Validating DSGE models with VARs and Large-Dimensional Dynamic Factor Models.

Marco Lippi*

February 22, 2019

Abstract

A popular validation procedure for DSGE models consists in comparing the structural shocks and impulse-response functions obtained by estimation-calibration of the DSGE with those obtained in a SVAR identified by means of some of the DSGE restrictions. The variables included in the VAR are all or part of the variables in the DSGE, usually augmented with measurement errors. The paper argues that this practice can be seriously misleading. For generic values of the parameters of the DSGE, the shocks estimated in the VAR are not "made of" the corresponding structural shocks plus measurement error. Rather, each of the VAR shocks is contaminated by non-corresponding structural shocks. We argue that Large-Dimensional Dynamic Factor Models are free from this drawback and are the natural model to use in validation procedures for DSGEs.

JEL classification: C32, C55, E32.

Keywords: Dynamic Stochastic General Equilibrium, SVAR, Large-Dimensional Dynamic Factor Models, Validation of Dynamic Stochastic General Equilibrium models.

*mlippi.eief@gmail.com – Einaudi Institute for Economics and Finance, Roma, Italy.

1 Introduction

The present paper argues against the use of SVARs for validation of Dynamic Stochastic General Equilibrium (DSGE) models. We show that this practice can be seriously misleading. For generic values of the parameters of the DSGE, the shocks estimated in the SVAR are not "made of" the corresponding structural shocks plus measurement error. Rather, each of the VAR shocks is contaminated by non-corresponding structural shocks. We argue that Large-Dimensional Dynamic Factor Models (DFM) are free from this drawback and are the natural model to use in validation procedures for DSGEs.

Our negative argument, regarding VAR models, can be illustrated as follows. Let the DSGE consist of only one variable y_t, one unit-variance shock u_t and the equation

$$y_t = (2.5 + 1.2\,L)v_t, \tag{1}$$

and suppose that y_t is measured with an error η_t, which is a white noise process with $\sigma_\eta^2 = 2.31$, orthogonal to the white noise v_t at all leads and lags, so that we observe

$$x_t = (2.5 + 1.2\,L)v_t + \eta_t. \tag{2}$$

Elementary time-series theory shows that

$$x_t = (3 + L)V_t, \tag{3}$$

where V_t is a unit-variance white noise. Now, what is V_t? For example, if y_t is the rate of change of productivity and v_t the technology shock, can we say that V_t is just $v_t + e\eta_t$ for some e, so that we can can claim that, after all, V_t is the technology shock with a measurement error? The answer is an emphatic no. From (2) and (3) we obtain

$$V_t = \frac{2.5 + 1.2\,L}{3 + L} v_t + \frac{1}{3 + L}\eta_t. \tag{4}$$

Thus V_t is a moving average including all past values of v_t and η_t, not a combination of their current values only.

The situation is much worse in multivariate DSGEs. For example, suppose that the DSGE contains $m \geq 2$ variables and two shocks, a demand shock v_{1t} and a supply shock v_{2t}, and that the variables are observed with measurement errors. Then the shock V_{1t}, the one that has been indentified in the VAR as the "demand shock", the identification restriction being one of those holding in the DSGE model, is dynamically contaminated, like in (4), not only by the measurement errors, but by the supply shock also.

Our positive argument is that none of these phenomena occur in a Large-Dimensional Dynamic Factor Model (DFM). We argue that (i) the variables of a DSGE model (free of the measurement error) can be estimated by a DFM, (ii) the DSGE structural shocks and impulse-response functions can be identified in the DFM using some of the DSGE restrictions. Thus the DFM shocks and impulse-response functions are a natural tool for validation of DSGEs. Moreover, the vector of the common components is singular, like the vector of the variables in a DSGE, so that neither model has to tackle the fundamentalness problem.

In Section 2 we briefly review DSGE models and their VARMA representation, singularity (more variables than shocks) as a general feature of DSGE models, measurement errors as the natural way to reconcile singularity of the model with observed variables, validation by means of VARs.

In Section 3 we study in detail the contamination effects outlined above. We show that contamination occurs for generic values of the parameters of the the DSGE model. We also show that non-fundamentalness in a block of the DSGE variables can be a source of contamination.

In Section 4 we give a short presentation of DFMs and develop our thesis that VAR models should be replaced by DFMs in DSGE validation. Section 5 concludes.

Although our topic here is quite far from Benedikt Pötscher's works on the consequences of data-driven model selection on subsequent inference, the present contribution is very close in spirit to his insistence on rigorous analysis in order to distinguish between "facts and fiction".

2 DSGE models

Let us start with the log-linearized solution of a DSGE model. The variables of interest are gathered in an m-dimensional vector

$$\mathbf{y}_t = (y_{1t}\ y_{2t}\ \cdots\ y_{mt}).$$

Well-known facts about $\mathbf{y}_t$ are the following:

(1) The vector $\mathbf{y}_t$ evolves according to a VARMA equation (see e.g. Hannan and Deistler (1988), Fernández-Villaverde et al. (2007), Morris (2016)):

$$\mathbf{C}(L)\mathbf{y}_t = \mathbf{D}(L)\mathbf{v}_t, \tag{5}$$

where $\mathbf{C}(L)$ is a stable $m \times m$ polynomial matrix in the lag operator L, $\mathbf{D}(L)$ is a $m \times p$ polynomial matrix, $\mathbf{v}_t$ is a p-dimensional orthonormal white noise, the shocks driving the system. The underlying economic theory implies restrictions on the polynomials $\mathbf{C}(L)$ and $\mathbf{D}(L)$ and therefore on the impulse-response functions $\mathbf{C}(L)^{-1}\mathbf{D}(L)$.

(2) The parameters of a DSGE model, i.e. the coefficients of the entries of $\mathbf{C}(L)$ and $\mathbf{D}(L)$, are determined by calibration, ML estimation or a mixture of the two techniques.

(3) It is also well known that DSGE models are "misspecified in the sense that they are, in general, too simple to capture the complex probabilistic nature of the data", Canova (2007), p. 160. Nevertheless, the impulse-response functions and the shocks resulting from their estimation-calibration can be compared with those obtained from a Structural VAR (SVAR), which uses the covariance-structure of the actual data and is identified by some of the DSGE restrictions. This comparison, *validation* by VARs, can be used to modify the DSGE if a mild difference emerges between the SVAR impulse-response functions and those predicted by the theory, or to reject the DSGE model if such difference is dramatic.

(4) Lastly, a general feature of DSGE models is that $p < m$, i.e. the vector $\mathbf{y}_t$ is *dynamically singular*, see e.g. Sargent (1989), Canova (2007), pp. 230-2. Assuming stationarity for $\mathbf{y}_t$, this is equivalent to the singularity of the spectral density of $\mathbf{y}_t$ at all the frequencies $\theta \in [-\pi,\ \pi]$. Of course the actual data for the variables $\mathbf{y}_t$ do not exhibit dynamic singularity. However, if it is assumed that observed data contain measurement errors, the singularity in the model is no longer inconsistent with the observed data (see e.g. Sargent (1989)).

Let us denote by $\boldsymbol{\eta}_t$ the m-dimensional vector representing the measurement errors. We assume that the measurement errors are additive, so that the observed variables, denoted by $\mathbf{x}_t$, are obtained as follows:

$$\mathbf{x}_t = \mathbf{y}_t + \boldsymbol{\eta}_t = \frac{\mathbf{D}(L)}{\mathbf{C}(L)}\mathbf{v}_t + \boldsymbol{\eta}_t = \mathbf{B}(L)u_t + \boldsymbol{\eta}_t, \tag{6}$$

that is

$$\mathbf{C}(L)\mathbf{x}_t = \mathbf{D}(L)\mathbf{v}_t + \mathbf{C}(L)\boldsymbol{\eta}_t. \tag{7}$$

Standard assumptions are:

Assumption 1. The variables v_{ht} and $\eta_{k\tau}$ are orthogonal for all $h = 1, 2, \ldots, p$, $k = 1, 2, \ldots, m$, $t \in \mathbb{Z}$, $\tau \in \mathbb{Z}$.

Assumption 2. The vector $\boldsymbol{\eta}_t$ is white noise with a non-singular variance-covariance matrix.

Note that $\mathbf{v}_t$ is orthonormal white noise, the usual assumption on the structural shocks, whereas we only assume that the second moments of the variables η_{ht} are positive.

3 Validation by means of a VAR model

We discuss validation of a DSGE model by means of a VAR by using a very simple specification for $\mathbf{C}(L)$ and $\mathbf{D}(L)$, namely that $\mathbf{C}(L) = \mathbf{I}$ and that $\mathbf{D}(L) = \mathbf{B}(L)$ is a moving average of order one:

$$\mathbf{B}(L) = \mathbf{B}_0 + \mathbf{B}_1 L\,,$$

so that the DSGE model is $\mathbf{y}_t = \mathbf{B}(L)\mathbf{v}_t$ and

$$\mathbf{x}_t = \mathbf{B}(L)\mathbf{v}_t + \boldsymbol{\eta}_t = (\mathbf{B}_0 + \mathbf{B}_1 L)\mathbf{v}_t + \boldsymbol{\eta}_t. \tag{8}$$

Under Assumptions 1 and 2, $\mathbf{x}_t$ has an MA(1) representation

$$\mathbf{x}_t = \mathbf{A}(L)\mathbf{V}_t = (\mathbf{A}_0 + \mathbf{A}_1 L)\mathbf{V}_t, \tag{9}$$

where (i) $\mathbf{V}_t$ is an orthonormal m-dimensional white noise, (ii) $\det[A(L)]$ has no roots inside the unit circle. Under (i) and (ii), the orthonormal white noise $\mathbf{V}_t$ and the matrix $\mathbf{A}(L)$ are identified up to multiplication by an orthogonal matrix. For these statements see Appendix (III), (a) and (b).

Condition (ii) implies that representation (9) fulfills the definition of *fundamentalness*, namely that $\mathbf{V}_t$ lies in the space spanned by current and past values of $\mathbf{x}_t$. We also say that $\mathbf{V}_t$ is fundamental for $\mathbf{x}_t$. Also, it will be useful to observe that for $m = 2$, under (i) and (ii), assuming that $a_{12}(0) = 0$, where $a_{12}(L)$ is the $(1,\ 2)$ entry of $\mathbf{A}(L)$, identifies $\mathbf{A}(L)$ up to a change of sign in the first column, the second column or both.

Non-singularity of $\boldsymbol{\eta}_t$ and orthogonality of $\boldsymbol{\eta}_t$ to $\mathbf{v}_\tau$ for all t and τ imply more that (ii), namely that $\det\mathbf{A}(L)$ has no roots inside *or on* the unit circle, see again Appendix (III), (a). As a consequence, $\mathbf{x}_t$ has the (infinite) VAR representation

$$\mathbf{A}(L)^{-1}(L)\mathbf{x}_t = \mathbf{V}_t.$$

Equating the right hand sides of (8) and (9), and denoting by $\mathbf{A}_{\mathrm{ad}}(L)$ the adjoint matrix of $\mathbf{A}(L)$, we have

$$\det[\mathbf{A}(L)]\mathbf{V}_t = \mathbf{A}_{\mathrm{ad}}(L)\mathbf{B}(L)\mathbf{v}_t + \mathbf{A}(L)_{\mathrm{ad}}\boldsymbol{\eta}_t. \tag{10}$$

We assume that some of the theory-based restrictions of the DSGE take the form of zeros in the matrix $\mathbf{B}_0$. Such restrictions are used to identify the VAR model in the validation procedure, so that a correspondence is established between the structural shocks and the VAR shocks. For example, if the shock v_{2t} is the supply shock and the entry $(1,\ 2)$ in $\mathbf{B}_0$ is zero, the "supply shock" in the VAR is identified by imposing that the same entry in $\mathbf{A}_0$ is zero. We might expect that the VAR supply shock is "made of" the structural supply shock and the measurement error. We show that things are worse. The VAR supply shock is also contaminated by the other structural shocks.

The contamination problem is discussed using only population entities and their moving average representations. Note that in this context the VAR equation for $\mathbf{x}_t$ is not really needed. Representations (8) and (9), and the resulting (10) are sufficient to study the relationship between $\mathbf{V}_t$, $\mathbf{v}_t$ and $\boldsymbol{\eta}_t$. Of course in empirical situations an approximation of the matrix $\mathbf{A}(L)$ will be obtained by inverting the estimated VAR.

Lastly, the examples of shock contamination given below are sufficient to make the main point of the present paper. Contamination of the impulse-response functions can be studied by the same methods, with the same results, see part (II) of the Appendix.

3.1 VAR dimension and number of structural shocks are equal

Assume that $m = p = 2$, so that the vector $(y_{1t}\ y_{2t})'$ in the DSGE is not singular. This case is not very interesting *per se* but its results are used in the sequel, see part (I) in the Appendix, which is used in Section 3.2, and Section 3.3.

To fix ideas, the shocks v_{1t} and v_{2t} are a demand and a supply shock respectively. Moreover, the supply shock v_{2t} has no contemporaneous effect of the first variable, so that we write $b_{12}(L) = f_{12}L$. Equating the right-hand sides of (8) and (9), we have in this case:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} b_{11}(L) & f_{12}L \\ b_{21}(L) & b_{22}(L) \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = \begin{pmatrix} a_{11}(L) & g_{12}L \\ a_{21}(L) & a_{22}(L) \end{pmatrix} \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix}, \tag{11}$$

where the matrix $\mathbf{A}(L)$ has been identified such that V_{2t} can be labeled as the VAR supply shock. Equation (10) takes the form:

$$\det[\mathbf{A}(L)] \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} a_{22}(L) & -g_{12}L \\ -a_{21}(L) & a_{11}(L) \end{pmatrix} \begin{pmatrix} b_{11}(L) & f_{12}L \\ b_{21}(L) & b_{22}(L) \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix},$$

where $\boldsymbol{\epsilon}_t = \mathbf{A}_{\mathrm{ad}}(L)\boldsymbol{\eta}_t$.

The conditions for non-contamination of V_{1t} by v_{2t} and of V_{2t} by v_{1t} are

$$\begin{aligned} a_{22}(L)f_{12}L - b_{22}(L)g_{12}L &= 0 \\ a_{21}(L)b_{11}(L) - a_{11}(L)b_{21}(L) &= 0, \end{aligned}$$

that is

$$\frac{a_{21}(L)}{b_{21}(L)} = \frac{a_{11}(L)}{b_{11}(L)} = \alpha(L), \qquad \frac{g_{12}}{f_{12}} = \frac{a_{22}(L)}{b_{22}(L)} = \beta(L). \tag{12}$$

Note that $\beta(L)$ is a constant. Thus:

$$\begin{pmatrix} a_{11}(L) & Lg_{12} \\ a_{21}(L) & a_{22}(L) \end{pmatrix} = \begin{pmatrix} \alpha(L)b_{11}(L) & \beta(L)f_{12}L \\ \alpha(L)b_{21}(L) & \beta(L)b_{22}(L) \end{pmatrix}. \tag{13}$$

From (11) and (13) we obtain

$$\begin{pmatrix} b_{11}(L) & f_{12}L \\ b_{21}(L) & b_{22}(L) \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = \begin{pmatrix} \alpha(L)b_{11}(L) & \beta(L)f_{12}L \\ \alpha(L)b_{21}(L) & \beta(L)b_{22}(L) \end{pmatrix} \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix}.$$

Equating the spectral densities,

$$\begin{pmatrix} |b_{11}(z)|^2 + |f_{12}|^2 & b_{11}(z)b_{21}(\bar z) + z f_{12} b_{22}(\bar z) \\ b_{11}(\bar z)b_{21}(z) + \bar z f_{12} b_{22}(z) & |b_{21}(z)|^2 + |b_{22}(z)|^2 \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$
$$= \begin{pmatrix} |\alpha(z)|^2 |b_{11}(z)|^2 + |\beta(z)|^2 |f_{12}|^2 & |\alpha(z)|^2 b_{11}(z) b_{21}(\bar z) + z|\beta(z)|^2 f_{12} b_{22}(\bar z) \\ |\alpha(z)|^2 b_{11}(\bar z) b_{21}(z) + \bar z |\beta(z)|^2 f_{12} b_{22}(z) & |\alpha(z)|^2 |b_{21}(z)|^2 + |\beta(z)|^2 b_{22}(z)|^2 \end{pmatrix},$$

where $z = e^{-i\theta}$, $\theta \in [-\pi,\ \pi]$, $\tilde\alpha(z) = |\alpha(z)|^2 - 1$, $\tilde\beta(z) = |\beta(z)|^2 - 1$, σ_h^2 is the second moment of η_{ht}. Equating entries:

$$\begin{pmatrix} |b_{11}(z)|^2 & |f_{12}|^2 \\ |b_{21}(z)|^2 & |b_{22}(z)|^2 \\ b_{11}(z)b_{21}(\bar z) & z f_{12} b_{22}(\bar z) \end{pmatrix} \begin{pmatrix} \tilde\alpha(x) \\ \tilde\beta(z) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ 0 \end{pmatrix} \tag{14}$$

(the fourth equation is just the conjugate of the third and is therefore omitted). The linear system (14), in the unknowns $\tilde\alpha(z)$ and $\tilde\beta(z)$ has a solution only if the 3×2 matrix on the-right hand side of (14), call it $\mathbf{M}(z)$, has the same rank as the matrix

$$\mathbf{N}(z) = \begin{pmatrix} |b_{11}(z)|^2 & |f_{12}|^2 & \sigma_1^2 \\ |b_{21}(z)|^2 & |b_{22}(z)|^2 & \sigma_2^2 \\ b_{11}(z)b_{21}(\bar z) & z f_{12} b_{22}(\bar z) & 0 \end{pmatrix}.$$

Now, our DSGE model has nine parameters, the seven coefficients of $\mathbf{B}(L)$ plus the two second moments of $\boldsymbol{\eta}_t$. Assume that the parameter vector belongs to an open set $\Pi \subset \mathbb{R}^9$. Adding z, which varies on the unit circle $\mathcal{C}$, the matrices $\mathbf{M}(z)$ and $\mathbf{N}(z)$ are parameterized on the set $\Pi \times \mathcal{C}$, which is the closure of an open subset of $\mathbb{R}^{10}$. It is very easy to see that the subset of $\Pi \times \mathcal{C}$ where the rank of $\mathbf{M}(z)$ equals the rank of $\mathbf{N}(z)$ is nowhere dense in $\Pi \times \mathcal{C}$. Thus *generically* the system (14) has no solution, that is, generically the supply (demand) shock of the VAR is contaminated by the demand (supply) shock of the DSGE.

3.2 VAR dimension is greater than number of structural shocks

This is the standard case, in which the vector $\mathbf{y}_t$ is singular. We have again the demand shock v_{1t} and the supply shock v_{2t} and augment model (11) with a third variable which loads both shocks with one period lag:

$$\begin{aligned} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} &= \begin{pmatrix} b_{11}(L) & f_{12}L \\ b_{21}(L) & b_{22}(L) \\ f_{31}L & f_{32}L \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}(L) & g_{12}L & a_{13}(L) \\ a_{21}(L) & a_{22}(L) & a_{23}(L) \\ g_{31}L & g_{32}L & a_{33}(L) \end{pmatrix} \begin{pmatrix} V_{1t} \\ V_{2t} \\ V_{3t} \end{pmatrix}, \end{aligned} \tag{15}$$

Again, the restrictions of the DSGE have been reproduced in the VAR model. With three zero restrictions, the latter is just identified. The DGSE has eighteen parameters: $18 - 3$ for the matrix $\mathbf{B}(L)$ plus the 3 second moments of $\boldsymbol{\eta}_t$. We assume that the parameter vector belongs to an open subset of $\mathbb{R}^{18}$.

Define $\mathbf{K}(L) = \mathbf{A}_{\mathrm{ad}}(L)$. Using equation (10), if the shock V_{1t} does not load v_{2t} and the shock V_{2t} does not load the shock v_{1t}, i.e if there is no contamination, then:

$$\begin{aligned} Lk_{11}(L)f_{12} + k_{12}(L)b_{22}(L) + Lk_{13}(L)f_{13} &= 0 \\ k_{21}(L)b_{11}(L) + k_{22}(L)b_{21}(L) + Lk_{23}(L)f_{31} &= 0. \end{aligned} \tag{16}$$

In the Appendix, part (I), we sketch a proof that generically equations (16) are not fulfilled in Π.

3.3 No measurement errors, blocks of variables, non-fundamentalness

An alternative to measurement errors to reconcile the singularity of the DSGE with observed data consists in selecting blocks of variables so that the number of shocks and the number of variables are equal, see Canova (2007), p. 232-3.

Assume that $\boldsymbol{\eta}_t = 0$, so that $\mathbf{x}_t = \mathbf{y}_t$, and that from a DSGE with $p = 2$ we have selected the variables y_{1t} and y_{2t}. Assuming that they are modeled like in (11),

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} b_{11}(L) & f_{12}L \\ b_{21}(L) & b_{22}(L) \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} = \begin{pmatrix} a_{11}(L) & g_{12}L \\ a_{21}(L) & a_{22}(L) \end{pmatrix} \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix}.$$

Because $\mathbf{V}_t$ is fundamental by definition, if $\mathbf{v}_t$ is fundamental the matrices $\mathbf{B}(L)$ and $\mathbf{A}(L)$ are equal up to a change of sign in the first column, the second column or both (see the observations following the definition of fundamentalness in Section 2). Thus of course equation (12) is fulfilled and no contamination occurs.

Suppose that $\mathbf{v}_t$ is non-fundamental, i.e. $\det[\mathbf{B}(L)]$ has a root of modulus less than unity, call it z^*, and that equation (12) is fulfilled. Because (i) $\det[\mathbf{A}(L)]$ has no roots inside the unit circle, (ii) $\det[\mathbf{A}(L)] = \alpha(L)\beta(L)\det[\mathbf{B}(L)]$, (iii) $\beta(L)$ is a constant, then $\alpha(L)$ has a pole at z^*. On the other hand the entries of $\mathbf{A}(L)$ have no poles of modulus less than unity so that both $b_{11}(L)$ and $b_{21}(L)$ have a root at z^*. In conclusion, non-fundamentaleness is allowed for $\mathbf{v}_t$ but only in a special form, namely the entries of the first column of $\mathbf{B}(L)$ must share a root of modulus less than unity. From (12) we obtain

$$\alpha(L) = \gamma \frac{L - z^*}{1 - z^* L}.$$

We have, setting $\delta = \beta(L)$:

$$\begin{aligned} \mathbf{y}_t = \mathbf{A}(L)\mathbf{V}_t &= \begin{pmatrix} \gamma \frac{1 - z^* L}{L - z^*} b_{11}(L) & \delta f_{12}L \\ \gamma \frac{1 - z^* L}{L - z^*} b_{21}(L) & \delta b_{22}(L) \end{pmatrix} \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix} \\ &= \mathbf{B}(L) \begin{pmatrix} \gamma \frac{1 - z^* L}{L - z^*} & 0 \\ 0 & \delta \end{pmatrix} \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix} = \mathbf{B}(L)\mathbf{v}_t. \end{aligned}$$

It easily seen that

$$\begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \gamma^{-1} \frac{L - z^*}{1 - z^* L} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}.$$

Thus, although no contamination occur in this case, the shock V_t is an infinite moving average of v_{1t}. On the other hand, if $\mathbf{B}(L)$ is non-fundamental and (12) is not fulfilled then contamination occurs.

The non-fundamentalness issue for DSGE linearized solutions can be easily described in general. Let us go back to model (5):

$$\mathbf{C}(L)\mathbf{y}_t = \mathbf{D}(L)\mathbf{v}_t.$$

As recalled in Section 2, the vector $\mathbf{y}_t$ is dynamically singular, i.e. m, the dimension of $\mathbf{y}_t$, is greater than p, the dimension of $\mathbf{v}_t$. Singularity of $\mathbf{y}_t$ implies that *generically* $\mathbf{v}_t$ is fundamental for $\mathbf{y}_t$. This important result has been proved in Anderson and Deistler (2008a) and Anderson and Deistler (2008b). An elementary illustration is the following:

$$\begin{aligned} y_{1t} &= b_{1,0}v_t + b_{1,1}v_{t-1} \\ y_{2t} &= b_{2,0}u_t + b_{2,1}v_{t-1}. \end{aligned}$$

Here $m = 2$ and $p = 1$. If $b_{1,0}b_{2,1} - b_{1,1}b_{2,0} \neq 0$, we obtain

$$v_t = \frac{1}{b_{1,0}b_{2,1} - b_{1,1}b_{2,0}}(b_{2,1}y_{1t} - b_{1,1}y_{2t}),$$

so that v_t lies in the space spanned by current and past values of $\mathbf{y}_t$. Thus, apart from he lower-dimensional subset of $\mathbb{R}^4$ where $b_{1,0}b_{2,1} - b_{1,1}b_{2,0} = 0$, the shock v_t is fundamental for the vector $\mathbf{y}_t$.

However, fundamentalness of $\mathbf{v}_t$ for $\mathbf{y}_t$ does not imply that $\mathbf{v}_t$ is fundamental for a p-dimensional block. In the example above, if $y_{1t} = v_t - 4v_{t-1}$ and $y_{2t} = v_t - 0.5v_{t-1}$, v_t is fundamental for $\mathbf{y}_t$ and for the block containing only y_{2t}, but non-fundamental for the block containing only y_{1t}.

In conclusion, fundamentalness is not an issue for the whole DSGE model. However, assuming no measurement errors, if a block of p variables is selected to be used for validation by means of a VAR, then the block should be carefully analyzed to ascertain if fundamentalness of the shocks for the block is warranted by the theory.

3.4 No measurement errors, more structural shocks than VAR dimension

As in the previous section, there are no measurement errors: $\mathbf{x}_t = \mathbf{y}_t$. Suppose that the VAR is misspecified in that its dimension is less than the number of structural shocks. For example, assume that there are two different demand shocks in the DSGE, v_{1t} and v_{2t}, and one supply shock v_{3t}, but the block selected for VAR estimation includes only the two variables y_{1t} and y_{2t}. Thus

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} b_{11}(L) & b_{12}(L) & f_{13}L \\ b_{21}(L) & b_{22}(L) & b_{23}(L) \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{pmatrix} = \begin{pmatrix} a_{11}(L) & g_{12}L \\ a_{21}(L) & a_{22}(L) \end{pmatrix} \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix},$$

so that:

$$\det[A(L)] \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} a_{22}(L) & -g_{12}L \\ -a_{21}(L) & a_{11}(L) \end{pmatrix} \begin{pmatrix} b_{11}(L) & b_{12}(L) & f_{13}L \\ b_{21}(L) & b_{22}(L) & b_{23}(L) \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{pmatrix}.$$

The conditions for non-contamination of the supply shock by the demand shocks are:

$$\begin{aligned} a_{21}(L)b_{11}(L) - a_{11}(L)b_{21}(L) &= 0 \\ a_{21}(L)b_{12}(L) - a_{11}(L)b_{22}(L) &= 0, \end{aligned}$$

that is

$$b_{21}(L) = \gamma(L)b_{11}(L), \quad b_{22}(L) = \gamma(L)b_{12}(L). \tag{17}$$

Now observe that $b_{11}(L)v_{1t}+b_{12}(L)v_{2t}$ can be represented as $\tilde{b}(L)\tilde{v}_t$ where $\tilde{v}_t$ is a unit-variance white noise, so that, if (17) holds,

$$\begin{aligned} b_{11}(L)v_{1t} + b_{12}(L)v_{2t} &= \tilde{b}(L)\tilde{v}_t \\ b_{21}(L)v_{1t} + b_{22}(L)v_{2t} &= \gamma(L)\tilde{b}(L)\tilde{v}_t \end{aligned}$$

and the DSGE model has the representation

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} \begin{pmatrix} \tilde{b}(L) & b_{13}(L) \\ \gamma(L)\tilde{b}(L) & b_{23}(L) \end{pmatrix} \begin{pmatrix} \tilde{v}_t \\ v_{3t} \end{pmatrix},$$

with only one demand shock and the original supply shock. Thus, if there is a genuine couple of demand shocks, i.e. condition (17) does not hold, the supply shock V_{2t} gets contaminated by the demand shocks v_{1t} and v_{2t}.

Lastly, even when (17) is satisfied, the aggregate demand shock $\tilde{v}_t$, defined by $\tilde{b}(L)\tilde{v}_t = b_{11}(L)v_{1t} + b_{12}(L)v_{2t}$, though depending only on the demand shocks, is a linear combination of current and past values of them, not only of their current values (see the same observation for the simple example in the Introduction).

4 Large-Dimensional Dynamic Factor Models

4.1 General definitions

An argument to dismiss the results of the previous section might be that the coefficients of the matrix $\mathbf{A}(L)$ and the shocks $\mathbf{V}_t$ are continuous functions of the parameters of the DSGE, including the second moments of $\boldsymbol{\eta}_t$. As a consequence, if the measurement errors are small, then after all the representation $\mathbf{x}_t = \mathbf{A}(L)\mathbf{V}_t$ is close to $\mathbf{y}_t = \mathbf{B}(L)\mathbf{v}_t$ and therefore validation of the DSGE by means of a VAR is acceptable. This is fairly reasonable.

However, we claim that a DFM is an alternative tool that can be used to clean the variables $\mathbf{x}_t$ from the error $\boldsymbol{\eta}_t$, so obtaining an estimate of $\mathbf{y}_t$ and that such an estimation is fairly simple.

To fix ideas let us consider a dataset of macroeconomic time series, call it $\mathbf{X}_t$, which includes those that are typical of DSGEs, aggregate income, prices, industrial production, rate of interest, etc. plus sectoral and regional economic indicators. We assume that the dataset contains a number of variables, call it n, which is large as compared to T, the number of observations for each time series, so that estimating a VAR is unfeasible. This feature, an n comparable in size to T, is embodied in definitions and the asymptotic analysis, in which both T and n tend to infinity (thus Large-Dimensional DFMs). The general form of the DFM is the following:

$$\begin{aligned} x_{it} &= \chi_{it} + \xi_{it} \\ \chi_{it} &= \mu_{i1}(L)u_{1t} + \mu_{i2}(L)u_{2t} + \cdots + \mu_{iq}(L)u_{qt}, \end{aligned} \tag{18}$$

for $t \in \mathbb{Z}$ and $n \in \mathbb{N}$, where:
(i) The vector $\mathbf{u}_t = (u_{1t}\ u_{2t}\ \cdots\ u_{qt})'$ is an orthonormal white noise, the vector of the *common shocks*, also called the *dynamic factors.*
(ii) The polynomials $\mu_{ij}(L)$ are rational functions of L with no poles inside the unit circle.

(iii) The variables ξ_{it}, called the *idiosyncratic components*, are zero-mean stationary. Moreover, they are orthogonal to the common shocks at all leads and lags, i.e. $\xi_{it} \perp u_{j\tau}$ for all t, $\tau \in \mathbb{Z}$, $i \in \mathbb{N}$. As a consequence they are orthogonal to the variables χ_{it}, which are called the *common components.*
(iv) Idiosyncratic components for different i's are *weakly correlated.* This is an asymptotic definition whose details are not needed here. It requires, for example, that the mean of the ξ's tends to zero as n tends to infinity:

$$\lim_{n\to\infty} \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\xi_{it}\right]^2 = 0.$$

This is obviously true if the ξ's are mutually orthogonal with an upper bound for the variance, but is also true if some "local" non-zero covariance among the ξ's is allowed.
(v) The common shocks are *pervasive,* i.e. they affect all the variables x_{it}, with possibly a finite number of exceptions.

For a statement of the assumptions, representation and estimation results, see Forni and Reichlin (1998), Forni et al. (2000), Forni and Lippi (2001), Stock and Watson (2002b), Stock and Watson (2002a). In these papers, and in the many others in this literature, it is proved that the shocks $\mathbf{u}_t$ and the common components χ_{it} can be estimated by taking some averages over the x's and letting n and T tend to infinity. The weak correlation property of the ξ's, see (iv) above, ensures that in such averages only the common components survive as n tends to infinity.

The idiosyncratic components are interpreted as a cause of variation of the x's that are specific to one or just a few variables, like *regional* or *sectoral shocks*, plus *measurement errors.* In particular, for the big aggregates like income, consumption, investment, in which all local or sectoral shocks have been averaged out, the variable ξ_{it} can be interpreted as only containing measurement error.

On the other hand, the common shocks $\mathbf{u}_t$, as they are pervasive, see (v) above, are interpreted as *macroeconomic* causes of variation.

A common additional assumption in the literature on DFMs is that the space spanned by the common components χ_{it}, for a given t, call it $\mathcal{S}_t$, has finite dimension r. As a consequence, $\mathcal{S}_t$ has a finite stationary basis $\mathbf{F}_t = (F_{1t}\ F_{2t}\ \ \cdots\ F_{rt})'$ such that

$$x_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} + \xi_{it}. \tag{19}$$

The variables F_{jt} are called the *static factors* and (19) the *static representation* of the DFM. For example, if $q = 1$ and

$$x_{it} = \mu_{i,0}u_t + \mu_{i,1}u_{t-1} + \xi_{it},$$

we set $F_{1t} = u_t$, $F_{2t} = u_{t-1}$, and the static representation is

$$x_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \xi_{it},$$

with $\lambda_{i1} = \mu_{i,0}$, $\lambda_{i2} = \mu_{i,1}$. We see that the static representation is obtained by replacing the dynamics with "artificial" static factors, so that the dynamics of the common components has been moved into the static factors:

$$\begin{pmatrix} F_{1t} \\ F_{2t} \end{pmatrix} = \begin{pmatrix} 1 \\ L \end{pmatrix} u_t, \quad \text{or} \quad \begin{pmatrix} 1 & 0 \\ -L & 1 \end{pmatrix}\begin{pmatrix} F_{1t} \\ F_{2t} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} u_t. \tag{20}$$

The example above is sufficient to motivate the assumption that

$r > q$, i.e. the number of static factors is greater than the number of dynamic factors,

and therefore that the vector $\mathbf{F}_t$ is singular. The moving average representation on the left in (20) has the generalization

$$\mathbf{F}_t = \mathbf{G}(L)\mathbf{u}_t, \tag{21}$$

where $\mathbf{G}(L)$ is an $r \times q$ matrix of rational functions of L, thus a non-square matrix. Anderson and Deistler, in the papers cited in Section 3.3, show that, for generic values of the coefficients of the rational functions in $\mathbf{G}(L)$, the singular vector $\mathbf{F}_t$ has an autoregressive representation

$$\mathbf{H}(L)\mathbf{F}_t = \mathbf{G}(0)\mathbf{u}_t, \tag{22}$$

where $\mathbf{H}(L)$ is an $r \times r$ stable polynomial matrix of *finite* degree. This implies of course the result mentioned in Section 3.3, that representation (21) is generically fundamental.

In conclusion, under the assumption that $\mathcal{S}_t$ has finite dimension r, the DFM can be represented in the form:

$$\begin{aligned} x_{it} &= \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} + \xi_{it} \\ \mathbf{H}(L)\mathbf{F}_t &= \mathbf{R}\mathbf{u}_t, \end{aligned} \tag{23}$$

where $\mathbf{H}(L)$ is an $r \times r$ stable polynomial matrix of finite degree $\mathbf{R}$ is an $r \times q$ matrix.

Let us insist that, under the assumptions of singularity for $\mathbf{F}_t$ and rationality for the functions $\mu_{ij}(L)$, Anderson and Deistler results imply fundamentalness of $\mathbf{v}_t$ and the finite degree of $\mathbf{H}(L)$, so that representation (23) is quite general.

Estimation of model (23) requires three steps.

(I) Firstly the dimensions q and r must be determined. From the vast literature on the topic we only mention here Bai and Ng (2002), the first paper to provide a criterion for r, consistent for n and T tending to infinity, and Hallin and Liška (2007) for q.

(II) Once r and q have been specified, the factors $\mathbf{F}_t$ and the loadings λ_{ij} can be estimated consistently by taking the first r principal components of the observations x_{it}, $i = 1, 2, \dots, n$, $t = 1, 2, \dots, T$.

(III) The estimated factors are used to estimate the non-standard VAR in (23), and therefore $\mathbf{H}(L)$, $\mathbf{R}$ and the dynamic factors $\mathbf{u}_t$. Estimates of $\mu_{ij}(L)$ are easily obtained. Defining $\mathbf{G}(L) = \mathbf{H}(L)^{-1}\mathbf{R}$,

$$\chi_{it} = (\lambda_{i1}\ \lambda_{i2}\ \cdots\ \lambda_{ir})\mathbf{F}_t = (\lambda_{i1}\ \lambda_{i2}\ \cdots\ \lambda_{ir})\mathbf{G}(L)\mathbf{u}_t = (\mu_{i1}(L)\ \mu_{i2}(L)\ \cdots\ \mu_{iq}(L))\mathbf{u}_t,$$

so that, under the assumption of finite dimension for $\mathcal{S}_t$, we have obtained an estimate of model (18).

Lastly, let us point out an important difference beween Large-Dimensional DFMs and standard Factor Models in which the number of variables is given, the model is estimated by maximum likelihood and the asymptotic analysis is conducted for T tending to infinity. The latter require for identification that the idiosyncratic components are mutually orthogonal whereas in the Large-Dimensional DFM we only need weak correlation, see (iv) above. But measurement errors in variables belonging to the same group, real variables like income and consumption for example, might well be correlated in macroeconomic datasets. Thus the assumption of weak correlations seems more realistic. Estimation by maximum likelihood of a model of the form (7) has been suggested in Sargent (1989). Giannone et al. (2006) apply this idea to estimate a simple DSGE under the assumption of orthogonal idiosyncratic components.

4.2 Comparing DSGE and DFM

The static representation (19) and the static factors $\mathbf{F}_t$ are useful for the estimation of the DFM. However, if we are interested in structural analysis we must revert to the original representation (18) and the dynamic factors $\mathbf{u}_t$.

Our claim is, as stated above, that if x_{it} is a macroeconomic variable like aggregate income, investment, consumption, the idiosyncratic component ξ_{it} can be interpreted as the measurement error, so that the common component χ_{it} is the cleaned version of x_{it}, the variable that should be considered in structural analysis.

On the other hand, as argued in Stock and Watson (2005) and Forni et al. (2009), identification techniques applied in SVAR or DSGE analysis can be easily used for identifying DFMs.

Let us concentrate on the DSGE model. We assume that the common components of the first m variables of the DFM are the variables of the DSGE: $\boldsymbol{\chi}_t = \mathbf{y}_t$, where $\boldsymbol{\chi}_t = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{mt})'$. Moreover, to fix ideas, let us assume that $p = 2$, a demand and a supply shock, and that the number of shocks in the DFM has been correctly determined, that is $q = 2$. Then we have two rational moving average representations for $\mathbf{y}_t$:

$$\mathbf{y}_t = \mathbf{B}(L)\mathbf{v}_t = \boldsymbol{\mu}(L)\mathbf{u}_t,$$

where $\boldsymbol{\mu}(L)$ has $\mu_{ij}(L)$ in the $(i,\ j)$ entry. Both representations are singular, so that generically both are fundamental. As a consequence, the white noise vectors $\mathbf{u}_t$ and $\mathbf{v}_t$ differ for an orthogonal matrix:

$$\mathbf{v}_t = \mathbf{S}\mathbf{u}_t,$$

where $\mathbf{S}$ is a $q \times q$ orthogonal matrix (2×2 in our case), see Appendix (III), (b). If the DSGE assumes that the shock v_{2t} has no contemporaneous impact on the variable $y_{1t} = \chi_{1t}$, the matrix $\mathbf{S}$ is identified by the condition

$$\mu_{11}(0)s_{21} + \mu_{12}(0)s_{22} = 0,$$

see again Section 2. We believe that this elementary example is sufficient to make the point that no modification is required to apply an identifying restriction from DSGE or SVAR analysis within a DFM.

Thus DFMs can be used to validate both SVAR and DSGE models:

(i) The criteria for determining the number of shocks in the DFM can be used as a data-driven evaluation for the dynamic dimension (number of shocks) of SVAR and DSGE models.

(ii) Regarding SVARs, one can be interested in the shape of the impulse-response functions estimated using the error-free macroeconomic variables χ_{it}. Important papers adopting this approach are Bernanke and Boivin (2003), Bernanke et al. (2005), Boivin et al. (2009). Forni and Gambetti (2010) use a DFM to study the effect of monetary policy shocks on real exchange rate and stock prices. They find that in the DFM neither the delayed overshooting puzzle nor the price puzzle occur.

(iii) The use of DFMs for validation of DSGE models is obvious. The variables $\mathbf{y}_t$ in equation (5) are estimated in the DFM. The shocks of the DFM can be identified using some of the theoretical restrictions of the DSGE. The corresponding impulse-response functions can be compared with those of the DSGE. None of the contamination problems outlined in Section 3 arises.

5 Conclusions

The example in the Introduction, equations (2), (3), and (4), warning against the common misconception that measurement errors have no dramatic effect on the processes governing observable variables, goes back as far as Granger and Morris (1976). The dynamic contamination effects of measurement errors studied in Section 3 are special cases of the dynamic contamination effects of aggregation, as analyzed in Forni and Lippi (1997) (incidentally, the interest of the present writer for DFMs was spurred by the negative results obtained in that book).

DFMs have been applied extensively for forecasting. However, as we have seen, under reasonable assumptions they can be used for validation of DSGEs. For this purpose, provided that the number of dynamic factors is correctly determined, their advantages with respect to VARs are that neither non-fundamentalness nor contamination of shocks can occur. Although little explored as yet, the application of DFMs to macroeconomic analysis has a sound theoretical basis and is therefore very promising.

References

Anderson, B. D. and M. Deistler (2008a). Generalized linear dynamic factor models–A structure theory. *IEE Conference on Decision and Control*.

Anderson, B. D. and M. Deistler (2008b). Properties of zero-free transfer function matrices. *SICE Journal of Control, Measurement and System Integration 1*, 284–292.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bernanke, B. S. and J. Boivin (2003, April). Monetary policy in a data-rich environment. *Journal of Monetary Economics 50*(3), 525–546.

Bernanke, B. S., J. Boivin, and P. S. Eliasz (2005). Measuring the effects of monetary policy: A Factor-Augmented Vector Autoregressive (FAVAR) approach. *The Quarterly Journal of Economics 120*, 387–422.

Boivin, J., M. P. Giannoni, and I. Mihov (2009, March). Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data. *American Economic Review 99*(1), 350–384.

Canova, F. (2007). *Methods for applied Macroeconomics.* Princeton, New Jersey: Princeton University Press.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007). ABCs (and Ds) of Understanding VARs. *American Economic Review 97*(3), 1021–1026.

Forni, M. and L. Gambetti (2010). The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics 57*, 203–216.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). Opening the black box: Structural factor models versus structural VARs. *Econometric Theory 25*, 1319–1347.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The Generalized Dynamic Factor Model: Identification and estimation. *The Review of Economics and Statistics 82*, 540–554.

Forni, M. and M. Lippi (1997). *Aggregation and the microfoundations of dynamic macroeconomics.* Oxford: Oxford University Press.

Forni, M. and M. Lippi (2001). The Generalized Dynamic Factor Model: Representation theory. *Econometric Theory 17*, 1113–1141.

Forni, M. and L. Reichlin (1998). Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics. *Review of Economic Studies 65*(3), 453–473.

Giannone, D., L. Reichlin, and L. Sala (2006). VARs, common factors and the empirical validation of equilibrium business cycle models. *Journal of Econometrics 132*(1), 257–279.

Granger, C. W. J. and M. J. Morris (1976). Time series modelling and interpretation. *Journal of the Royal Statistical Society. Series A. 139*(2), 246–257.

Hallin, M. and R. Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association 102*, 603–617.

Hannan, E. J. and M. Deistler (1988). *The statistical theory of linear systems.* New York: Wiley.

Lütkepohl, H. (1984). Linear transformations of vector ARMA processes. *Journal of Econometrics 26*(3), 283–293.

Morris, S. D. (2016). VARMA representation of DSGE models. *Economics Letters 138*, 30–33.

Rozanov, Y. A. (1967). *Stationary Random Processes.* San Francisco: Holden Day.

Sargent, T. J. (1989). Two Models of Measurements and the Investment Accelerator. *Journal of Political Economy 97*(2), 251–287.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*, 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*, 147–162.

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. Working Paper 11467, NBER.

Appendix

(I) Consider the first of equations (16):

$$\begin{aligned} Lk_{11}(L)f_{12} + k_{12}(L)b_{22}(L) + Lk_{13}(L)f_{13} &= L[a_{22}(L)a_{33}(L) - La_{23}(L)g_{32}]f_{12} \\ &- [a_{11}(L)a_{33}(L) - La_{13}(L)g_{31}]b_{22}(L) + L[La_{21}(L)g_{32} - La_{22}(L)g_{31}]f_{13} \\ &= \zeta_0 + \zeta_1 L + \zeta_2 L^2 + \zeta_3 L^3 = 0. \end{aligned}$$

This condition is equivalent to

$$\zeta_s = 0, \quad \text{for } s = 0,1,2,3. \tag{24}$$

It is easily seen that ζ_s is a polynomial function of (i) the coefficients of the entries of $\mathbf{B}(L)$, (ii) the coefficients of the entries of $\mathbf{A}(L)$. The method employed in Forni and Lippi (1997), Chapters 6 and 10, can be easily adapted to show that the coefficients of the entries of $\mathbf{A}(L)$ are analytic functions of the coefficients of the entries of $\mathbf{B}(L)$ and the three second moments of $\boldsymbol{\eta}_t$ (see Forni and Lippi (1997), Section 10.1). Therefore ζ_s is an analytic function of $\mathbf{p} \in \Pi$. As a consequence, assuming that Π is open and connected, $\zeta_s = 0$, for $s = 0,1,2,3$, holds either on the whole Π or on a nowhere dense subset (see Forni and Lippi (1997), Section 10.2). Thus, it is sufficient to find a point in $\mathbf{p}^* \in \Pi$ such that $\zeta_s \neq 0$, for some s, to obtain that generically (24) does not hold in Π.

Finding a point $\mathbf{p}^*$ is fairly easy. Let $\tilde{\Pi}$ be the subset of Π which contains all parameter vectors such that the third row of $\mathbf{B}(L)$ vanishes and assume that $\tilde{\Pi}$ is not empty. For $\mathbf{p} \in \tilde{\Pi}$, we firstly obtain the fundamental representation for $(x_{1t}\ x_{2t})'$:

$$\begin{pmatrix} b_{11}(L) & L\,b_{12}(L) \\ b_{21}(L) & b_{22}(L) \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = \begin{pmatrix} a_{11}(L) & L\,a_{12}(L) \\ a_{21}(L) & a_{22}(L) \end{pmatrix} \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix}, \tag{25}$$

then the fundamental representation for the whole vector:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} b_{11}(L) & L\,b_{12}(L) \\ b_{21}(L) & b_{22}(L) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{pmatrix} = \begin{pmatrix} a_{11}(L) & L\,a_{12}(L) & 0 \\ a_{21}(L) & a_{22}(L) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} V_{1t} \\ V_{2t} \\ V_{3t} \end{pmatrix},$$

where $V_{3t} = \eta_{3t}/\sigma_3^2$. As $\tilde{\Pi}$ is non-empty, the parameters of the model on the left-hand side of (25) lie in an open connected non-empty subset of $\mathbb{R}^9$. Thus the results of Section 3.1 apply and generically in $\tilde{\Pi}$ contamination occurs, so that $\zeta_s \neq 0$ for some s.

(II) We give here a very simple example to illustrate the contamination occurring in the impulse-response functions. Consider the model

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} b_{11} & f_{12}L \\ 0 & b_{22} \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}, \tag{26}$$

which is a special case of model (11). Both x_{1t} and x_{2t} are white noise and their covariance is zero. However, as the covariance between x_{1t} and $x_{2,t-1}$ is $f_{12}b_{22}$, the vector $\mathbf{x}_t$ is not a white noise in general. As a candidate for the representation $\mathbf{x}_t = \mathbf{A}(L)\mathbf{V}_t$, consider

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} & g_{12}L \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix}. \tag{27}$$

As $\mathbf{V}_t$ is assumed to be orthonormal, equating covariances between (26) and (27) we obtain the following three equations for the entries of $\mathbf{A}(L)$:

$$\begin{aligned} a_{11}^2 + g_{12}^2 &= b_{11}^2 + f_{12}^2 + \sigma_1^2 \\ a_{22}^2 &= b_{22}^2 + \sigma_2^2 \\ g_{12}a_{22} &= f_{12}b_{22}. \end{aligned}$$

The system is easily solved:

$$\begin{aligned} a_{11}^2 &= b_{11}^2 + f_{12}^2 + \sigma_1^2 - f_{12}^2 \frac{b_{22}^2}{b_{22}^2 + \sigma_2^2} \\ a_{22}^2 &= b_{22}^2 + \sigma_2^2 \\ g_{12}^2 &= f_{12}^2 \frac{b_{22}^2}{b_{22}^2 + \sigma_2^2} \\ \mathrm{sign}(g_{12}a_{22}) &= \mathrm{sign}(f_{12}b_{22}). \end{aligned} \tag{28}$$

Thus, representation (27), with its coefficients determined in (28), produces the same covariance matrices as (26). Moreover, $\det[\mathbf{A}(L)]$ has no roots. Therefore (27) is the unique fundamental representation for $\mathbf{x}_t$ with an orthonormal white noise and fulfilling the condition that the polynomial in entry $(1,\ 2)$ vanishes for $L = 0$ (up to a change of sign for V_{1t}, for V_{2t} or for both, this corresponding to the multiple solutions of (28)).

We see that a_{22} depends on b_{22} and the size of the measurement error η_{2t}. However, unless $\sigma_2^2 = 0$, both a_{11} and g_{12} are contaminated by b_{22}. Using the technique briefly illustrated in part (I), example (26) could be used to show that contamination occurs generically in model (11).

(III) (a) The spectral density of $\mathbf{x}_t$, as defined in (8), is

$$\mathbf{\Sigma}^x(\theta) = (\mathbf{B}_0 + \mathbf{B}_1 e^{-i\theta})(\mathbf{B}_0 + \mathbf{B}_1 e^{i\theta})' + \mathbf{\Sigma}^\eta,$$

where $\mathbf{\Sigma}^\eta$ is the covariance matrix of $\boldsymbol{\eta}_t$. Assumption 2 implies that $\mathbf{\Sigma}^x(z)$ is non singular for all $z \in \mathbb{C}$. Moreover, the covariance function of $\mathbf{x}_t$, that is $\mathrm{E}(\mathbf{x}_t\mathbf{x}_{t-k}')$, vanishes for $|k| > 1$. Therefore $\mathbf{x}_t$ has a Wold representation $\mathbf{x}_t = \mathbf{A}(L)\mathbf{V}_t$, where (i) $\mathbf{V}_t$ is orthonormal white noise, (ii) $\mathbf{A}(L)$ is an MA(1), (iii) $\mathbf{A}(z)$ has no roots inside or on the unit circle (see Rozanov (1967), pp. 43-50; see also Lütkepohl (1984)).

(b) Let $\mathbf{w}_t$ be an r-dimensional stochastic vector and suppose that

$$\mathbf{w}_t = \boldsymbol{\alpha}(L)\mathbf{v}_t = \boldsymbol{\beta}(L)\mathbf{u}_t,$$

where $\mathbf{v}_t$ and $\mathbf{u}_t$ are q-dimensional and orthonormal white noises and $q \leq r$. Suppose that both $\mathbf{v}_t$ and $\mathbf{u}_t$ are fundamental for $\mathbf{w}_t$. Then $\mathbf{v}_t = \mathbf{S}\mathbf{u}_t$, where $\mathbf{S}$ is a $q \times q$ orthogonal matrix (see Rozanov (1967), pp. 56-7; see also Forni et al. (2009), Section 3.2).