{\Phi}^{(j)}] &= -\int_{-\pi}^{\pi}\left[\boldsymbol{\Delta}'(e^{i\lambda})\otimes\boldsymbol{\Delta}'(e^{-i\lambda})\right]\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\right]\frac{\partial vec[\mathbf{G}'_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})]}{\partial\phi_j}d\lambda \\ &= -\int_{-\pi}^{\pi}\left[\boldsymbol{\Delta}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\otimes\boldsymbol{\Delta}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\right]\mathbf{K}_{NN}\frac{\partial vec[\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})]}{\partial\phi_j}d\lambda. \end{aligned}$$

Therefore, we can finally write

$$\begin{aligned} \boldsymbol{\Pi} = &\int_{-\pi}^{\pi}\{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})]/\partial\boldsymbol{\phi}\}\left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\boldsymbol{\Delta}(e^{-i\lambda})\otimes\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\boldsymbol{\Delta}(e^{i\lambda})\right]d\lambda \\ \times\boldsymbol{\Upsilon}\times&\int_{-\pi}^{\pi}\left[\boldsymbol{\Delta}'(e^{-i\lambda})\mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\otimes\boldsymbol{\Delta}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})\right]\{\partial vec[\mathbf{G}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})]/\partial\boldsymbol{\phi}'\}d\lambda \end{aligned}$$

because

$$\mathbf{K}_{N+1,N+1}\boldsymbol{\Upsilon}\mathbf{K}_{N+1,N+1}=\mathbf{K}_{N+1,N+1}E[vec(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')\otimes vec'(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')]\mathbf{K}_{N+1,N+1}=E[vec(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')\otimes vec'(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')].$$

The $\boldsymbol{\Phi}^{(j)}$ matrices simplify considerably in restricted VARMA models with no latent variables because the matrix $\boldsymbol{\Delta}(L)$ is square and the integrals of the derivatives of the spectral density with respect to the dynamic parameters are all 0. In the general case, we can once again use the Woodbury formula in (10) to express $\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda;\boldsymbol{\phi})$ in terms of its constituents under the assumption that neither $G_{xx}(\lambda)$ nor $\mathbf{G}_{\mathbf{uu}}(\lambda)$ are singular at any frequency.