

# Identification and Estimation in Non-Fundamental Structural VARMA Models\*

Christian Gouriéroux  
University of Toronto  
Toulouse School of Economics  
CREST

Alain Monfort  
CREST  
Banque de France

Jean-Paul Renne  
University of Lausanne

(Revised version, April 2019)

## Abstract

The basic assumption of a structural vector autoregressive moving-average (SVARMA) model is that it is driven by a white noise whose components are uncorrelated or independent and can be interpreted as economic shocks, called “structural” shocks. When the errors are Gaussian, independence is equivalent to non-correlation and these models face two identification issues. The first identification problem is “static” and is due to the fact that there is an infinite number of linear transformations of a given random vector making its components uncorrelated. The second identification problem is “dynamic” and is a consequence of the fact that, even if a SVARMA admits a non invertible moving average (MA) matrix polynomial, it may feature the same second-order dynamic properties as a VARMA process in which the MA matrix polynomials are invertible (the fundamental representation). The aim of this paper is to explain that these difficulties are mainly due to the Gaussian assumption, and that both identification challenges are solved in a non-Gaussian framework if the structural shocks are assumed to be instantaneously and serially independent. We develop new parametric and semi-parametric estimation methods that accommodate non-fundamentality in the moving average dynamics. The functioning and performances of these methods are illustrated by applications conducted on both simulated and real data.

**JEL codes:** C01, C15, C32, E37.

**Keywords:** Structural VARMA, Fundamental Representation, Identification, Structural Shocks, Impulse Response Function.

---

\*The first author gratefully acknowledges financial support of the chair ACPR/Risk Foundation: Regulation and Systemic Risks. We thank M. Barrigozi, M. Deistler, M. Dungey, A. Guay, A. Hyvärinen, J. Jasiak, M. Lippi, H. Lütkepohl, N. Meddahi, A. Melino, H. Uhlig, A. de Paula and anonymous referees for helpful comments. We also thank participant to the 2017 European Meeting of the Econometric Society, to the 2017 annual congress of the Swiss Society of Economics and Statistics and participants to a seminar at UQÀM. The second author gratefully acknowledges financial support of the chair Fondation du risque “New Challenges for New Data”. The views expressed in this paper are ours and do not necessarily reflect the views of the Banque de France.

# 1 Introduction

The basic assumption of a structural VARMA model (SVARMA) is that it is driven by a white noise whose components are uncorrelated or independent and are interpreted as economic shocks, called “structural” shocks. When the errors are Gaussian, independence is equivalent to non-correlation and these models have to face two kinds of identification problems.

First the components of the white noise appearing in the reduced-form VARMA are instantaneously correlated and the shock vector must be derived from this white noise by a linear transformation eliminating these instantaneous correlations. The snag is that this can be done in an infinite number of ways and, in the Gaussian case, all the resulting standardized shock vectors have the same (standard Gaussian) distribution. There is a huge literature trying to solve this “static” identification issue by adding restrictions on the short-run impact of a shock (see e.g. [Bernanke, 1986](#); [Sims, 1980, 1986, 1989](#); [Rubio-Ramirez, Waggoner, and Zha, 2010](#)), or on its long-run impact (see e.g. [Blanchard and Quah, 1989](#); [Faust and Leeper, 1997](#); [Erceg and Gust, 2005](#); [Christiano, Eichenbaum, and Vigfusson, 2006](#)), as well as on the sign of some impulse response functions (see e.g. [Uhlig, 2005](#); [Chari, Kehoe, and McGrattan, 2008](#); [Mountford and Uhlig, 2009](#)).

A second identification issue comes from the fact that a stationary SVARMA process may feature a non-invertible moving average (MA) matrix lag polynomial, in the sense that it cannot be inverted into a matrix lag series. This is the case when the determinant of the matrix lag polynomial has some roots inside the unit circle. The latter situation, called non-fundamentalness, may occur for many reasons, in particular when the SVARMA is deduced from business cycle models (see e.g. [Kydland and Prescott, 1982](#); [Francis and Ramey, 2005](#); [Gali and Rabanal, 2005](#)), or from log-linear approximations of Dynamic Stochastic General Equilibrium (DSGE) models involving rational expectations or news shocks (see e.g. [Hansen and Sargent, 1991](#); [Smets and Wouters, 2003](#); [Christiano, Eichenbaum, and Vigfusson, 2007](#); [Fève, Matheron, and Sahuc, 2009](#); [Sims, 2012](#); [Leeper, Walker, and Yang, 2013](#); [Blanchard, L’Huillier, and Lorenzoni, 2013](#)). A non-fundamental SVARMA process has exactly the same second-order dynamic properties as another VARMA process with an invertible MA part (the fundamental representation) and, in the Gaussian case, both processes are observationally equivalent. This creates a dynamic identification problem, which is exacerbated by the fact that the standard Box-Jenkins approach –the Gaussian Pseudo Maximum Likelihood method based on a VAR approximation of the VARMA ([Box and Jenkins, 1970](#))– automatically provides a consistent estimation of the fundamental representation. Even though they can feature the exact same second-order dynamic properties, a fundamental and a non-fundamental SVARMA entail different Impulse Response Functions (IRFs). Using the Box-Jenkins approach may therefore lead to misspecified IRFs (see [Lippi and Reichlin, 1993, 1994](#)). As in the static identification literature, some papers propose dynamic identification by imposing

recursivity conditions (Mertens and Ravn, 2010), or by using the fact that some roots of the MA part are known (see e.g. Forni, Gambetti, Lippi, and Sala, 2017b, where zero is a root).

The aim of this paper is to explain that these difficulties are due to the Gaussian assumption underlying the Box-Jenkins type approaches, and that these identification problems disappear in a non-Gaussian framework when the structural shocks are assumed to be instantaneously and serially independent. We also introduce novel parametric and semi-parametric estimation approaches that accommodate non-fundamentalness in the multivariate moving-average dynamics.

In Section 2, we consider a vector autoregressive moving average process, with roots of the moving average polynomial that are not necessarily outside the unit circle. We stress that the economic shocks are not necessarily interpretable in terms of causal innovations. We review different examples of non-fundamental representations in the moving average dynamics given in the literature. Next we discuss the identification issues in the Gaussian case and point out that the standard Box-Jenkins approach based on Gaussian pseudo-likelihood and the Kalman filter algorithm suffer from these issues.

Section 3 is the core of the paper. We consider linear non-Gaussian SVARMA processes based on serially and instantaneously independent shocks (see e.g. Brockwell and Davis, 1991; Rosenblatt, 2000, for an introduction to linear processes). In the context of these models –called Strong Structural VARMA (or SSVARMA) models hereinafter– we explain that the standard static and dynamic identification problems encountered in the Gaussian SVARMA analysis disappear; we also discuss the identification of the structural shocks and of the Impulse Response Functions (IRFs). In Section 4 we present new parametric and semi-parametric estimation methods to improve upon the standard SVAR methodology. A key element is a novel filtering algorithm aimed at estimating the structural shocks from samples of endogenous variables; this procedure is shown to provide consistent estimates of the structural shocks, irrespective of the moving-average fundamentalness regime. The algorithm further makes it possible to compute truncated log-likelihood functions, opening the way to Maximum Likelihood (ML) estimation. We also propose a consistent 2-step semi-parametric approach which is less subject to the curse of dimensionality than the ML approach and which does not require particular assumptions about the distribution of the shocks.

Applications are provided in Section 5. First, we conduct Monte-Carlo analyses to illustrate the performances of our estimation approaches in the contexts of a univariate MA(1) process and of a bivariate VMA(1) process. These exercises suggest that, even for relatively small samples, our estimation approaches can recover the right fundamentalness regime when the shocks feature asymmetry or fat tails. Identification however weakens when the shocks' distributions get closer to the normal distribution (in particular when increasing the number of degrees of freedom of a Student distribution). Second, following Blanchard and Quah (1989), Lippi and Reichlin (1993, 1994), we study the joint dynamics of U.S. GNP growth and unemployment rates; our results

suggest that the data call for non-fundamental bivariate VARMA models. Section 6 concludes.

Proofs are gathered in the appendix. The special case of a one-dimensional MA(1) process is completely analysed in an online appendix, which also contains additional proofs and examples, as well as details on the numerical optimization procedure implemented in the codes.

## 2 Dynamic Linear Model and Non-Fundamentality

### 2.1 The dynamic SSVARMA model

Despite the standard Vector Autoregressive (VAR) terminology, the linear dynamic reduced-form structural models may have both autoregressive and moving average parts. The VARMA model is the following:

$$\Phi(L)Y_t = \Theta(L)\varepsilon_t, \quad (2.1)$$

where  $Y_t$  is a  $n$ -dimensional vector of observations at date  $t$ ,  $\varepsilon_t$  is a  $n$ -dimensional vector of errors,  $L$  the lag operator,

$$\Phi(L) = I - \Phi_1 L - \dots - \Phi_p L^p, \quad \Theta(L) = I - \Theta_1 L - \dots - \Theta_q L^q, \quad (2.2)$$

$I$  is the identity matrix, and the matrix autoregressive and moving average lag-polynomials are of degrees  $p$  and  $q$ , respectively.

Let us now introduce the following assumptions on model (2.1):

**Assumption A.1. Assumption on errors.**

*The errors can be written as  $\varepsilon_t = C\eta_t$ , where  $C$  is invertible, the  $\eta_t$ 's are independently and identically distributed, and the components  $\eta_{j,t}$  of  $\eta_t = C^{-1}\varepsilon_t$  are mutually independent with mean zero and unit variance, i.e.  $E(\eta_{j,t}) = 0$  and  $V(\eta_{j,t}) = 1$ ,  $j = 1, \dots, n$ .*

The independent random variables  $\eta_{j,t}$ ,  $j = 1, \dots, n$ , are called “structural shocks” and the following representation is called a strong structural VARMA, or SSVARMA, representation:

$$\Phi(L)Y_t = \Theta(L)C\eta_t. \quad (2.3)$$

Note that we assume not only the serial and instantaneous non-correlation of the  $\eta_{j,t}$ 's — as in the standard SVARMA — but also their independence.

**Assumption A.2. Assumption of left coprimeness on the lag-polynomials.**

If  $\Phi(L)$  and  $\Theta(L)$  have a left common factor  $C(L)$ , say, such that:  $\Phi(L) = C(L)\tilde{\Phi}(L)$ ,  $\Theta(L) = C(L)\tilde{\Theta}(L)$ , then  $\det(C(L))$  is independent of  $L$ .

This condition ensures that the VARMA representation is minimal in the sense that all possible simplifications have been already done (see [Hannan and Deistler, 1996](#), Chap 2 for more details). This condition greatly simplifies the discussions in the next sections. It is often forgotten in structural settings and it might be necessary to test for the minimality of the representation. This is out of the scope of this paper.<sup>1</sup>

The next assumption, on the roots of  $\det \Phi(z)$ , is also made to simplify our analysis.<sup>2</sup>

**Assumption A.3. Assumption on the autoregressive polynomial.**

All the roots of  $\det \Phi(z)$  have a modulus strictly larger than 1.

Under Assumptions [A.1–A.3](#), the linear dynamic system [\(2.1\)](#) has a unique strongly stationary solution, such that  $E(\|Y_t\|^2) < \infty$  (see e.g. the discussion in [Gouriéroux and Zakoian, 2015](#)). Also note that, if the right-hand side of [\(2.1\)](#) is  $\mu + \Theta(L)\varepsilon_t$ , the process  $Y_t - m \equiv Y_t - [\Phi(1)]^{-1}\mu$  satisfies [\(2.1\)](#) without intercept; we can therefore assume  $\mu = 0$ , or  $m = 0$ , without loss of generality.

**Assumption A.4. Assumption on the observable process.**

The observable process is the (strong) stationary solution of model [\(2.1\)](#) associated with the true values of  $\Phi$ ,  $\Theta$ ,  $C$  and with the true distribution of  $\varepsilon$ .

Since all the roots of  $\det(\Phi(z))$  lie outside the unit circle, it is easy to derive the inverse of the

<sup>1</sup>See [Deistler and Schrader \(1979\)](#) for a study of identifiability without coprimeness, and [Gouriéroux, Monfort, and Renault \(1989\)](#) for the test of coprimeness –i.e. common roots– for one-dimensional ARMA processes.

<sup>2</sup>This assumption excludes cointegrated variables. When  $Y_t$  is  $I(1)$  and, assuming that  $\beta$  is a  $n \times r$  matrix whose columns span the cointegrating space (of dimension  $r$ ), one can come back to the present (stationary) case by considering the following stationary vector of variables:  $W_t = [\tilde{Y}_t', Y_t' \beta']'$ , where  $\tilde{Y}_t = [\Delta Y_{1,t}, \dots, \Delta Y_{n-r,t}]'$ . Note that we can assume that the  $\Delta$  operator applies to the  $n - r$  first components of  $Y_t$  since we can always reorder and rename the components. [Engle and Granger \(1987\)](#)'s least square methodology provides a consistent estimate of the cointegration directions  $\beta$  (at rate  $1/T$ ). Let's denote this estimate by  $\hat{\beta}$ . The estimation approaches that are presented in [Section 4](#) can then be applied to the stationarized process  $\hat{W}_t = [\tilde{Y}_t', Y_t' \hat{\beta}]'$ ; the effect of estimating the matrix of cointegrating directions  $\beta$  can be neglected due to the high convergence speed of  $\hat{\beta}$ .

autoregressive polynomial operator  $\Phi(L)$  as a convergent one-sided series in the lag operator  $L$ :

$$\begin{aligned}\Phi(L)Y_t &= \Theta(L)\varepsilon_t \\ \iff Y_t &= \Phi(L)^{-1}\Theta(L)\varepsilon_t \equiv \Psi(L)\varepsilon_t = \sum_{k=0}^{\infty} \Psi_k L^k \varepsilon_t = \sum_{k=0}^{\infty} \Psi_k \varepsilon_{t-k} \\ &= \sum_{k=0}^{\infty} \Psi_k C \eta_{t-k} = \sum_{k=0}^{\infty} A_k \eta_{t-k},\end{aligned}\tag{2.4}$$

with  $A_k = \Psi_k C$ . Hence, the  $A_k$ 's are combinations of the  $\Psi_k$ 's, which determine the dynamics of the system, and of  $C$ , which defines the instantaneous impact of the structural shocks.

Moreover, when all the roots of  $\det(\Theta(z))$  lie outside the unit circle,  $Y_t$  has a one-sided autoregressive representation:

$$\Theta^{-1}(L)\Phi(L)Y_t \equiv \sum_{k=0}^{\infty} B_k L^k Y_t = \sum_{k=0}^{\infty} B_k Y_{t-k} = \varepsilon_t,$$

and

$$\eta_t = C^{-1}\Theta^{-1}(L)\Phi(L)Y_t,$$

where  $\Theta^{-1}(L)$  is the one-sided series operator involving positive powers of  $L$  and that satisfies  $\Theta^{-1}(L)\Theta(L) = I$ . In this case, we say that the operator  $\Theta(L)$  is invertible and that the SSVARMA model (2.3) is fundamental.

However, from the macroeconomic literature we know that VARMA models do not always have roots of the moving average located outside the unit circle (see Section 2.2). We also make the following assumption:

**Assumption A.5. Assumption on the moving average polynomial**

*The roots of  $\det(\Theta(z))$  are not on the unit circle.*

In the MA(1) case  $Y_t = (I - \Theta L)C\eta_t$ , this excludes eigenvalues of  $\Theta$  on the unit circle, and also the differencing of a cointegrated VARMA process. Under Assumption A.5, the roots of  $\det(\Theta(L))$  can be inside or outside the unit circle;  $\Theta(L)$  is then invertible in a general sense, since there exists a two-sided series  $B(L) = \sum_{k=-\infty}^{\infty} B_k L^k$  such that  $B(L)\Theta(L) = I$ , and  $Y_t$  then has a two-sided autoregressive representation. The structural shocks can then be recovered from all lagged, current, and future values of the observed variables.

Let us now study the consequences of “ill-located” roots of  $\det(\Theta(z))$ , that are roots located inside the unit circle. For expository purpose, let us consider a one-dimensional ARMA(1,1) process:

$$(1 - \varphi L)y_t = (1 - \theta L)\varepsilon_t, \quad (2.5)$$

where  $|\varphi| < 1$  and  $|\theta| > 1$ . We have:

$$y_t = (1 - \varphi L)^{-1}(1 - \theta L)\varepsilon_t, \quad (2.6)$$

and, therefore,  $y_t$  is a function of the present and past values of  $\varepsilon_t$ .

To get the (infinite) pure autoregressive representation of process  $(y_t)$ , we have to invert  $1 - \theta L$ . This leads to:

$$\begin{aligned} (1 - \varphi L)y_t &= \left(1 - \frac{1}{\theta}L^{-1}\right)(-\theta L\varepsilon_t) \\ \Leftrightarrow \left(1 - \frac{1}{\theta}L^{-1}\right)^{-1}(1 - \varphi L)y_t &= -\theta L\varepsilon_t. \end{aligned} \quad (2.7)$$

Formula (2.7) implies that  $\varepsilon_t$  is a function of the future values of  $y_t$ . Therefore,  $\varepsilon_t$  is not the causal innovation of  $y_t$ , defined by  $y_t - E(y_t|y_{t-1}, y_{t-2}, \dots)$ , the latter being a function of present and past values of  $y_t$  only. Formula (2.6) shows that the knowledge of  $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$  implies the knowledge of  $\{y_t, y_{t-1}, \dots\}$ , but since (2.7) shows that  $\varepsilon_t$  is not a function of  $\{y_t, y_{t-1}, \dots\}$ , it comes that the information set  $\{y_t, y_{t-1}, \dots\}$  is strictly included in the information set  $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$ .

To summarize, under Assumptions A.1–A.4, the error term in the VARMA representation is equal to the causal innovation of the process if the roots of  $\det(\Theta(z))$  are all outside the unit circle. Under this condition, the process  $(Y_t)$  has a fundamental VARMA representation.<sup>3,4</sup> In this case, at any date  $t$ , the information contained in the current and past values of  $Y_t$  coincides with the information contained in the current and past values of  $\varepsilon_t$ . Otherwise, if some roots of  $\det(\Theta(z))$  are inside the unit circle, the VARMA representation is non-fundamental. In the latter case,  $\varepsilon_t$  is not equal to the causal innovation, that is, it is not function of present and past observations of  $Y_t$  only.

Consider a non-fundamental SSVARMA process  $(Y_t)$  defined by (2.3).  $(Y_t)$  admits a fundamental representation of the form:  $\Phi(L)Y_t = \Theta^*(L)\varepsilon_t^*$ ,<sup>5</sup> where process  $(\varepsilon_t^*)$  is a weak white noise,

<sup>3</sup>See e.g. Hansen and Sargent (1980), p18, (1991), p79, and Lippi and Reichlin (1994) for the introduction of this terminology in the macroeconomic literature. The term “fundamental” is likely due to Kolmogorov and appears in Rozanov (1960), p367, and Rozanov (1967), p56, to define the “fundamental process”, that is, the second-order white noise process involved in the Wold decomposition of a weak stationary process.

<sup>4</sup>The terminology “fundamental” can be misleading, in particular since fundamental shock and structural shock are often considered as equivalent notions (see e.g. the description of the scientific works of Nobel prizes Sargent and Sims in Economic Sciences Prize Committee, 2011, or Evans and Marshall, 2011). Moreover in some papers (see Grassi, Ferroni, and Leon-Ledesma, 2015) a shock is called fundamental if its standard deviation is non-zero.

<sup>5</sup> This representation can be obtained from the non-fundamental SVARMA representation by making use of

i.e.  $\varepsilon_t^*$  and  $\varepsilon_s^*$ ,  $t \neq s$ , are uncorrelated but not independent, except in the Gaussian case. This process is the linear causal innovation of  $(Y_t)$  appearing in the Wold representation, but is in general different from the nonlinear causal innovation, except in special cases, such as the Gaussian case. In any case, it does not coincide with process  $(C\eta_t)$ . Now, consider the new (fundamental) process  $(Y_t^*)$  defined through  $\Phi(L)Y_t^* = \Theta^*(L)C^*\eta_t$ , where  $C^*$  satisfies  $V(\varepsilon_t^*) = C^*C^{*'$ . Processes  $(Y_t)$  and  $(Y_t^*)$  have the same (dynamic) second-order properties. As a result, the estimation methods focussing on second-order properties cannot distinguish between  $\Theta^*(L)$  and  $\Theta(L)$ . However, the dynamic responses of  $Y_t$  and  $Y_t^*$  to changes in  $\eta_t$  resulting from one or the other MA specification are different.<sup>6</sup>

## 2.2 Examples of non-fundamentalness

There exist different sources of non-fundamentalness in SVARMA models, that is, of “ill-located” roots of the moving average polynomial (see also the discussion in [Alessi, Barigozzi, and Capasso, 2011](#)). Let us consider some examples and highlight, in each case, the errors with a structural interpretation.

- i) **Lagged impact.** A first example of non-fundamentalness is the one provided by [Lippi and Reichlin \(1993\)](#) in their comment of the model estimated by [Blanchard and Quah \(1989\)](#). Suppose that the productivity process, denoted by  $y_t$ , is given by:

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1},$$

where  $\varepsilon_t$  denotes the productivity shock, reflecting for instance the introduction of a technological innovation. It may be realistic to assume that the maximal impact of the productivity shock is not instantaneous and is maximal with a lag, i.e.  $\theta > 1$ . The MA(1) process is then non-fundamental (or non-invertible).

- ii) **Advanced indicator (noise / news shocks).** The non-invertibility of the process may also arise from some modelling of the information structure. Consider a process  $(x_t)$  that summarizes

---

Blaschke matrices. Consider a square matrix of the lag operator denoted by  $B(L)$ .  $B(L)$  is a Blaschke matrix if and only if  $[B(L)]^{-1} = B^*(L^{-1})$ , where  $B^*(\cdot)$  is obtained from  $B(\cdot)$  by transposing and taking conjugate coefficients. See [Leeper, Walker, and Yang \(2013\)](#), p1123-1124 for a practical use of Blaschke matrices.

<sup>6</sup>This is easily illustrated in the context of a univariate MA(1) process. Consider for instance the MA(1) processes  $(y_t)$  and  $(y_t^*)$  respectively defined by  $y_t = \sigma\eta_t - \theta\sigma\eta_{t-1}$  and  $y_t^* = \sigma\theta\eta_t - \sigma\eta_{t-1}$ . Although  $(y_t)$  and  $(y_t^*)$  have the same second-order properties, they react differently to shocks to  $\eta_t$ . Whereas this shock implies increases in  $y_t$ ,  $y_{t+1}$  and  $y_{t+h}$ ,  $h > 1$ , by  $\sigma$ ,  $-\theta\sigma$  and 0, respectively, it implies increases in  $y_t^*$ ,  $y_{t+1}^*$  and  $y_{t+h}^*$  by  $\sigma\theta$ ,  $-\sigma$  and 0. In particular, if one of these two IRFs is decreasing (in absolute values), the other is hump-shaped.

“fundamentals”, which themselves capture changes in technology, preferences, endowments, or government policy (Chahrouh and Jurado, 2018). The dynamics of  $(x_t)$  is described by:

$$x_t = a(L)u_t, \quad (2.8)$$

where process  $(u_t)$  is a strong white noise. On date  $t$ , the consumer observes  $x_t$  as well as a “noisy” signal about the future value of fundamentals:

$$s_t = x_{t+1} + v_t, \quad (2.9)$$

where  $(v_t)$  is also a strong white noise, independent of  $(u_t)$ . This second equation describes the incremental information about future fundamentals (Barsky and Sims, 2012). If the econometrician also observes the signal, the model leads to the following bivariate moving average representation:

$$\begin{bmatrix} x_t \\ s_t \end{bmatrix} = \begin{bmatrix} La(L) & 0 \\ a(L) & 1 \end{bmatrix} \begin{bmatrix} u_t^a \\ v_t \end{bmatrix}, \quad (2.10)$$

where  $u_t^a = u_{t+1}$ .

The determinant of the moving average polynomial has a root equal to zero, which is within the unit circle. The moving average representation is therefore non-fundamental. The noise shock  $v_t$  has the nature of a demand shock: although it has no effect on fundamentals, it may affect the macroeconomy because it affects beliefs about future fundamentals through the signal  $s_t$ .<sup>7</sup> Hence, the existence of this noise may be used to account for belief-driven business cycles (see e.g. Forni, Gambetti, Lippi, and Sala, 2017b).

This type of analysis is also employed in models disentangling the permanent and transitory shocks affecting fundamentals (Lorenzoni, 2009; Blanchard, L’Huillier, and Lorenzoni, 2013), or own agent productivity and noisy signal on the aggregate productivity (Lorenzoni, 2009, Section 3). Chahrouh and Jurado (2018) have recently shown that the previous kind of “noise representation” closely relates to the so-called “news representation”, where people perfectly observe part of an exogenous fundamental in advance.<sup>8</sup>

**iii) Rational expectations.** Other sources of non-fundamentalness are the rational expectations introduced in the models. In the simple example of Hansen and Sargent (1991) the economic

<sup>7</sup>The “noise” terminology can be misleading: although “noise” mainly refers to  $v_t$  in this literature, both types of disturbances ( $u_t$  and  $v_t$ ) are strong white noises, i.e. sequences of i.i.d. variables.

<sup>8</sup>Basic references on the impacts of “news/noise” shocks include (Blanchard, L’Huillier, and Lorenzoni, 2013; Forni, Gambetti, Lippi, and Sala, 2017a,b; Chahrouh and Jurado, 2018).

variable  $y_t$  is defined as:

$$y_t = E_t \left( \sum_{h=0}^{\infty} \beta^h w_{t+h} \right), \text{ with } w_t = \varepsilon_t - \theta \varepsilon_{t-1}, 0 < \beta < 1, |\theta| < 1.$$

If the information set available to the economic agent at date  $t$  is  $I_t = (\varepsilon_t, \varepsilon_{t-1}, \dots)$ , we get:

$$y_t = (1 - \beta\theta)\varepsilon_t - \theta\varepsilon_{t-1}.$$

The root of the moving average polynomial is  $(1 - \beta\theta)/\theta$ . Depending on the values of  $\beta$  and  $\theta$ , the absolute value of this root is larger or smaller than 1. When the root is strictly smaller than 1, the model is non-fundamental. In such rational expectation models, the information sets of the economic agent and of the econometrician are assumed to be aligned.

**v) Prediction error.** When the variable of interest is interpreted as a prediction error, non-fundamentalness may also appear (see [Hansen and Hodrick, 1980](#)). For instance if  $y_t$  is the price of an asset at  $t$ ,  $E_{t-2}y_t$  can be interpreted as the futures price at  $t-2$  (if the agents are risk-neutral). The spread between the spot price and the futures price is:  $s_t = y_t - E_{t-2}y_t$  and, if  $y_t$  is a fundamental (invertible) MA(2) process:  $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} = \Theta(L)\varepsilon_t$ , then  $s_t = \varepsilon_t + \theta_1\varepsilon_{t-1} (= \Theta_1(L)\varepsilon_t)$  and the spread process is not necessarily fundamental. For example if  $\Theta(L) = (1 - \theta L)^2$  with  $|\theta| < 1$ , we have  $\Theta_1(L) = 1 - 2\theta L$ , which is not invertible as soon as  $|\theta| > 1/2$ .

**vi) Non-observability.** Non-fundamentalness can arise from a lack of observability. [Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson \(2007\)](#) give the example of a state-space representation of the surplus in a permanent income consumption model (see [Lof, 2013](#), Section 3, for another example). The state-space model is of the following type:

$$\begin{cases} c_t &= ac_{t-1} + (1 - 1/R)\varepsilon_t, & 0 < a < 1, \\ y_t &= -ac_{t-1} + 1/R\varepsilon_t, \end{cases}$$

where  $y_t - c_t$  is the surplus, the consumption  $c_t$  is latent,  $R > 1$  is a constant gross interest rate on financial assets, and  $\varepsilon_t$  is an i.i.d. labor income process. From the first equation, we deduce:

$$c_t = \frac{(1 - 1/R)}{1 - aL} \varepsilon_t,$$

and, by substituting in the second equation, the dynamics of  $y_t$  reads:

$$y_t = \left[ 1/R - a \frac{L(1-1/R)}{1-aL} \right] \varepsilon_t = \frac{1/R - aL}{1-aL} \varepsilon_t.$$

Thus the root of the moving-average lag-polynomial is equal to  $1/aR$ , and it is smaller than one when  $aR > 1$ .<sup>9</sup>

## 2.3 The limits of the Gaussian approach

### 2.3.1 Static identification issue

Let us first consider the popular case of a structural VAR process (SVAR), where  $\Theta(L) = I$ . The Gaussian SVAR process is defined by:

$$\Phi(L)y_t = C\eta_t,$$

where  $\Phi(0) = I$ , the roots of  $\det(\Phi(z))$  are outside the unit circle and where the process  $(\eta_t)$  is a Gaussian white noise, with  $E(\eta_t) = 0$  and  $V(\eta_t) = I$ . The model involves two types of parameters: whereas the sequence  $\Phi_k$ ,  $k = 1, \dots, p$ , characterizes the dynamic features of the model, matrix  $C$  defines (static) instantaneous effects.

In this case, the dynamic parameters  $\Phi_1, \dots, \Phi_p$  are characterized by the Yule-Walker equations; they are therefore identifiable, but the static parameter  $C$  is not, since replacing  $C$  by  $CQ$ , where  $Q$  is an orthogonal matrix, leaves the distribution of process  $(Y_t)$  unchanged. It is the *static identification problem*.

In order to solve this identification issue, additional short-run, long-run, or sign restrictions have been imposed in the literature [see e.g. the references in the introduction].<sup>10</sup> It turns out that, if at most one of the components of  $\eta_t$  is Gaussian, the identification problem disappears since  $C$  is then identifiable, up to sign change and permutation of its columns. This result, shown by [Comon \(1994\)](#) (Theorem 11) is a consequence of the Darmois-Skitovich characterization of the multivariate Gaussian distribution (see [Darmois, 1953](#); [Skitovich, 1953](#); [Ghurye and Olkin, 1961](#)). In this case,  $C$  can be estimated using Independent Component Analysis (ICA) algorithms (see [Hyvärinen, Karhunen, and Oja, 2001](#)), or Pseudo Maximum Likelihood techniques (see [Gouriéroux, Monfort,](#)

<sup>9</sup>This reasoning does not hold for  $a = 1$ , which is precisely the case considered in [Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson \(2007\)](#), where  $c_t$  and  $y_t$  are nonstationary co-integrated processes. Indeed their equation (5) assumes the stationarity of the  $y$  process and is not compatible with the assumption of a cointegrated model.

<sup>10</sup>An alternative consists in leaving the linear dynamic framework by considering Markov Switching SVAR [see [Lanne, Lütkepohl, and Maciejowska \(2010\)](#), [Lütkepohl \(2013\)](#), [Herwartz and Lütkepohl \(2017\)](#), [Velinov and Chen \(2014\)](#)]. In this paper we will stay in a pure SVARMA framework.

and Renne, 2017 or Lanne, Meitz, and Saikkonen, 2017). The identifiability of the static parameter  $C$  in the non-Gaussian case implies that the recursive approach proposed by Sims, imposing that  $C$  is lower-triangular, cannot be used in general to find independent shocks, but only uncorrelated shocks.

### 2.3.2 Dynamic identification issue

Let us now consider the general case of a SVARMA process:

$$\Phi(L)Y_t = \Theta(L)C\eta_t,$$

where  $\Phi(0) = \Theta(0) = I$ , the roots of  $\det(\Phi(z))$  lie outside the unit circle, the roots of  $\det(\Theta(z))$  can be inside or outside the unit circle, and  $(\eta_t)$  is a Gaussian white noise with  $E(\eta_t) = 0$  and  $V(\eta_t) = I$ .

Let us focus here on the identification of the dynamic parameters  $\Phi$  and  $\Theta$ . In the Gaussian case, the distribution of the stationary process  $(Y_t)$  depends on the dynamic and static parameters through the second-order moments of the process or, equivalently, through the spectral density matrix:

$$f(\omega) = \frac{1}{2\pi} \Phi^{-1}(\exp i\omega) \Theta(\exp i\omega) C C' \Theta(\exp -i\omega)' \Phi^{-1}(\exp -i\omega)'. \quad (2.11)$$

Using the equalities  $\Gamma_h - \Phi_1 \Gamma_{h-1} - \dots - \Phi_p \Gamma_{h-p} = 0$ ,  $\forall h \geq q + 1$ , with  $\Gamma_h = \text{cov}(Y_t, Y_{t-h})$ , it is readily seen that the coefficient matrices  $\Phi_1, \dots, \Phi_p$  are identifiable from the distribution of process  $(Y_t)$  (Gaussian or not), but several sets of coefficients  $(\Theta_1, \dots, \Theta_q, C)$  yield the same spectral density and the same distribution for the process  $(Y_t)$  in the Gaussian case; the different polynomials  $\Theta(L)$  are obtained from the fundamental one –the one with the roots of  $\det(\Theta(z))$  outside the unit circle– by use of Blaschke matrices.<sup>11</sup> The lack of identification of the dynamic parameters  $\Theta$  is called the *dynamic identification problem*. We see in Section 3 that this second identification problem also disappears in the non-Gaussian case.

The dynamic identification problem is simply illustrated in the univariate MA(1) model:  $y_t = \sigma \eta_t - \theta \sigma \eta_{t-1}$ , where  $(\eta_t)$  is a Gaussian white noise with  $E(\eta_t) = 0$ ,  $V(\eta_t) = 1$  and, for instance,  $0 < \theta < 1$ . If we replace  $\theta$  by  $\theta^* = 1/\theta$  and  $\sigma$  by  $\sigma^* = \sigma \theta$ , we get another process:

$$\begin{aligned} y_t^* &= \sigma^* \eta_t - \theta^* \sigma^* \eta_{t-1} \\ &= \sigma \theta \eta_t - \sigma \eta_{t-1}, \end{aligned}$$

<sup>11</sup>See Footnote 5 for the definition of Blaschke matrices.

which is also Gaussian and with the same covariance function as  $(y_t)$ , namely:

$$\Gamma_0 = \sigma^2(1 + \theta^2), \quad \Gamma_1 = -\theta\sigma^2 \quad \text{and} \quad \Gamma_h = 0, \text{ for } h \geq 2,$$

and, therefore, with the same distribution. In other words, the pairs  $(\theta, \sigma)$  and  $(1/\theta, \sigma\theta)$  give the same distributions for processes  $(y_t)$  and  $(y_t^*)$ . By contrast, we will see that, if  $\eta_t$  is non-Gaussian, the distributions of processes  $(y_t)$  and  $(y_t^*)$  are different, although their spectral density matrices are the same (see e.g. [Weiss, 1975](#); [Breidt and Davis, 1992](#); [Lii and Rosenblatt, 1992](#)).

Alternatively, the process  $(\eta_t^*)$  defined by  $y_t = \sigma\theta\eta_t^* - \sigma\eta_{t-1}^* \equiv \sigma^*(\eta_t^* - \frac{1}{\theta}\eta_{t-1}^*)$ , with  $\sigma^* = \sigma\theta$  is another Gaussian white noise with zero mean and unit variance. More generally in the MA( $q$ ) case,  $y_t = \theta(L)\sigma\eta_t$ , with  $\theta(0) = 1$ , we get other representations (i) by replacing  $\theta(L)$  by  $\theta^*(L)$ , where the latter is obtained from  $\theta(L)$  by inverting some roots (the complex roots being inverted by pairs), and (ii) by replacing  $\sigma$  by  $\sigma^*$ , where the value of  $\sigma^*$  is determined to give the same marginal variance to  $y_t$ . Then the processes  $(\eta_t^*)$  defined through  $y_t = \theta^*(L)\sigma^*\eta_t^*$  are Gaussian white noises with zero mean and unit variance. Among all these equivalent representations, a single one is fundamental. In the non-Gaussian case, if one of these  $(\eta_t^*)$  processes is a strong white noise, i.e. a serially independent process, the others will be only weak white noises, i.e. serially uncorrelated (see [Proposition 1](#) below). If the strong white noise process does not correspond to the fundamental representation, the weak white noise corresponding to the invertible polynomial  $\theta^*(L)$  is the linear innovation process associated with the Wold representation.

In the usual Box-Jenkins approach, the estimation of the parameters  $\Phi_1, \dots, \Phi_p, \Theta_1, \dots, \Theta_q, \Sigma = CC'$  is based on a truncated VAR approximation relying on the assumption that  $\Theta(L)$  is invertible (i.e. the roots of  $\det(\Theta(z))$  are outside the unit circle), namely a truncation of  $\Theta(L)^{-1}\Phi(L)Y_t = \varepsilon_t$ , with  $V(\varepsilon_t) = \Sigma = CC'$  (see e.g. [Galbraith, Ullah, and Zinde-Walsh, 2002](#)). In other words, a fundamental representation is a priori imposed without being tested. The introduction of multivariate non-fundamentalness tests is actually very recent (see e.g. [Forni and Gambetti, 2014](#); [Chen, Choi, and Escanciano, 2017](#)).<sup>12</sup>

<sup>12</sup>The test developed by [Chen, Choi, and Escanciano \(2017\)](#) exploits the non-normality of the shocks; [Forni and Gambetti \(2014\)](#) use information not included in the VAR specification.

### 3 Identification and Impulse Response Functions (IRFs) in the Non-Gaussian SSVARMA

#### 3.1 Identification of the parameters

Consider again the SSVARMA process:

$$\Phi(L)Y_t = \Theta(L)C\eta_t,$$

and assume that Assumptions A.1 to A.5 hold. Since  $\Phi(L)$  is invertible, we have:

$$Y_t = \Phi^{-1}(L)\Theta(L)C\eta_t = A(L)\eta_t, \quad (3.1)$$

with  $A(L) = \Phi^{-1}(L)\Theta(L)C$ .

In Subsection 2.3.1, we have seen that  $\Phi_1, \dots, \Phi_p$  are characterized by the Yule-Walker equations; they are therefore identified. What about  $\Theta(L)$  and  $C$ ? The next proposition is deduced from Theorem 1 in Chan, Ho, and Tong (2006) (based on Theorem 4 in Chan and Ho, 2004), and solves the dynamic identification issue in the non-Gaussian case.<sup>13,14</sup> Let us first introduce the following assumption:

**Assumption A.6.** *Each component of  $\eta_t$  has a non-zero  $r^{\text{th}}$  cumulant, with  $r \geq 3$ , and a finite moment of order  $s$ , where  $s$  is an even integer greater than or equal to  $r$ .*

Assumption A.6 is introduced to eliminate the Gaussian framework in which all cumulants of order  $r \geq 3$  are zero. More precisely, if the true distribution of  $\eta_{j,t}$  has moments of any order, the only distribution that does not satisfy Assumption A.6 is the Gaussian distribution. Assumption A.6 is also satisfied for asymmetric distributions, that is, if the true distribution of  $\eta_{j,t}$  has a non-zero skewness (and has a finite moment of order 4), or if it is symmetric, but has a non-zero (finite) excess kurtosis. This includes the Student distributions with degrees of freedom strictly larger than 4.

**Proposition 1.** *Under Assumptions A.1 to A.6, if we consider another stationary process  $(Y_t^*)$  defined by:*

$$\Phi(L)Y_t^* = \Theta^*(L)C^*\eta_t^*,$$

<sup>13</sup>See Findley (1986), Cheng (1992) for the one-dimensional case.

<sup>14</sup>A similar identification result has been recently derived when the components of  $\eta_t$  have fat tails (see Gouriéroux and Zakoian, 2015).

then the processes  $(Y_t)$  defined in (2.3) and  $(Y_t^*)$  are observationally equivalent if and only if:

$$\Theta(L) = \Theta^*(L) \text{ and } C = C^*,$$

where the last equality holds up to a permutation and sign change of the columns and  $\eta_t^* = \eta_t$  in distribution up to the same permutation and sign change of their components.

**Proof:** See Appendix A. □

In order to transform the local identifiability of  $C$  under Assumptions A.1 to A.6 into a global identifiability, we add the following normalization:<sup>15</sup>

**Assumption A.7.** *The components of the first row of  $C$  are nonnegative and in increasing order.*

Proposition 1 provides conditions under which the SSVARMA parameterization is identified. This identification result is however not constructive and does not explain how to estimate the correct –possibly non-fundamental– SSVARMA representation.<sup>16</sup> This task is the objective of the methods presented in Section 4 below.

## 3.2 Identification of the structural shocks and of the IRFs

What precedes implies that, under Assumptions A.1 to A.7, the structural shocks  $\eta_t$  are identified. Indeed, we have  $\eta_t = C^{-1}\Theta^{-1}(L)\Phi(L)Y_t$  where, under these assumptions,  $\Phi(L)$ ,  $\Theta(L)$  and  $C$  are identified. The only remaining problem is the labelling of these shocks, since these shocks are ordered by Assumption A.7, but not yet interpreted. The labelling will depend on the dynamic impact of these shocks on the different endogenous variables considered in the model, that is on the Impulse Response Functions (IRFs). This is completely different from standard structural approaches where the identification of the shocks themselves necessitates additional theory-based restrictions. Proposition 1 states that, under Assumptions A.1 to A.7, such restrictions are over-identifying restrictions.

The IRFs are also identified under Assumptions A.1 to A.7 since, in our linear setting, the IRFs correspond to the coefficients of the MA representation (2.4), which are deduced from  $\Theta(L)$ ,  $\Phi(L)$

---

<sup>15</sup>This normalization is an alternative to the one proposed by Lanne, Meitz, and Saikkonen (2017) in the context of non-Gaussian SSVAR (see their Subsection 3.3).

<sup>16</sup>Theorem 1 of Chen, Choi, and Escanciano (2017) is closely related to Proposition 1. Chen, Choi, and Escanciano (2017) exploit their Theorem 1 to define a test aimed at checking if the fundamental representation is the right one. They do not explain how to estimate the SSVARMA representation under non-fundamentalness.

and  $C$ . Specifically, let us denote by  $IRF_{j,h}$  the differential impact of a unit shock on  $\eta_{j,t}$  on  $Y_{t,h}$ , that is:<sup>17</sup>

$$IRF_{j,h} = E\left(Y_{t+h} | \eta_{j,t} = 1, \underline{\eta}_{t-1}\right) - E\left(Y_{t+h} | \underline{\eta}_{t-1}\right), \quad (3.2)$$

where  $\underline{\eta}_{t-1} = \{\eta_{t-1}, \eta_{t-2}, \dots\}$ . Then, from (2.4), we have:<sup>18</sup>

$$IRF_{j,h} = \Psi_h C^j, \quad (3.3)$$

where  $C^j$  denotes the  $j^{th}$  column of matrix  $C$ .

Equation (3.3) highlights that the IRFs hinge on the fundamentalness regime of the SSVARMA model. It shows indeed that the IRF depends on the (possibly non-invertible) MA coefficients through the  $\Psi_h$ 's.<sup>19</sup>

Besides, note that  $IRF_{j,h}$  can be written  $\Psi_h C E(\eta_t | \eta_{j,t} = 1)$  using the serial independence of the  $\eta_t$ 's only. Consequently, although the assumption of instantaneous zero correlation is not sufficient, we also get  $IRF_{j,h} = \Psi_h C^j$  under the weaker assumption of mean independence ( $E(\eta_{i,t} | \eta_{j,t}) = 0, \forall i \neq j$ ). However, the result of Proposition 1 on the identification of  $C$  – and therefore of the IRFs – is based on the assumption of instantaneous independence of the  $\eta_{j,t}$ 's; it has not been shown under the assumption of mean independence.

## 4 Estimation of Models with Non-Fundamentalness

In this section, we present parametric and semi-parametric estimation approaches to estimate (possibly) non-fundamental SSVARMA models. For a parametric SSVARMA model, we show that the parameters can be estimated by Maximum Likelihood (ML). More precisely, we explain how to compute a truncated log-likelihood function. Under regularity assumptions, the estimator resulting from the maximization of the latter function achieves asymptotic efficiency. When the distribution of the errors is left unspecified, we propose a two-step approach whose second step is based on moment restrictions. These restrictions pertain to moments of order higher than three. This semi-parametric approach, which also gives consistent and asymptotically normal estimators, is robust to a misspecification of the error distribution.

<sup>17</sup>Other IRF definitions can be found in the literature. In particular, [Koop, Pesaran, and Potter \(1996\)](#) discuss the following “traditional” definition:  $E(Y_{t+h} | \eta_{t+h} = 0, \dots, \eta_{t+1} = 0, \eta_t = e_j, \underline{\eta}_{t-1}) - E(Y_{t+h} | \eta_{t+h} = 0, \dots, \eta_t = 0, \underline{\eta}_{t-1})$ . In the linear case, these “traditional” IRFs are equivalent to those resulting from (3.2), called Generalized IRF ([Koop, Pesaran, and Potter, 1996](#); [Pesaran and Shin, 1998](#)).

<sup>18</sup>If the SSVARMA is non-fundamental, then  $\underline{\eta}_{t-1}$  is not a function of  $Y_{t-1}$ , and, therefore, is not observable (even if the parameters were known). This is however not a problem for the computation of  $IRF_{j,h}$  since, thanks to the linearity of the model and the serial independence assumptions,  $\underline{\eta}_{t-1}$  does not appear in  $IRF_{j,h}$ .

<sup>19</sup>Footnote 6 shows for instance that, in the MA(1) case, if the IRF associated with the fundamental representation is monotonously decreasing, then the one associated with the non-fundamental representation is hump-shaped.

When the dimension of the vector of endogenous variables  $Y_t$  increases, the number of parameters specifying the SSVARMA representation (2.3) increases at a much faster rate, giving rise to a curse of dimensionality problem. As a result, in practice, there is a tradeoff between the dimension  $n$  and the degrees  $p$  and  $q$  in VARMA modelling. Contrary to our ML approach, the numerical complexity of the semi-parametric approach does not depend on the autoregressive order  $p$  (because the estimate of  $\Phi(L)$  results from linear regressions). Nevertheless, both methods are subject to the curse of dimensionality when augmenting the moving-average order  $q$ . In most applications,  $q$  is equal to 0, 1 or 2. In the following, we focus on VARMA( $p,1$ ) models. Subsection 4.1.4 details how to extend the ML approach to the estimation of VARMA( $p,q$ ) models with  $q > 1$ .

## 4.1 Maximum Likelihood estimation of parametric SSVARMA models

For illustrative purpose, we will first discuss the case of a one-dimensional MA(1) process before considering the general framework of a SSVARMA process. Derivations of truncated log-likelihood functions and associated asymptotic results of ML estimators in the context of possibly noninvertible univariate MA( $q$ ) and ARMA( $p,q$ ) processes can notably be found in [Lii and Rosenblatt \(1992, 1996\)](#), or [Wu and Davis \(2010\)](#). The approach exposed below extends the previous studies to the multivariate case.

### 4.1.1 The Maximum Likelihood approach in the MA(1) context

We consider the MA(1) process:

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1}, \quad (4.1)$$

where the  $\varepsilon_t$ 's are independent and identically distributed.

Suppose that we observe  $\{y_1, \dots, y_T\}$ . If the common distribution of the  $\varepsilon_t$ 's is  $N(0, \sigma^2)$ , the model is not identifiable (see Section 2.3). If  $\varepsilon_t$  satisfies Assumptions A.1 to A.6, i.e. in particular if it is not Gaussian, then Proposition 1 states that the model is identifiable.<sup>20</sup> Let us denote by  $g(\varepsilon; \gamma)$  the common p.d.f. of the  $\varepsilon_t$ 's, where  $\gamma$  is an unknown parameter, and let us consider three cases, depending on the position of  $|\theta|$  with respect to 1:

**i)** When  $|\theta| < 1$ , we can invert equation (4.1) in the standard way in order to get  $\varepsilon_t$  as a function of parameter  $\theta$  and of current and lagged values of process  $Y$  as:

$$\varepsilon_t(\theta) = \sum_{h=0}^{\infty} \theta^h y_{t-h} = \theta^t \varepsilon_0(\theta) + \sum_{h=0}^{t-1} \theta^h y_{t-h}. \quad (4.2)$$

<sup>20</sup>See Section 1 of the online appendix for a more detailed discussion of non-identifiability of a MA(1) process and the links with invertibility.

The truncated log-likelihood function is defined as the log-likelihood based on the truncated observable approximation of  $\varepsilon_t(\theta)$ , that is:

$$\begin{aligned} L_1(y_1^T; \theta, \gamma) &= \sum_{t=1}^T \log g \left( \sum_{h=0}^{t-1} \theta^h y_{t-h}; \gamma \right) \\ &= \sum_{t=1}^T \log g \left( \varepsilon_t(\theta) - \theta^t \varepsilon_0(\theta); \gamma \right), \end{aligned} \quad (4.3)$$

where we use the notation  $y_1^T = \{y_1, \dots, y_T\}$ .

ii) When  $|\theta| > 1$ , equation (4.1) can still be inverted, but in reverse time. We get:

$$\begin{aligned} y_t = \varepsilon_t - \theta \varepsilon_{t-1} &\Leftrightarrow -\frac{y_{t+1}}{\theta} = \varepsilon_t - \frac{1}{\theta} \varepsilon_{t+1} \\ &\Leftrightarrow \varepsilon_t(\theta) = \frac{1}{\theta^{T-t}} \varepsilon_T(\theta) - \sum_{h=1}^{T-t} \frac{1}{\theta^h} y_{t+h}. \end{aligned} \quad (4.4)$$

The truncated log-likelihood function is then defined as:

$$\begin{aligned} L_2(y_1^T; \theta, \gamma) &= \sum_{t=0}^{T-1} \log \left\{ \frac{1}{|\theta|} g \left( -\sum_{h=1}^{T-t} \frac{1}{\theta^h} y_{t+h}; \gamma \right) \right\} \\ &= \sum_{t=0}^{T-1} \log \left\{ \frac{1}{|\theta|} g \left( \varepsilon_t(\theta) - \frac{1}{\theta^{T-t}} \varepsilon_T(\theta); \gamma \right) \right\} \end{aligned} \quad (4.5)$$

where the factor  $1/|\theta|$  comes from the Jacobian formula.

iii) Let us now discuss the case  $|\theta| = 1$ . Though this case is not consistent with Assumption A.5, it has to be considered for analyzing the continuity of the likelihood function on the unit circle.

Focusing on the regimes when truncating the log-likelihood function gives the misleading impression of a lack of continuity of the exact log-likelihood function w.r.t.  $\theta$  at  $|\theta| = 1$ , since  $\varepsilon_t(\theta)$  is not continuous (and not even defined) for  $|\theta| = 1$ . This exact log-likelihood is however continuous.<sup>21</sup> Indeed, using equation (4.2), the joint p.d.f. of  $y_1^T$  given  $\varepsilon_0$  can be written:

$$\prod_{t=1}^T g \left( \sum_{h=0}^{t-1} \theta^h y_{t-h} + \theta^t \varepsilon_0; \gamma \right).$$

<sup>21</sup>An exact log-likelihood is for instance used in the Gaussian case, with  $|\theta| < 1$ , by [Chen, Davis, and Song \(2011\)](#) to analyze the properties of the ML estimator of a moving-average parameter close to non-invertibility.

Since the distribution of  $\varepsilon_0$  is  $g(\varepsilon_0; \gamma)$ , the exact log-likelihood is:

$$\mathcal{L}(y_1^T; \theta, \gamma) = \log \left\{ \int \prod_{t=1}^T g \left( \sum_{h=0}^{t-1} \theta^h y_{t-h} + \theta^t \varepsilon; \gamma \right) g(\varepsilon; \gamma) d\varepsilon \right\}. \quad (4.6)$$

Hence, the exact log-likelihood is generally a differentiable function of  $\theta$ . By contrast, the truncated log-likelihood function, given by:

$$L(y_1^T; \theta, \gamma) = L_1(y_1^T; \theta, \gamma) \mathbb{1}_{|\theta| < 1} + L_2(y_1^T; \theta, \gamma) \mathbb{1}_{|\theta| \geq 1}, \quad (4.7)$$

is only right-differentiable at  $\theta = 1$  (and left-differentiable at  $\theta = -1$ ).

In the simple MA(1) case, the maximum likelihood estimation is conducted by directly maximizing the truncated log-likelihood function (4.7). If  $|\theta| \neq 1$ , the standard asymptotic theory applies (see Proposition 2 below).

#### 4.1.2 The SSVARMA( $p,1$ ) case

Let us consider the SSVARMA( $p,1$ ) model:

$$\Phi(L)Y_t = \varepsilon_t - \Theta\varepsilon_{t-1}, \quad (4.8)$$

where the errors  $\varepsilon_t$  are given by:

$$\varepsilon_t = C\eta_t, \text{ say,}$$

where the  $\eta_t$  are independently, identically distributed, have independent components, and are such that  $E(\eta_t) = 0$  and  $V(\eta_t) = I$ . The distribution of the  $\eta_t$ 's is parameterized with  $\gamma$ , say, where parameter  $\gamma$  is identifiable from the distribution of  $\eta$ . The p.d.f. of the errors  $\varepsilon_t$  is therefore of the form  $g(\varepsilon, \Gamma)$ , with  $\Gamma = (C, \gamma)$ . We assume that this model satisfies Assumptions A.1 to A.7.

Our objective here is to solve system (4.8) in order to get  $\varepsilon_t$ 's as linear combinations of  $Y_{-p+1}^T = \{Y_{-p+1}, \dots, Y_T\}$  and of “state variables” – similar to  $\varepsilon_0$  in (4.2) and to  $\varepsilon_T$  in (4.4) – that are such that small changes in these “state variables” only have a small influence on these linear combinations for  $t$  sufficiently far from the sample bounds (1 and  $T$ ). For expository purpose, we denote these solutions by  $\varepsilon_t$ , without explicitly mentioning their dependence on parameters  $\Phi_1, \dots, \Phi_p, \Theta$ , and on observable variables.

Our approach makes use of the real Schur decomposition of matrix  $\Theta$ :<sup>22</sup>

$$\Theta = A_{\Theta} U_{\Theta} A'_{\Theta} = A_{\Theta} \begin{bmatrix} U_1^{\Theta} & U_{1,2}^{\Theta} & \cdots & & U_{1,K}^{\Theta} \\ 0 & U_2^{\Theta} & U_{2,3}^{\Theta} & \cdots & U_{2,K}^{\Theta} \\ \vdots & \ddots & \ddots & & \vdots \\ & & 0 & U_{K-1}^{\Theta} & U_{K-1,K}^{\Theta} \\ 0 & \cdots & & 0 & U_K^{\Theta} \end{bmatrix} A'_{\Theta}, \quad (4.9)$$

where  $A_{\Theta}$  is orthogonal,  $U_{\Theta}$  is upper block-triangular, and the diagonal blocks  $(U_k^{\Theta}, k \in \{1, \dots, K\})$  are either  $1 \times 1$ , or  $2 \times 2$  blocks, the  $2 \times 2$  blocks corresponding to complex conjugate complex eigenvalues of  $\Theta$ . Under Assumption A.5, the eigenvalues of  $U_{\Theta}$  are not on the unit circle. We denote by  $n_k$  the dimension of  $U_k^{\Theta}$  (with  $n_k \in \{1, 2\}$ ). We assume, without loss of generality, that  $U_1^{\Theta}, \dots, U_s^{\Theta}$  have eigenvalues with modulus strictly larger than 1, whereas  $U_{s+1}^{\Theta}, \dots, U_K^{\Theta}$  have eigenvalues with modulus strictly lower than 1.

Left-multiplying  $\Phi(L)Y_t = \varepsilon_t - \Theta \varepsilon_{t-1}$  by  $A_{\Theta}^{-1} = A'_{\Theta}$ , we get:

$$W_t = \varepsilon_t^* - U_{\Theta} \varepsilon_{t-1}^*, \quad (4.10)$$

where  $W_t = A'_{\Theta} \Phi(L)Y_t$  and  $\varepsilon_t^* = A'_{\Theta} \varepsilon_t$ . Note that both  $W_t$  and  $\varepsilon_t^*$  depend on parameters and on observable variables.

Let us denote by  $\varepsilon_t^{*(1)}$  and  $\varepsilon_t^{*(2)}$  the two vectors that are such that  $\varepsilon_t^* = [\varepsilon_t^{*(1)'}, \varepsilon_t^{*(2)'}]'$ . The dimension of  $\varepsilon_t^{*(1)}$ , that is the non-fundamentalness order, is equal to  $m = n_1 + \dots + n_s$ . The fundamentalness order is equal to  $n - m$ . In the same way, we define  $W_t^{(1)}$  and  $W_t^{(2)}$  that are such that  $W_t = [W_t^{(1)'}, W_t^{(2)'}]'$ ,  $W_t^{(1)}$  being of dimension  $m$ .

With a clear block decomposition of  $U_{\Theta}$ , equation (4.10) writes:

$$\begin{bmatrix} \varepsilon_t^{*(1)} \\ \varepsilon_t^{*(2)} \end{bmatrix} = \begin{bmatrix} W_t^{(1)} \\ W_t^{(2)} \end{bmatrix} + \begin{bmatrix} U_{\Theta}^{(1)} & U_{\Theta}^{(12)} \\ 0 & U_{\Theta}^{(2)} \end{bmatrix} \begin{bmatrix} \varepsilon_{t-1}^{*(1)} \\ \varepsilon_{t-1}^{*(2)} \end{bmatrix},$$

<sup>22</sup>One could also use the real Jordan decomposition for this purpose. Formulas would then actually be slightly simpler. However, the real Jordan decomposition is less commonly available in programming softwares (typically in R). The relative numerical instability of the real Jordan decomposition may account for its absence from usual packages (see e.g. Söderlind, 1999).

which leads to:

$$\boldsymbol{\varepsilon}_t^{*(2)} = \mathbf{W}_t^{(2)} + \mathbf{U}_\Theta^{(2)} \mathbf{W}_{t-1}^{(2)} + \dots + \mathbf{U}_\Theta^{(2)t-1} \mathbf{W}_1^{(2)} + \mathbf{U}_\Theta^{(2)t} \boldsymbol{\varepsilon}_0^{*(2)} \quad (4.11)$$

$$\begin{aligned} \boldsymbol{\varepsilon}_t^{*(1)} &= \left\{ (\mathbf{U}_\Theta^{(1)})^{-1} \right\}^{T-t} \boldsymbol{\varepsilon}_T^{*(1)} - (\mathbf{U}_\Theta^{(1)})^{-1} \mathbf{W}_{t+1}^{(1)} - \dots - \left\{ (\mathbf{U}_\Theta^{(1)})^{-1} \right\}^{T-t} \mathbf{W}_T^{(1)} \\ &\quad - (\mathbf{U}_\Theta^{(1)})^{-1} \mathbf{U}_\Theta^{(12)} \boldsymbol{\varepsilon}_t^{*(2)} - \dots - \left\{ (\mathbf{U}_\Theta^{(1)})^{-1} \right\}^{T-t} \mathbf{U}_\Theta^{(12)} \boldsymbol{\varepsilon}_{T-1}^{*(2)}. \end{aligned} \quad (4.12)$$

Equation (4.11) shows that, once  $\boldsymbol{\varepsilon}_0^{*(2)}$  is known, the  $\boldsymbol{\varepsilon}_t^{*(2)}$ 's can be computed by forward recursions. Since the eigenvalues of  $\mathbf{U}_\Theta^{(2)}$  are strictly smaller than one, the effect of the ‘‘initial state variable’’  $\boldsymbol{\varepsilon}_0^{*(2)}$  tends to zero when  $t$  tends to infinity. The truncated version of  $\boldsymbol{\varepsilon}_t^{*(2)}$  is

$$\boldsymbol{\varepsilon}_t^{*(2)} - \mathbf{U}_\Theta^{(2)t} \boldsymbol{\varepsilon}_0^{*(2)} = \mathbf{W}_t^{(2)} + \mathbf{U}_\Theta^{(2)} \mathbf{W}_{t-1}^{(2)} + \dots + \mathbf{U}_\Theta^{(2)t-1} \mathbf{W}_1^{(2)}. \quad (4.13)$$

The right-hand side of the previous equation demonstrates that this truncated version of  $\boldsymbol{\varepsilon}_t^{*(2)}$  depends on model parameters and on observed variables only.

Equation (4.12) shows that  $\boldsymbol{\varepsilon}_t^{*(1)}$  linearly depends on  $\{\mathbf{W}_{t+1}^{(1)}, \dots, \mathbf{W}_T^{(1)}\}$ , on  $\{\boldsymbol{\varepsilon}_t^{*(2)}, \dots, \boldsymbol{\varepsilon}_{T-1}^{*(2)}\}$  and on  $\boldsymbol{\varepsilon}_T^{*(1)}$ . Since the eigenvalues of  $(\mathbf{U}_\Theta^{(1)})^{-1}$  are strictly smaller than one, the effect of the ‘‘end state variable’’  $\boldsymbol{\varepsilon}_T^{*(1)}$  tends to zero when  $T - t$  tends to infinity. The truncated version of  $\boldsymbol{\varepsilon}_t^{*(1)}$  is obtained by omitting the term in  $\boldsymbol{\varepsilon}_T^{*(1)}$  and by replacing the  $\boldsymbol{\varepsilon}_{t+k}^{*(2)}$ 's,  $k = 0, \dots, T - 1 - t$ , by their truncated versions. This truncated version of  $\boldsymbol{\varepsilon}_t^{*(1)}$ , which depends on model parameters and on observed variables only, is equal to:

$$\begin{aligned} \boldsymbol{\varepsilon}_t^{*(1)} - \left\{ (\mathbf{U}_\Theta^{(1)})^{-1} \right\}^{T-t} \boldsymbol{\varepsilon}_T^{*(1)} + \\ \underbrace{(\mathbf{U}_\Theta^{(1)})^{-1} \left[ \mathbf{U}_\Theta^{(12)} + \dots + \left\{ (\mathbf{U}_\Theta^{(1)})^{-1} \right\}^{T-t-1} \mathbf{U}_\Theta^{(12)} \mathbf{U}_\Theta^{(2)T-t-1} \right]}_{\equiv B_{\Theta, T-t}} \mathbf{U}_\Theta^{(2)t} \boldsymbol{\varepsilon}_0^{*(2)}. \end{aligned} \quad (4.14)$$

Hence, while  $\boldsymbol{\varepsilon}_t$  is given by:

$$\boldsymbol{\varepsilon}_t = \mathbf{A}_\Theta \boldsymbol{\varepsilon}_t^*,$$

its truncated version is:

$$\begin{aligned} &\mathbf{A}_\Theta \left( \boldsymbol{\varepsilon}_t^* - \begin{bmatrix} \left\{ (\mathbf{U}_\Theta^{(1)})^{-1} \right\}^{T-t} \boldsymbol{\varepsilon}_T^{*(1)} + \mathbf{B}_{\Theta, T-t} \mathbf{U}_\Theta^{(2)t} \boldsymbol{\varepsilon}_0^{*(2)} \\ \mathbf{U}_\Theta^{(2)t} \boldsymbol{\varepsilon}_0^{*(2)} \end{bmatrix} \right) \\ &= \mathbf{A}_\Theta \boldsymbol{\varepsilon}_t^* - \mathbf{A}_\Theta \begin{bmatrix} \left\{ (\mathbf{U}_\Theta^{(1)})^{-1} \right\}^{T-t} & \mathbf{B}_{\Theta, T-t} \mathbf{U}_\Theta^{(2)t} \\ 0 & \mathbf{U}_\Theta^{(2)t} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_T^{*(1)} \\ \boldsymbol{\varepsilon}_0^{*(2)} \end{bmatrix}. \end{aligned} \quad (4.15)$$

Note that for any value of  $\Theta$  such that there is no eigenvalue of  $U_\Theta$  on the unit circle, processes  $\varepsilon_t$  and  $\varepsilon_t^*$  –seen as functions of the observed variables and of model parameters– are strictly stationary.

The online appendix develops (4.11) and (4.12) in the bivariate VMA(1) case; in this simple case, one can notably investigate the conditions under which a (structural) shock of interest can be approximated by means of a VAR model (see Sims, 2012; Forni and Gambetti, 2014; Beaudry, Fève, Guay, and Portier, 2015).

### 4.1.3 Truncated Maximum Likelihood estimator in the SSVARMA( $p,1$ ) case

As in the univariate case (Subsection 4.1.1), we base the definition of the truncated log-likelihood on the truncated version of  $\varepsilon_t$ . For convenience, we denote by  $\Lambda = \{\Phi_1, \dots, \Phi_p, \Theta, \Gamma\}$  the parameter to be estimated, with  $\Lambda \in \mathcal{L}$ , where  $\mathcal{L}$  is the parameter set. The truncated log-likelihood is obtained by applying the Jacobian formula. The form of the Jacobian is derived in Appendix B.2. We get the following truncated log-likelihood expression:

$$L_T(\Lambda) = -T \sum_{i=1}^n \log |\lambda_i(\Theta)| \mathbb{1}_{|\lambda_i(\Theta)| > 1} + \sum_{t=1}^T \log g \left( A_\Theta \varepsilon_t^* - A_\Theta \begin{bmatrix} \{(U_\Theta^{(1)})^{-1}\}^{T-t} & B_{\Theta, T-t} U_\Theta^{(2)t} \\ 0 & U_\Theta^{(2)t} \end{bmatrix} \begin{bmatrix} \varepsilon_T^{*(1)} \\ \varepsilon_0^{*(2)} \end{bmatrix}; \Gamma \right) \quad (4.16)$$

where  $\lambda_i(\Theta)$  is the  $i^{\text{th}}$  eigenvalue of  $\Theta$  and where the  $\varepsilon_t^*$  are functions of parameter  $\Lambda$  and of observations.

The truncated maximum likelihood estimator is defined as the parameter  $\hat{\Lambda}_T = \{\hat{\Phi}_1, \dots, \hat{\Phi}_p, \hat{\Theta}, \hat{\Gamma}\}$  that maximizes  $L_T(\Lambda)$ .

We can also introduce the untruncated log-likelihood defined by:

$$L_T^u(\Lambda) = -T \sum_{i=1}^n \log |\lambda_i(\Theta)| \mathbb{1}_{|\lambda_i(\Theta)| > 1} + \sum_{t=1}^T \log g(A_\Theta \varepsilon_t^*; \Gamma). \quad (4.17)$$

The untruncated log-likelihood depends on values of  $Y$  that are not observed. Nevertheless, in order to derive the asymptotic properties of the truncated maximum likelihood estimator  $\hat{\Lambda}_T$ , it will prove useful to consider the solution  $\hat{\Lambda}_T^u$  that maximizes the untruncated log-likelihood  $L_T^u(\Lambda)$ . This solution is called the abstract estimator by Bates and White (1985).

**Lemma 1.** *Under Assumptions A.1 to A.5 and under*

(a.1) *The parameter set  $\mathcal{L}$  is compact.*

(a.2) For any given non-fundamentalness regime, a Schur decomposition continuous in  $\Theta$  is selected.

(a.3) The density functions of the error terms  $\eta_{i,t}$  are strictly positive, differentiable, and such that the derivative of the log-density is Lipschitz with a Lipschitz coefficient that is a continuous function of  $\Gamma$ .

(a.4) There exists  $\delta > 2$  such that  $E(|\eta_{j,t}|^\delta) < \infty$ , for  $i = 1, \dots, n$ ,

then the difference  $\left| \frac{1}{T}L_T(\Lambda) - \frac{1}{T}L_T^u(\Lambda) \right|$  tends almost surely (a.s.) to zero uniformly on the parameter set  $\mathcal{L}$ .

**Proof:** See Online Appendix, Section 3. □

Assumption (a.1) allows for decomposing  $\mathcal{L}$  as  $\cup_s \mathcal{L}_s$ , where  $\mathcal{L}_s$  is the restriction of  $\mathcal{L}$  to the parameters with a given (non)fundamentalness regime.<sup>23</sup> Each  $\mathcal{L}_s$  is itself compact. The number of nonfundamentalness regimes depends on the dimension of  $Y$ . For instance, for  $n = 2$ , we have 5 regimes: one is purely fundamental, two are purely nonfundamental and the last two are mixed (see Appendix B.1). In each regime  $s$ , the nonfundamentalness order  $m(s)$  is constant.

Assumptions (a.3) and (a.4) are compatible with a variety of tail magnitudes, in particular (mixtures of) Gaussian, exponential, Student –with a degree of freedom strictly larger than 3 to satisfy (a.4)– or Pareto tails –with a shape parameter strictly larger than 2 to satisfy (a.4).

The advantage of considering the untruncated log-Likelihood is the possibility to write it with the notations used by [Bates and White \(1985\)](#), that is:

$$\frac{1}{T}L_T^u(\Lambda) \equiv \frac{1}{T} \sum_{t=1}^T q_t(Y; \Lambda),$$

where  $Y$  denotes the complete trajectory of  $Y_t$ , from  $-\infty$  to  $+\infty$ , and  $q_t(Y; \Lambda) = q_0(L^{-t}Y; \Lambda)$ , where  $L$  denotes the lag operator.<sup>24</sup> In particular, the process  $(q_t(Y; \Lambda))$  is strictly stationary and the mean of  $q_t$ , if it exists, is independent of time. Then we can directly apply ([Bates and White, 1985](#), Theorem 2.5):

<sup>23</sup>The partition of the parameter space into regimes is the smallest partition of the parameter space –excluding those parameterizations entailing one or several  $\Theta$ 's eigenvalues of modulus equal to one– with connected components. In the bivariate case, this partition can be obtained by taking the preimage of an  $\mathbb{R}^2$  partition based on the trace and determinant of  $\Theta$  (see Appendix B.1).

<sup>24</sup>The lag operator is denoted by  $T$  in [Bates and White \(1985\)](#), who call it measure-preserving one-to-one operator.

**Lemma 2.** (*Bates and White, 1985, Theorem 2.5*)

Under Assumptions **A.1** to **A.5**, (a.1), (a.2) and

(a.5) There exists an integrable function  $\bar{q}_t(Y)$  such that  $\bar{q}_t(Y) = \bar{q}_0(L^{-t}Y)$  and

$$\sup_{\Lambda \in \mathcal{L}} q_t(Y, \Lambda) \leq \bar{q}_t(Y),$$

then  $\frac{1}{T}L_T^u(\Lambda) - E_0(q_t(Y; \Lambda))$  tends almost surely to zero when  $T$  tends to infinity, uniformly on  $\mathcal{L}$ .

Taken together, Lemmas **1** and **2** prove the uniform a.s. convergence of  $\frac{1}{T}L_T(\Lambda) - E_0(q_t(Y; \Lambda))$ . We can then deduce the consistency of both the estimator  $\hat{\Lambda}_T$  and its untruncated version  $\hat{\Lambda}_T^u$  by applying the standard Wald-Jennrich argument ([Wald, 1949](#); [Jennrich, 1969](#)), which leads to the following proposition:

**Proposition 2.** Under Assumptions **A.1** to **A.7**, (a.1) to (a.5) and

(a.6) The true parameter value  $\Lambda_0 \in \mathcal{L}$ ,

then  $\hat{\Lambda}_T$  (resp.  $\hat{\Lambda}_T^u$ ) converges a.s. to  $\Lambda_0$ .

**Proof:** Indeed  $\hat{\Lambda}_T$  (resp.  $\hat{\Lambda}_T^u$ ) converges a.s. to the solution of the limiting optimization problem:

$$\Lambda_0^* = \operatorname{argmax}_{\Lambda \in \mathcal{L}} E_0[q_t(Y; \Lambda)],$$

which, by the identification Assumptions **A.5** and **A.7**, is equal to  $\Lambda_0$ . □

In the above, we have focused on the proof of the strong consistency of the truncated ML estimator. When the true value  $\Lambda_0$  is in interior of  $\mathcal{L}$ , it is in the interior of one of the  $\mathcal{L}_s$ , and the asymptotic normality of the estimator is deduced by standard expansion of the first-order condition within  $\mathcal{L}_s$ .

#### 4.1.4 Extension to the SSVARMA( $p, q$ ) case

If  $\Theta(L)$  is of order  $q > 1$ , one can go back to the previous case. For this, define  $\tilde{\epsilon}_t = [\epsilon_t', \dots, \epsilon_{t-q+1}']'$  and  $\tilde{Y}_t = [Y_t', \mathbf{0}_{1 \times (n-1)q}]'$ . With these notations, the model reads:

$$\tilde{Y}_t = \tilde{\Phi}_1 \tilde{Y}_{t-1} + \dots + \tilde{\Phi}_p \tilde{Y}_{t-p} + (I - \tilde{\Theta}L)\tilde{\epsilon}_t, \quad \text{with } \tilde{\Theta} = \begin{bmatrix} \Theta_1 & \Theta_2 & \dots & & \Theta_q \\ I & 0 & \dots & & \\ 0 & I & 0 & \dots & \\ \vdots & & \ddots & & \\ 0 & \dots & 0 & I & 0 \end{bmatrix}, \quad (4.18)$$

and with  $\tilde{\Phi}_k = [\mathbf{u}\mathbf{u}'] \otimes \Phi_k$ , where  $\mathbf{u}$  is a  $q$ -dimensional vector equal to  $[1, 0, \dots, 0]'$ . The eigenvalues of  $\tilde{\Theta}$  are the roots of  $\det \Theta(z)$ .

Although the representation (4.18) makes it possible to get approximations of the  $\varepsilon_t$ 's, it does not imply that  $(\tilde{Y}_t)$  is a VARMA( $p, 1$ ) process because the process  $(\tilde{\varepsilon}_t)$  is not a white noise.

## 4.2 Semi-parametric estimation of a non-fundamental SSVARMA model

Let us consider a SSVARMA( $p, 1$ ) model:

$$Y_t = \mu + \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + C_0 \eta_t + C_1 \eta_{t-1}, \quad (4.19)$$

satisfying Assumptions A.1–A.7. For the sake of notational simplicity, we have replaced  $C$  by  $C_0$  and  $-\Theta_1 C$  by  $C_1$ . We assume that the roots of the determinant of the autoregressive polynomial are outside the unit circle, but the roots of the determinant of the moving average polynomial may be inside or outside the unit circle. We denote by  $f_j$  the common probability density function of the independent  $\eta_{j,t}$ 's,  $t = 1, \dots, T$ . We have to consistently estimate the (true values of the) parameters  $\mu, \Phi_1, \dots, \Phi_p, C_0, C_1$  as well as the (true) functional parameters  $f_j, j = 1, \dots, n$ . We consider below a 2-step moment method.

The first step consists in 2-Stage Least Squares (2SLS) regressions using  $Y_{t-2}, \dots, Y_{t-1-k}$  as instruments ( $k \geq p$ ), exploiting the fact that the latter are independent of  $Z_t = C_0 \eta_t + C_1 \eta_{t-1}$ . We denote by  $\hat{\mu}, \hat{\Phi}_1, \dots, \hat{\Phi}_p$  the corresponding 2SLS parameter estimates (see Subsection 5.1 of the Online Appendix for more details about this 2SLS estimator).

Once  $\mu$  and  $\Phi$  are estimated, the associated residuals

$$\hat{Z}_t \equiv Y_t - \hat{\mu} - \hat{\Phi}_1 Y_{t-1} - \dots - \hat{\Phi}_p Y_{t-p}, \quad (4.20)$$

are consistent approximations of  $Z_t = C_0 \eta_t + C_1 \eta_{t-1}$ . Then, in a second step, we proceed with the estimation of  $C_0$  and  $C_1$  in a pure moving average framework.<sup>25</sup>

### 4.2.1 Moment restrictions for $C_0, C_1$

Let us first consider the estimation of the moving average parameters as if the true  $Z_t$ 's were observed (this assumption is relaxed in the next subsection). To address the lack of second-order identifiability of  $C_0, C_1$ , the econometric literature has proposed to exploit moments of order 3 and/or 4 (see e.g. Bonhomme and Robin, 2009, in the special case  $C_1 = 0$ , Gospodinov and Ng,

<sup>25</sup>If  $C_1 = 0$ ,  $C_0$  can be directly estimated by ICA (see e.g. Chen, Choi, and Escanciano, 2017; Gouriéroux, Monfort, and Renne, 2017; Lanne, Meitz, and Saikkonen, 2017).

2015, in the one-dimensional case, [Lobato and Velasco, 2018](#), in the one-dimensional case through bi-spectral and tri-spectral density functions).

Loosely speaking, it is appropriate to consider (cross) moments of order 3 if some components  $\eta_{j,t}$  of  $\eta_t$  are skewed (i.e. with non-zero third-order cumulants  $\kappa_{3,j}$ ) and it is appropriate to use (cross) moments of order 4 if  $\eta_t$  features kurtotic components (with non-zero fourth-order cumulants  $\kappa_{4,j}$ ). The description of the moment conditions associated with cumulants is provided in [Appendix C](#).<sup>26</sup>

This approach leads to a set of moment restrictions:

$$E [h(Z_t, Z_{t-1}; \beta)] = 0, \quad (4.21)$$

where

$$\beta = [\text{vec} C_0', \text{vec} C_1', \kappa_{31}, \dots, \kappa_{3n}, \kappa_{41}, \dots, \kappa_{4n}]'. \quad (4.22)$$

The calibrated moments concern linear combinations of  $Z_t$  and  $Z_{t-1}$  such as:

$$E ([u'Z_t + v'Z_{t-1}]^2), \quad E ([u'Z_t + v'Z_{t-1}]^3), \quad \text{or} \quad E ([u'Z_t + v'Z_{t-1}]^4), \quad (4.23)$$

for different pairs  $(u, v)$ . We get an infinite set of moment restrictions but, in practice, a finite set of pairs has to be selected (see [Section 5](#)). Denoting by  $r$  the dimension of  $h(Z_t, Z_{t-1}; \beta)$ , the order condition is  $r \geq 2n^2 + 2n$ . The rank condition is challenging to analyse in the multivariate case.<sup>27</sup> In practice, difficulties in inverting the asymptotic covariance matrix of the estimator – derived in [Subsection 5.2](#) of the [Online Appendix](#) – constitute a signal of non-identification.

In the spirit of the approach proposed by [Lobato and Velasco \(2018\)](#) who work in the frequency domain, additional types of moment restrictions might be introduced:<sup>28</sup>

- i) We might consider similar moment restrictions based on a different lag order, i.e.  $E([u'Z_t + v'Z_{t-k}]^2)$ ,  $E([u'Z_t + v'Z_{t-k}]^3)$  or  $E([u'Z_t + v'Z_{t-k}]^4)$  for some  $k > 1$ .
- ii) We might also consider quantities such as:  $E([u'Z_t + v'Z_{t-1} + w'Z_{t-2}]^2)$ ,  $E([u'Z_t + v'Z_{t-1} + w'Z_{t-2}]^3)$  or  $E([u'Z_t + v'Z_{t-1} + w'Z_{t-2}]^4)$ .

These additional moment restrictions are likely to bring additional information when  $q > 1$ .

<sup>26</sup>See [Boudt, Cornilly, and Verdonck \(2018\)](#) or [Guay and Normandin \(2018\)](#) for a counting of up-to-order-4 comoments in the multivariate static case.

<sup>27</sup>Existing results are only for  $\text{ARMA}(p, q)$  processes ([Gospodinov and Ng, 2015](#), Lemma 3, and [Lobato and Velasco, 2018](#), Theorem 1). In [Gospodinov and Ng \(2015\)](#), the conditions pertain to specific cumulants of order  $p + q$ ; in [Lobato and Velasco \(2018\)](#), the conditions are on multi-dimensional integrals of spectral densities. [Guay and Normandin \(2018\)](#) have recently derived rank conditions in the context of moment-based estimations of SVAR models.

<sup>28</sup>The approach of [Lobato and Velasco \(2018\)](#) based on bi-spectral and tri-spectral densities seems difficult to extend in the multivariate framework.

### 4.2.2 2-step moment method

The first step of our approach, namely the 2SLS estimation, provides us with an estimator of  $\alpha = [\mu', \text{vec}(\Phi)']'$ . The second step of the approach involves the sample counterparts of (4.21), after replacement of  $Z_t$  by  $\hat{Z}_t = Y_t - \hat{\mu} - \hat{\Phi}_1 Y_{t-1} - \dots - \hat{\Phi}_p Y_{t-p}$ . In others words, the moment restrictions (4.21) become:

$$E[h(z_t(\alpha), z_{t-1}(\alpha); \beta)] = 0, \quad (4.24)$$

with  $z_t(\alpha) = y_t - \mu - \Phi_1 y_{t-1} - \dots - \Phi_p y_{t-p}$ . In the first step,  $\hat{\alpha}_T$  is estimated by IV (see 4.2.1); then  $\hat{\beta}_T$  is estimated by minimization of a moment-based criterion made precise in Subsection 5.2 of the Online Appendix (exploiting Appendix C). This minimization implies that

$$\frac{1}{T} \sum_{t=1}^T h(z_t(\hat{\alpha}_T), z_{t-1}(\hat{\alpha}_T); \hat{\beta}_T) \approx 0.$$

Subsection 5.2 of the Online Appendix derives the asymptotic accuracy of this 2-step moment estimator.

### 4.2.3 Nonparametric estimation of the error distribution

Once  $\mu$ ,  $\Phi_1, \dots, \Phi_p$ ,  $C_0$  and  $C_1$  have been estimated, consistent approximations of the errors  $\eta_t$  are based on the truncated estimates of the  $\varepsilon_t^*$ 's, obtained by recursively applying equations (4.11) and (4.12) of Subsection 4.1.2 (replacing  $\varepsilon_0^{*(2)}$  and  $\varepsilon_T^{*(1)}$  by zero vectors). The p.d.f.  $f_j$  can then be estimated by a kernel density estimator applied to the residuals  $\hat{\eta}_{j,t}$ ,  $t = 1, \dots, T$ .

## 5 Applications

### 5.1 Monte Carlo exercises

This subsection illustrates the finite-sample performances of the estimation approaches by means of Monte-Carlo experiments. Subsections 5.1.1 and 5.1.2 respectively deal with MA(1) and VMA(1) processes.

#### 5.1.1 Univariate case

In this subsection, we focus on MA(1) processes:

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1}, \quad (4.25)$$

where the  $\varepsilon_t$ 's are serially independent,  $E(\varepsilon_t) = 0$  and  $V(\varepsilon_t) = 1$ .

We consider different sample sizes ( $T = 100, 200$  and  $500$ ) and different true distributions of the errors  $\varepsilon_t$ . Four distributions are used: the Gaussian distribution, a mixture of Gaussian distributions and two Student distributions with respective degrees of freedom of 5 and 10. In all simulations, we use  $\theta = -2$ . Hence, data generating processes are non-fundamental.

In order to get the intuition behind our approach, it is instructive to look at the joint distribution of  $y_t$  and  $y_{t-1}$ . Figure 1 displays contour plots associated with these distributions in the context of the four different distributions used for  $\varepsilon_t$ . While the black solid lines correspond to the non-fundamental case, the grey lines represent the (pseudo) distribution that would prevail under the fundamental case, i.e. with  $\theta = -1/2$  and  $V(\varepsilon_t) = 4$ . In the purely Gaussian case (Panel (a)), the two distributions coincide, since the two processes are observationally equivalent. By contrast, in the other three cases –Panels (b), (c) and (d)– the two distributions are different. The case of the mixture of Gaussian distributions is particularly illustrative. For this distribution, and in the non-fundamental case, the shock  $\varepsilon_t$  is drawn from  $\mathcal{N}(\mu_1, \sigma_1^2)$  with probability  $p$  and from  $\mathcal{N}(\mu_2, \sigma_2^2)$  with probability  $1 - p$ . We set:  $\mu_1 = -0.7$ ,  $\mu_2 = 0.7$ ,  $\sigma_1 = 0.32$ ,  $\sigma_2 = 0.95$ ,  $p = 0.5$ , which results in a zero-mean unit-variance distribution with order-3 and order-4 cumulants of 0.85 and 0, respectively.<sup>29</sup>

For the ML approach (Subsections 4.1), and for each of the four distributions, we estimate five parameters:  $\theta$ , the variance of  $\varepsilon_t$  and three parameters specifying a mixture of Gaussian distributions with zero mean and unit variance. As far as the Student distributions are concerned, we proceed under the assumption that we do not know the true distribution, which is generally the case in practice. Hence, when the true distributions are Student, the ML approach is, more precisely, a Pseudo Maximum Likelihood (PML) approach.<sup>30</sup> This exercise is in the spirit of [Lii and Rosenblatt \(1996\)](#) and [Wu and Davis \(2010\)](#), who study the performances of their ML estimators when using a misspecified distribution in the computation of the log-likelihood function.

In the context of our variant of the GMM approach (Subsection 4.2), we consider the order-2, order-3 and order-4 moments given in (4.23), with  $(u, v) \in \{(1, 0), (1, 2), (2, 1)\}$ .<sup>31</sup> Hence, we use 9 restrictions to estimate, for each model, the four following parameters:  $\theta$ , the variance of  $\varepsilon_t$ ,  $\kappa_3$  and  $\kappa_4$ .

In our discussion of the results, we focus on the estimates of  $\theta$ . Figure 2 shows the distributions of the estimators of  $\theta$  resulting from both approaches. Each of the four rows of plots corresponds to one of the four considered distributions for  $\varepsilon_t$ , that are those distributions represented on Figure 1.

<sup>29</sup>In their Monte-Carlo experiment, [Gospodinov and Ng \(2015\)](#) also consider a distribution characterized by  $\kappa_3 = 0.85$  and  $\kappa_4 = 0$ .

<sup>30</sup>When the true distribution is Gaussian or a Gaussian mixture, the approach is not a PML, but a standard ML approach. Note that the Gaussian distribution is a special case of Gaussian mixture.

<sup>31</sup>We do not use  $(u, v) = (1, 1)$  because, for these values of  $u$  and  $v$ , the moments of  $uy_t + vy_{t-1}$  are the same as those of  $vy_t + uy_{t-1}$ , thereby preventing the identification of the parameters. Indeed, if  $y_t = \varepsilon_t - \theta\varepsilon_{t-1}$  and  $y_t^* = -\theta\varepsilon_t + \varepsilon_{t-1}$ , then  $y_t + y_{t-1}$  and  $y_t^* + y_{t-1}^*$  have the same distribution, and therefore the same moments.

On each panel, the three curves correspond to the three considered sample sizes ( $T = 100, 200$  and  $500$ ). The (finite sample) distributions are often bimodal; one mode being close to the true value of  $\theta$  –indicated by a vertical bar on each panel– and the other being close to  $1/\theta$ .

The results illustrate the fact that, the closer the distribution of the generated shocks is to a normal one, the weaker the identification. Let us focus for instance on the fraction of estimated fundamental processes, i.e. with  $\theta$  estimates that are lower than 1 in absolute value (whereas the data generating process is non-fundamental since  $|\theta| > 1$ ). In the Gaussian case (Panels a.1 and a.2 of Figure 2), about half of the estimated processes are fundamental, irrespective of the sample size or of the estimation approach. Among the three remaining distributions, the one leading to the largest fractions of estimated fundamental processes is the Student  $t(10)$ . For the latter distribution and in the ML case, even for a relatively large sample size ( $T = 200$ ), about a third of the estimated  $\theta$ 's have a modulus lower than 1, whereas this proportion is of about 12% in the Student- $t(5)$  case and close to 0% in the Gaussian mixture case. Identification is easier in the Student  $t(5)$  and in the Gaussian mixture cases (Panels b.1, b.2, c.1 and c.2) since, in these cases, the differences between the distributions of  $(y_{t-1}, y_t)$  in the fundamental versus non-fundamental regimes are more marked than for the other two distributions, as illustrated by Figure 1.

Table 1 reports summary statistics associated with the different estimators. The results show that our GMM approach is less efficient than the (P)ML one. Indeed, Root Mean Squared Errors (RMSEs) are lower with the (P)ML approach. This is the case even when the (P)ML uses a misspecified distribution, i.e. when the true distributions are Student- $t$ . These lower RMSEs are accounted for by far smaller standard deviations of the estimator distributions for the (P)ML approach, which more than compensates the fact that (P)ML-estimator biases are often larger than in the GMM case.

To the best of our knowledge, this study is the first one to compare GMM and ML estimates in the context of non-fundamental univariate MA processes. As regards the GMM approach, our results are comparable to the results in the literature, such as the ones reported by [Gospodinov and Ng \(2015\)](#) for  $T = 500$ , or by [Lobato and Velasco \(2018\)](#) with spectral-density-based estimator for  $T = 100, 200$  and Student  $t(5)$  distribution.<sup>32</sup>

The last two columns of Table 1 aim at assessing the validity of the asymptotic distribution of the  $\theta$  estimators. These columns indicate the fractions of times (among the  $N$  simulations) where the true value of  $\theta$  lies within the interval  $[\hat{\theta} - \phi_\alpha \sigma_{asy}, \hat{\theta} + \phi_\alpha \sigma_{asy}]$ , where  $\sigma_{asy}$  denotes the estimate of the asymptotic standard deviation of the estimator  $\hat{\theta}$  and where  $\phi_\alpha$  is such that

<sup>32</sup>Our GMM estimator however appears to be less efficient than the spectral-density-based estimator of [Lobato and Velasco \(2018\)](#): For  $T = 200$  and when the true distribution is  $t(5)$ , [Lobato and Velasco \(2018\)](#) report that the absolute value of their  $\theta$  estimator is on the right side of 1 in 93% of the cases (last column in their Table 1), while we get an equivalent percentage of 71% with our GMM estimator. This percentage is of 88% for our PML estimator. Our PML RMSEs are close to those reported by [Lobato and Velasco \(2018\)](#) in this case ( $T = 200$ ,  $t(5)$  distribution,  $\theta = -2$ ).

$P(-\phi_\alpha < X < \phi_\alpha) = \alpha$ , if  $X \sim \mathcal{N}(0, 1)$ . The closer to  $\alpha$  the reported fractions, the better the asymptotic approximation of the estimator distribution. For each of the two considered values of  $\alpha$ , two ratios are reported: while the first one is based on all simulated sample paths, the second one (in italics) excludes those paths where the wrong fundamentalness regime is estimated. Indeed, when the wrong fundamentalness regime is estimated, it is extremely likely that the true value of  $\theta$  is not in the estimated confidence interval; this is because  $\sigma_{asy}$  does not integrate the fact that the wrong fundamentalness regime may have been estimated. As expected, the results indicate that the adequacy between the ratios and the  $\alpha$ 's is better in the second case, that is when the right fundamentalness regime is estimated (the proportion of the latter cases is reported in the sixth column of Table 1). Let us stress that while this second case points to the validity of the asymptotic approximation of the estimator distribution, it is not consistent with practical situations, where the right fundamentalness regime is usually unknown.

### 5.1.2 Bivariate case

Let us now consider the following bivariate VMA(1) process:

$$Y_t = C\eta_t - \Theta C\eta_{t-1}, \quad (4.26)$$

with

$$C = \begin{bmatrix} 0 & 1 \\ 1 & 0.5 \end{bmatrix} \quad \text{and} \quad \Theta = \begin{bmatrix} -0.5 & 0 \\ 1 & -2 \end{bmatrix}. \quad (4.27)$$

Both components of  $\eta_t$  are independently drawn. The first structural shock,  $\eta_{1,t}$ , is drawn from a mixture of Gaussian distributions featuring a skewness of 2 and an excess kurtosis of 6.<sup>33</sup> The second structural shock, that is  $\eta_{2,t}$ , is drawn from a Student's distribution with 6 degrees of freedom, i.e. with a skewness of 0 and an excess kurtosis of 3.

For each simulated path of  $Y_t$ , we implement both the GMM and the ML approaches. As regards the ML approach, we proceed as if we did not know that the true distribution of  $\eta_{2,t}$  was a Student's distribution and we use mixtures of Gaussian distributions for both  $\eta_{1,t}$  and  $\eta_{2,t}$ . Strictly speaking, this approach is therefore a PML one. Whereas twelve parameters are estimated in the GMM case, fourteen are involved in the ML case: both cases involve the estimation of the components of  $C$  and  $\Theta$  (8 parameters); the remaining parameters to be estimated are  $\kappa_{3,i}$  and  $\kappa_{4,i}$ ,  $i \in \{1, 2\}$  in the GMM case (where  $i = 1$  for the first structural shock and  $i = 2$  for the second one), and  $p_i$ ,  $\mu_{1,i}$  and  $\sigma_{1,i}$ ,  $i \in \{1, 2\}$  in the PML case (constrained by the zero mean and unit variance conditions). For the GMM approach, we consider the order-2, order-3 and order-4 moments given

<sup>33</sup> Specifically, thus first shock is drawn from  $\mathcal{N}(\mu_1, \sigma_1^2)$  with probability  $p$  and from  $\mathcal{N}(\mu_2, \sigma_2^2)$  with probability  $1 - p$ , where  $\mu_1 = 2.12$ ,  $\mu_2 = -0.24$ ,  $\sigma_1 = 1.41$ ,  $\sigma_2 = 0.58$ ,  $p = 0.1$ .

in (4.23) with:

$$(u_1, u_2, v_1, v_2) \in \{(2, 0, 0, 0), (0, 2, 0, 0), (1, 0, 2, 0), (2, 0, 1, 0), (1, 0, 0, 2), (2, 0, 0, 1), (0, 1, 0, 2), (0, 2, 0, 1), (0, 1, 2, 0), (0, 2, 1, 0)\}, \quad (4.28)$$

which results in 30 restrictions.<sup>34</sup>

Figure 3 shows the densities of the estimators of the  $\Theta_{i,j}$ 's. As in the univariate case, some distributions are bi-modal, in particular for the shorter samples and for  $\Theta_{1,1}$  and  $\Theta_{2,2}$ . These densities also show that the PML estimates are on average more precise than the GMM ones. Table 2 documents the latter point: its third column indicates in particular that the RMSEs are lower in the PML case. The sixth column reports the fraction of replications for which the “right fundamentalness regime” is estimated. Here, we consider that an estimated model features the “right fundamentalness regime” if the estimate of  $\Theta$  has one eigenvalue on each side of the unit circle and if the absolute values of the estimates of  $\Theta_{1,1}$  and  $\Theta_{2,2}$  are respectively below and above one, as the true values.<sup>35</sup> It appears that the PML estimations result in the right fundamentalness regime in more than 90% of the cases as soon as  $T$  is larger than 200. The GMM approach is less successful, with 67% of correct regime identification for  $T = 500$ .

## 5.2 Bivariate real-data example: GNP growth and unemployment

For comparison with the literature, we consider the two-variable model of Blanchard and Quah (1989), referred to as BQ hereafter. The two stationarized endogenous variables are the U.S. real GNP growth and the (detrended) unemployment rate. BQ fit an 8-lag VAR model to these data for the period from 1948Q2 to 1987Q4 (159 dates) and impose long-run restrictions to identify demand and supply shocks. Specifically, they impose that the demand shock has no long-run impact on real GNP. That is, in their model, the contribution of supply disturbances to the variance of output tends to unity as the horizon increases.

Using the same dataset and analyzing the location of the (complex roots) of the 8-lag VAR of BQ, Lippi and Reichlin (1994)'s results suggest that this VAR approximates a VARMA(1,1) model. Further, Lippi and Reichlin (1994) explore the influence of inverting the roots of the moving average polynomial of this VARMA(1,1) model on the IRFs. They illustrate that fundamental and non-fundamental versions of the model have different implications, notably in terms of first impacts of the shocks and of variance decompositions. However, their analysis does not allow them to statistically pinpoint the most suitable model among the different versions they obtain (the

<sup>34</sup>See Footnote 31 for the reason why the  $u_i$ 's and  $v_i$ 's are not in  $\{0, 1\}$  only.

<sup>35</sup>Because (the true value of)  $\Theta$  is triangular, its eigenvalues coincide with its diagonal elements. It is not the case for its estimate.

non-fundamental ones and the fundamental one). As explained in Section 3.1, if the underlying structural shocks are non-Gaussian and independent, the data-generating SSVARMA process, be it fundamental or not, is identifiable. Preliminary indications regarding the non-normality of the system can be obtained by applying normality tests on the endogenous variable themselves: the Lütkepohl and Theilen (1991)'s multivariate tests, based on standardized third and fourth moments of possibly correlated variables, point to the non-Gaussianity of the vectors of endogenous variables.<sup>36</sup>

We employ both the ML and the 2SLS-GMM approaches described in Subsections 4.1 and 4.2 to fit VARMA( $p,1$ ) models to BQ's dataset. As in Blanchard and Quah (1989) and Lippi and Reichlin (1993, 1994), the two endogenous variables follow a stationary process and are not cointegrated.<sup>37</sup> Given the relative shortness of the sample, we expect the 2SLS-GMM estimates to be quite imprecise (see Subsection 5.1.2). However, an advantage of this method is that its numerical complexity does not depend on the order  $p$  of  $\Phi(L)$  (because the estimate of  $\Phi(L)$  results from linear least-square regressions). We therefore proceed as follows: for  $p \in \{1, \dots, 6\}$ , we apply the 2SLS-GMM approach and we then use the resulting estimates as starting values for the numerical optimization of the truncated log-likelihood function.<sup>38</sup>

In the context of the ML approach, the distributions of the independent shocks  $\eta_{j,t}$ , for  $j \in \{1, 2\}$ , are assumed to be Gaussian mixtures.<sup>39</sup> Specifically, we assume that  $\eta_{j,t}$  is drawn from  $\mathcal{N}(\mu_{j,1}, \sigma_{j,1}^2)$  with a probability  $p_j$  and from  $\mathcal{N}(\mu_{j,2}, \sigma_{j,2}^2)$  with a probability  $1 - p_j$ . Therefore, if we have  $E(\eta_{j,t}) = 0$  and  $V(\eta_{j,t}) = 1$ , the distribution of the two-dimensional vector  $\eta_t$  is completely defined by  $\gamma = [\mu_{1,1}, \mu_{2,1}, \sigma_{1,1}, \sigma_{2,1}, p_1, p_2]'$ .<sup>40</sup>

Our preferred model features four lags ( $p = 4$ ).<sup>41</sup> Parameter estimates resulting from both the 2SLS-GMM and the ML approach are reported in Table 3. For both approaches, the roots of

<sup>36</sup>The p-value of this test, whose null hypothesis is that the considered variables are Gaussian, is of 3.5%.

<sup>37</sup>As noted by Lippi and Reichlin (1994), Footnote 1 p. 310, assuming that the dynamics of the two endogenous variables is described by a stationary VAR model (as done in Blanchard and Quah, 1989; Lippi and Reichlin, 1993, 1994) amounts to precluding cointegration.

<sup>38</sup>In the 2SLS step of the 2SLS-GMM approach, we use  $k = p + 3$  (i.e. we regress  $Y_{t-1}$  on  $Y_{t-2}, \dots, Y_{t-p-4}$ ), see Subsection 5.1 of the Online Appendix. Then, in the second (GMM) step, we consider the order-2, order-3 and order-4 moments given in (4.23) for the  $(u, v)$  pairs given in (4.28).

<sup>39</sup>Because the true distribution of the shocks is not known in practice, the choice of a flexible parametric type of distributions is important. Though relatively parsimonious, Gaussian mixtures accommodate various interesting features. In particular, they can attain any possible (*skewness*, *kurtosis*) pair in the whole domain ( $kurtosis \geq skewness^2 + 1$ ) and can feature bi-modality.

<sup>40</sup>For the numerical optimization of the truncated log-likelihood function, starting values for the elements of  $\gamma$  are obtained as follows: after the 2SLS-GMM estimation, we apply the backward/forward filtering algorithm underlying function  $a_t$  (see Subsection 4.1.2) to obtain estimates of the structural shocks  $\eta_{j,t}$ . Then, for  $j \in \{1, 2\}$ , we estimate parameters  $(\mu_{1,i}, \sigma_{1,i}, p_i)$  by MLE, which is very fast.

<sup>41</sup>The Schwarz (1978), Hannan and Quinn (1979) and Akaike (1974) and criteria respectively call for  $p = 2$ ,  $p = 3$  and  $p = 5$ . However, contrary to the  $p = 4$  case, the Box and Pierce (1970)'s and Ljung and Box (1978)'s tests reject the null hypothesis of no auto-correlation in the estimated series of  $\eta_{j,t}$  when  $p$  is equal to 2, 3 or 5.

$\det(\Phi(z))$  are outside the unit circle; the estimated processes are therefore stationary. Besides, for both estimation methods, the two roots of  $\det(I - \Theta z)$  are on each side of the unit circle (1.77 and 0.41 for the ML approach, 2.34 and  $-0.03$  for the GMM approach). Hence, the estimated processes are non-fundamental. Figure 4 shows the ML estimated parametric distributions of the structural shocks (Gaussian mixtures, represented by dotted lines). These estimated distributions are fairly close to kernel-based densities associated with the  $\eta_{j,t}$ 's estimates derived from the backward/forward filtering algorithm underlying function  $a_t$ , introduced in Subsection 4.1.2.

It is difficult to check ex-ante the relevance of the assumptions on the structural shocks, in particular regarding the independence assumption A.1. Tests can however be applied to the estimated structural shocks. In the present case, the application of Box and Pierce (1970)'s and Ljung and Box (1978)'s tests on the estimated series of structural shocks do not detect auto-correlation in the estimated series of  $\eta_{j,t}$ ,  $|\eta_{j,t}|$ , nor in the  $\eta_{j,t}^2$ , which is a necessary condition for independence. The normality hypothesis of the  $\eta_{j,t}$ 's is rejected by the Shapiro and Wilk (1965)'s test, with p-values below 2%.<sup>42</sup> A limitation of these residual tests is however that they do not take into account the uncertainty regarding the estimation of the  $\eta_{j,t}$ 's; this is left for future research.

Figure 5 displays the impulse response functions resulting from the ML approach and compares them with those obtained with long-run restrictions *à la* Blanchard and Quah (1989). Because the ML approach does not rely on restrictions based on economic theory, the resulting structural shocks have no a priori economic interpretation. However, for comparing the different approaches, we bring the BQ supply shock closer to the ML-estimated shocks that accounts for the largest part of the GNP long-run variance. We call this shock the “long-run shock”.

Figure 5 shows that the BQ and ML response functions present similarities. Nevertheless, in the ML approach, both structural shocks –Shock 1 (first column of charts) and Shock 2 (second column of charts)– have a long-run impact on GNP. This is not the case with the BQ approach where, by construction, the effect of Shock 1 on GNP converges to zero. The fact that none of the two structural shocks identified within a bivariate VAR has a zero impact on output in the long-run was also obtained by Cochrane (1994) for instance.<sup>43</sup> The responses of unemployment to Shock 1 (bottom-left plot) and of GNP to Shock 2 (upper-right plot) broadly feature the same shapes for both approaches, but do not agree on the delay before maximal impact. The bottom-right plot shows that, according to the ML approach, the instantaneous response of unemployment to the second (long-run) shock is close to zero, reflecting the fact that the estimated coefficient  $C_{2,2}$  is small and not statistically different from zero (see Table 3). Obtaining a  $C$  coefficient equal to zero

<sup>42</sup>Because no auto-correlation is detected in the residual, there is no need for using autocorrelation-robust normality tests such as those of Bai and Ng (2005) and of Lobato and Velasco (2004).

<sup>43</sup>The bivariate model of Cochrane (1994) describes the joint dynamics of GNP and consumption growth. The structural shocks are identified by means of a short-run restriction (consumption does not respond contemporaneously to a GNP shock).

is reminiscent of what could be obtained in the context of an identification strategy *à la* Sims – imposing that one structural shock has no contemporaneous impact on the unemployment rate. However, the results presented here could not have been obtained in this context since the standard short-run restriction methodology (*à la* Sims) is based on VAR models and therefore necessarily leads to fundamental models. By contrast, our approach remains a priori agnostic with respect to the fundamentalness regime.

## 6 Concluding remarks

We have shown in this paper that the static and dynamic identification problems associated with Gaussian structural VARMA processes disappear if (i) the structural shocks are non-Gaussian and if (ii) these shocks are serially and instantaneously independent. In other words, non-Gaussian strong structural VARMA (SSVARMA) models are identified. We further propose parametric and semi-parametric estimation methods able to consistently estimate possibly non-fundamental representation in the moving average dynamics. Our analysis also highlights the fact that, in the context of non-Gaussian SSVARMA, the impulse response functions are defined in a non-ambiguous way.

Conventional econometric toolboxes available to macroeconomists have often been conceived for linear forecasting purposes and for SVARMA models with shocks that are uncorrelated, but not independent. Because it focuses on second-order properties only, the standard SVARMA literature and associated toolboxes often introduce potentially over-identifying assumptions that may entail misspecification and misleading interpretation of SVARMA residuals. As shown in the parametric and semi-parametric analysis developed in Section 4 and in the applications presented in Section 5, such potential pitfalls can be addressed in the non-Gaussian case, provided that the independence assumptions are valid and that the appropriate estimation methods are used.

The methods developed in this paper might be extended in several directions, which are left for future research. First testing procedures, in particular tests of fundamentalness, may be obtained. Second the properties of estimation methods can be analyzed in a neighbourhood of the Gaussian assumption (see e.g. [Gouriéroux and Jasiak, 2016](#)), or in a neighbourhood of unit roots (see e.g. [Chen, Davis, and Song, 2011](#), for non-invertible MA process). Third, the asymptotic properties of our ML estimate in the presence of cointegration may be derived. Fourth, the identification and estimation results might be extended to the case of more errors than observables. Indeed, while identification results exist when the errors are not Gaussian (see e.g. Th 3.1. in [Eriksson and Koivunen, 2004](#), for the static case, [Gouriéroux and Zakoian, 2015](#), for stable multivariate processes, or [Gagliardini and Gouriéroux, 2018](#), for a non-Gaussian factor model), the possibility to identify the dynamics when the number of shocks is larger than the number of endogenous variables and the errors are not Gaussian would be important in the discussion of the effect of

omitted variables (see e.g. [Giannone and Reichlin, 2006](#); [Lütkepohl, 2014](#)).

## A Proof of Proposition 1

Let us first recall Theorem 1 in [Chan, Ho, and Tong \(2006\)](#).

**Theorem.** *Let  $Y_t$  and  $Y_t^*$  be two non-Gaussian processes defined by:*

$$\begin{aligned} Y_t &= \sum_{k=-\infty}^{\infty} A_k \varepsilon_{t-k}, \\ Y_t^* &= \sum_{k=-\infty}^{\infty} A_k^* \varepsilon_{t-k}^*, \end{aligned}$$

where the processes  $\varepsilon_t$  and  $\varepsilon_t^*$  are strong white noises with independent components.

Then,  $Y_t$  and  $Y_t^*$  are observationally equivalent if and only if

$$\varepsilon_{j,t-m(j)}^* = \sigma_j \varepsilon_{\pi(j),t} \quad (\text{equality in distribution}), \quad (\text{a.1})$$

$$A_{k,j}^* = \frac{1}{\sigma_j} A_{k-m(j),\pi(j)}, \quad (\text{a.2})$$

where  $\pi$  is a permutation and  $A_{k,j}$  (respectively  $A_{k,j}^*$ ) is the  $j^{\text{th}}$  column of  $A_k$  (respectively  $A_k^*$ ) provided the following condition holds:

*the components of  $\varepsilon_t$  (resp.  $\varepsilon_t^*$ ) have non-zero  $r^{\text{th}}$  cumulant, with  $r \geq 3$  and a finite even moment of order  $s$  greater than, or equal to,  $r$ .*

If the moving averages are one-sided ( $A_k = A_k^* = 0, \forall k < 0, A_0 \neq 0, A_0^* \neq 0$ ) and  $\varepsilon_t$  (resp.  $\varepsilon_t^*$ ) is replaced by  $\eta_t$  (resp.  $\eta_t^*$ ), where the components of  $\eta_t$  (resp.  $\eta_t^*$ ) have a unit variance, this implies that  $m(j) = 0$  and  $\sigma_j = \pm 1$  for all  $j$ .

In our case, we have:

$$\begin{aligned} Y_t &= \Psi(L)C\eta_t, \\ Y_t^* &= \Psi^*(L)C^*\eta_t^*, \end{aligned}$$

with

$$\begin{aligned} \Psi(L) &= \Phi^{-1}(L)\Theta(L) = I + \Psi_1 L + \Psi_2 L^2 + \dots, \\ \Psi^*(L) &= \Phi^{-1}(L)\Theta^*(L) = I + \Psi_1^* L + \Psi_2^* L^2 + \dots \end{aligned}$$

Therefore, we have:

$$\begin{aligned} A_k &= \Psi_k C \quad \text{with } \Psi_0 = I, \\ A_k^* &= \Psi_k^* C^* \quad \text{with } \Psi_0^* = I. \end{aligned}$$

The previous theorem implies that the moving average matrix coefficients  $A_k$  are identified up to a permutation and a sign change of the columns. That is, there exists a permutation matrix  $P$  and a diagonal matrix  $D$ , whose diagonal elements are either  $-1$  or  $1$ , that are such that:

$$\Psi_k C = \Psi_k^* C^* P D, \quad \forall k.$$

For  $k = 0$ , this gives  $C = C^* P D$ , which further implies that  $\Psi_k = \Psi_k^*$  for all  $k$ . The  $\Psi_k$  are therefore identified and  $C$  is identified up to a permutation and a sign change of its columns. Since  $\Phi(L)$  and  $\Psi(L)$  are identified,  $\Theta(L) = \Phi(L)\Psi(L)$  is also identified.

## B Truncated Maximum Likelihood Approach

### B.1 The nonfundamentalness regimes for $n = 2$

When  $n = 2$ , the eigenvalues of  $\Theta$  are the solutions of the second-order polynomial equation:

$$\lambda^2 - Tr(\Theta)\lambda + det(\Theta) = 0.$$

Consider the case of roots with modulus equal to one. Three limiting cases have to be considered:

- i) One eigenvalue is equal to 1. This arises if  $1 - Tr(\Theta) + det(\Theta) = 0$ .
- ii) One eigenvalue is equal to  $-1$ . This arises if  $1 + Tr(\Theta) + det(\Theta) = 0$ .
- iii) Two complex eigenvalues have a modulus equal to 1. This arises if the second-order polynomial equation is of the form:  $\lambda^2 - 2\cos(a)\lambda + 1 = 0$ . This implies  $|Tr(\Theta)| < 2$  and  $det(\Theta) = 1$ .

The three frontiers above define 5 regions, as illustrated on Figure 6: Region I is purely fundamental; Regions III and V are purely nonfundamental; Regions II and IV are mixed. The dashed line represents the pairs  $(Tr(\Theta), det(\Theta))$  for which the discriminant  $Tr(\Theta)^2 - 4det(\Theta)$  is zero; the area that is below (respectively above) this parabola corresponds to real (resp. complex) eigenvalues for  $\Theta$ .

## B.2 Derivation of the Jacobian

Replacing the  $\varepsilon_t^{*(2)}$ 's appearing in equation (4.12) by their expressions given in (4.11), one obtains:

$$\underbrace{\begin{bmatrix} \varepsilon_0^{*(1)} \\ \vdots \\ \varepsilon_{T-1}^{*(1)} \\ \varepsilon_1^{*(2)} \\ \vdots \\ \varepsilon_T^{*(2)} \end{bmatrix}}_{=\varepsilon^*} = \underbrace{\begin{bmatrix} J_1 & J_{12} \\ 0 & J_2 \end{bmatrix}}_{=J} \underbrace{\begin{bmatrix} W_1^{(1)} \\ \vdots \\ W_T^{(1)} \\ W_1^{(2)} \\ \vdots \\ W_T^{(2)} \end{bmatrix}}_{=W} + \underbrace{\begin{bmatrix} \left\{ (U_\Theta^{(1)})^{-1} \right\}^T & M_0 \\ \vdots & \vdots \\ (U_\Theta^{(1)})^{-1} & M_{T-1} \\ \hline 0 & U_\Theta^{(2)} \\ \vdots & \vdots \\ 0 & (U_\Theta^{(2)})^T \end{bmatrix}}_{=M} \begin{bmatrix} \varepsilon_T^{*(1)} \\ \varepsilon_0^{*(2)} \end{bmatrix}, \quad (\text{a.3})$$

where  $J_1$  is an upper block triangular matrix with  $\left\{ -(U_\Theta^{(1)})^{-1} \right\}$  matrices on its diagonal, where  $J_2$  is lower block triangular with identity matrices on its diagonal, and where:

$$M_t = -(U_\Theta^{(1)})^{-1} U_\Theta^{(12)} (U_\Theta^{(2)})^t - \dots - \left\{ (U_\Theta^{(1)})^{-1} \right\}^{T-t} U_\Theta^{(12)} (U_\Theta^{(2)})^{T-1}.$$

Let's consider the vector  $V = [V_1', \dots, V_T']'$  where  $V_t = \Phi(L)Y_t$ . If  $Y = [Y_1', \dots, Y_T']'$ , we have  $V = DY + D_0Y^0$  where  $D$  is a lower triangular matrix whose diagonal is filled with ones (and therefore  $\det D = 1$ ), where  $D_0$  is matrix of dimension  $(nT) \times (np)$ , and where  $Y^0 = [Y_0', \dots, Y_{-p+1}']'$ .

By definition of  $W_t$ , we have  $A_\Theta W_t = \Phi(L)Y_t$  and, therefore,  $A_\Theta' V_t = W_t$ . Hence, with  $\tilde{W} = [W_1', W_2', \dots, W_T']'$ , we obtain:

$$\tilde{W} = (I \otimes A_\Theta') V = (I \otimes A_\Theta') (DY + D_0Y^0).$$

Let's denote by  $P$  the permutation that is such that  $W = P\tilde{W}$ , where  $W$  is defined in equation (a.3). Consider the case where  $\varepsilon_0^{*(2)}$  and  $\varepsilon_T^{*(1)}$ , thought of as state variables, are equal to zero. We obtain, using the notations of equation (a.3):

$$\varepsilon^* = JW = JP\tilde{W} = JP(I \otimes A_\Theta') (DY + D_0Y^0). \quad (\text{a.4})$$

Matrices  $A_\Theta$  and  $P$  being orthogonal, we have  $|\det A_\Theta| = |\det P| = 1$ . Since we also have  $\det D = 1$ , we get:

$$|\det(JP[I \otimes A_\Theta']D)| = |\det(J)| = \frac{1}{|\det(U_\Theta^{(1)})|^T}.$$

Denoting by  $\lambda_i(\Theta)$  the  $i^{\text{th}}$  eigenvalue of  $\Theta$ , the effect of the Jacobian on the truncated log-

likelihood is:

$$\log(|\det(U_{\Theta}^{(1)})|) = \sum_{i=1}^n \log |\lambda_i(\Theta)| \mathbb{1}_{|\lambda_i(\Theta)| \geq 1}. \quad (\text{a.5})$$

## C 2SLS-GMM: Cumulant-based identification

The first step provides consistent estimates of  $\mu$ ,  $\Phi_1, \dots, \Phi_p$  (see Subsection 5.1 of the Online Appendix). Therefore we can focus on the identification and the derivation of moments on the pure VMA process:  $Z_t = C_0 \eta_t + C_1 \eta_{t-1}$ .

The pairwise log-Laplace transform of  $(Z_t, Z_{t-1})$  is:

$$\begin{aligned} & \log E[\exp(u'Z_t + v'Z_{t-1})] \quad (\text{a.6}) \\ &= \log E[\exp(u'(C_0 \eta_t + C_1 \eta_{t-1}) + v'(C_0 \eta_{t-1} + C_1 \eta_{t-2}))] \\ &= \log (E[\exp(u'C_0 \eta_t)] \times E\{\exp[(u'C_1 + v'C_0) \eta_{t-1}]\} \times E[\exp(v'C_1 \eta_{t-2})]) \\ &= \sum_{j=1}^n \log E[\exp(u'C_{0j} \eta_{j,t})] + \sum_{j=1}^n \log E\{\exp[(u'C_{1j} + v'C_{0j}) \eta_{j,t-1}]\} + \sum_{j=1}^n \log E[\exp(v'C_{1j} \eta_{j,t-2})], \end{aligned}$$

using that the  $\eta_{j,t}$ 's are mutually and serially independent.

Let us denote by  $\Psi_{\xi}$  the log-Laplace transform of a given random variable  $\xi$  defined by:

$$\Psi_{\xi}(u) = \log E[\exp(w\xi)].$$

If the moments of  $\xi$  exist up to order 4, and if  $E(\xi) = 0$ , the log-Laplace transform can be expanded as:

$$\Psi_{\xi}(w) \approx \frac{w^2}{2} E(\xi^2) + \frac{w^3}{6} E(\xi^3) + \frac{w^4}{24} [E(\xi^4) - 3E(\xi^2)]. \quad (\text{a.7})$$

In particular, for the  $\eta_j$ 's, for which  $E(\eta_j) = 0$  and  $E(\eta_j^2) = 1$ , we have:

$$\Psi_{\eta_j}(w) \approx \frac{w^2}{2} + \frac{w^3}{6} \kappa_{3j} + \frac{w^4}{24} \kappa_{4j}, \quad (\text{a.8})$$

with  $\kappa_{3j} = E(\eta_{j,t}^3)$  and  $\kappa_{4j} = E(\eta_{j,t}^4) - 3$ .

Using (a.7) with  $\xi = u'Z_t + v'Z_{t-1}$  and (a.8), and taking the expansions of both sides of (a.6)

up to order 4, we get the following restrictions, holding for any pair  $(u, v)$ :

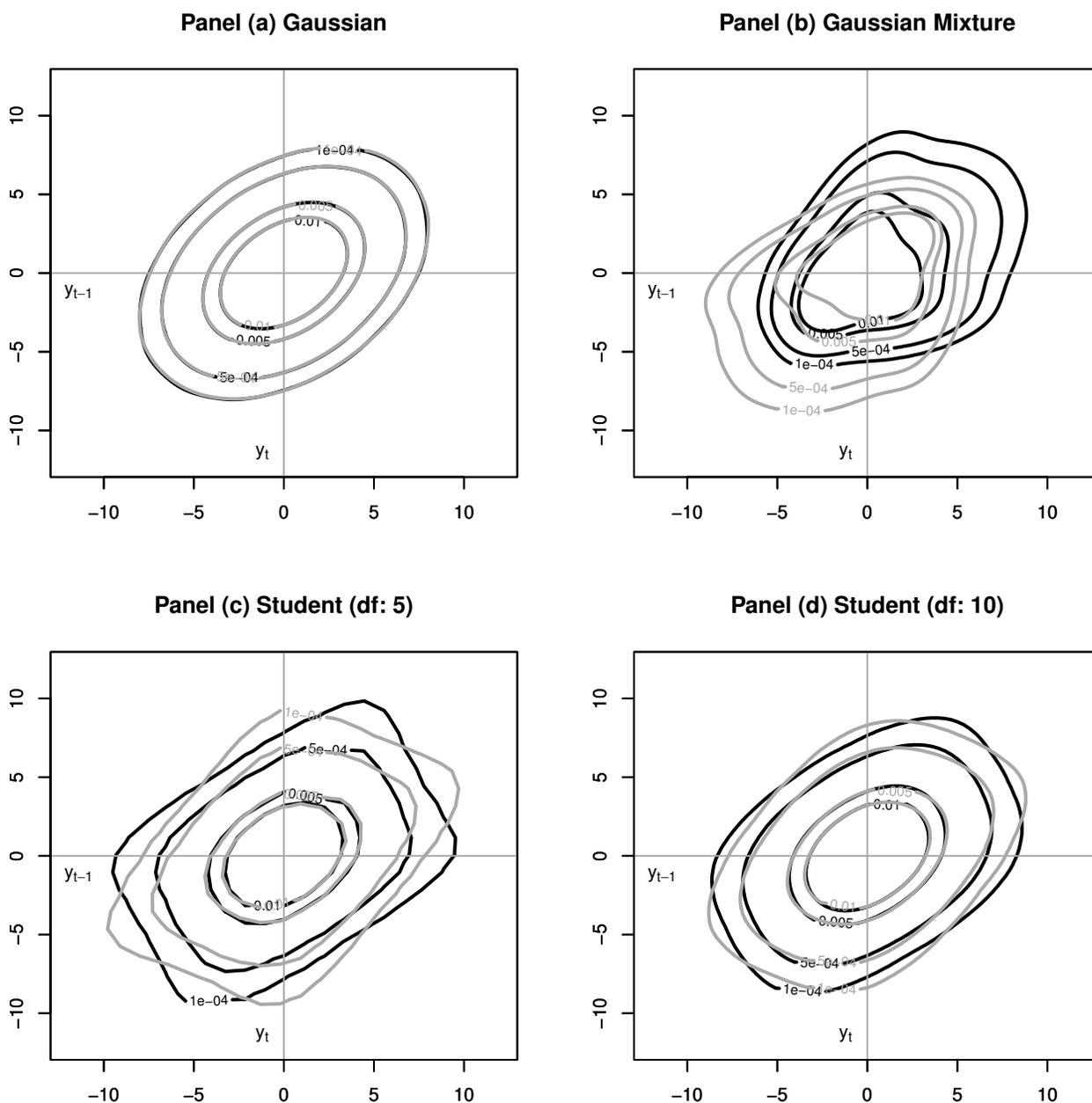
$$E[(u'Z_t + v'Z_{t-1})^2] = \sum_{j=1}^n [(u'C_{0j})^2 + (u'C_{1j} + v'C_{0j})^2 + (v'C_{1j})^2] \quad (\text{order 2})$$

$$E[(u'Z_t + v'Z_{t-1})^3] = \sum_{j=1}^n \kappa_{3j} [(u'C_{0j})^3 + (u'C_{1j} + v'C_{0j})^3 + (v'C_{1j})^3] \quad (\text{order 3})$$

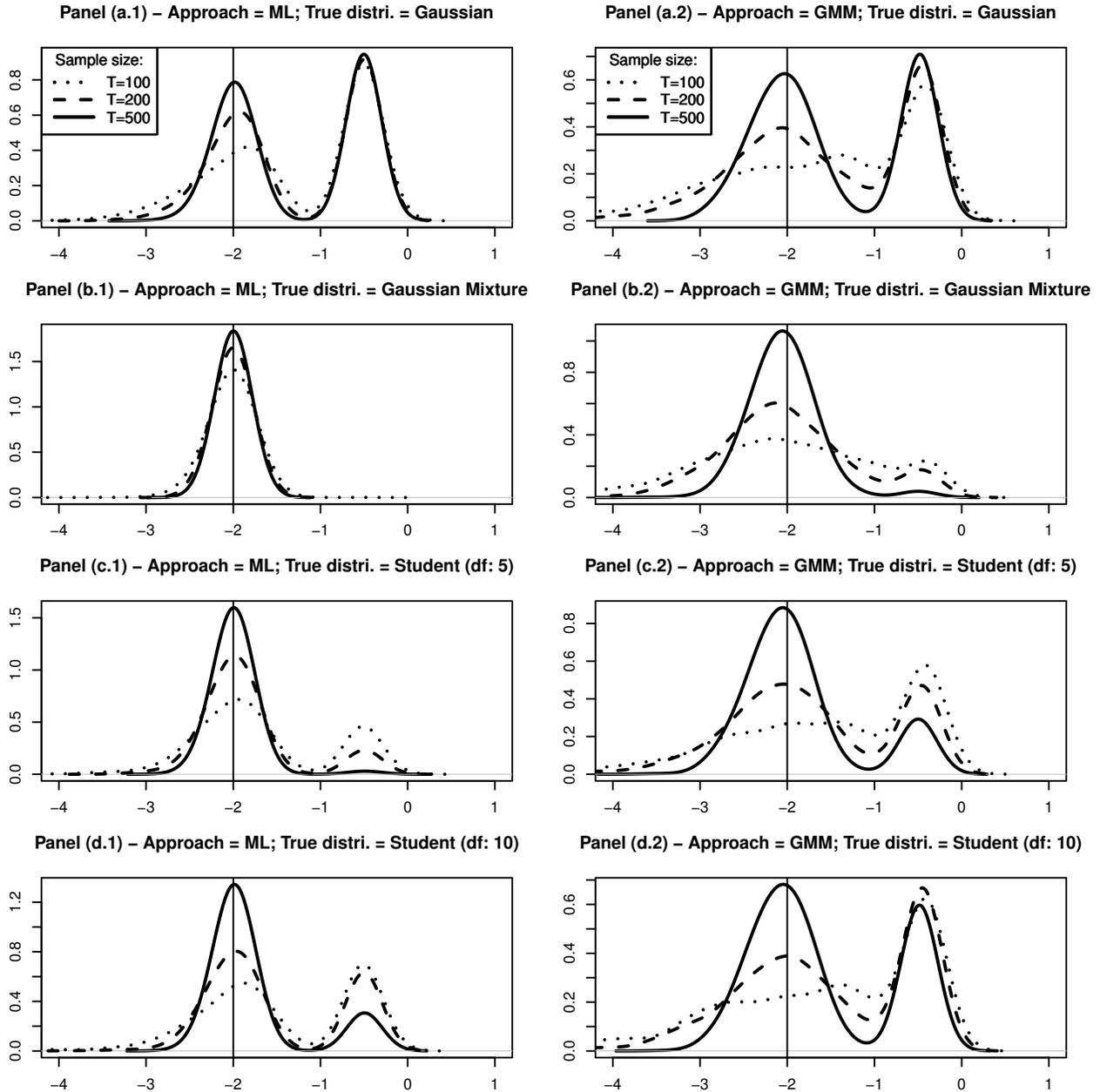
$$E[(u'Z_t + v'Z_{t-1})^4] = \sum_{j=1}^n \kappa_{4j} [(u'C_{0j})^4 + (u'C_{1j} + v'C_{0j})^4 + (v'C_{1j})^4] \\ + 3 \left( \sum_{j=1}^n [(u'C_{0j})^2 + (u'C_{1j} + v'C_{0j})^2 + (v'C_{1j})^2] \right)^2. \quad (\text{order 4})$$

At order 2, the system concerns the information contained in  $E(Z_t Z_t')$ ,  $E(Z_t Z_{t-1}')$ , which is not sufficient to identify  $C_0$  and  $C_1$ . The other equations provide additional information whenever appropriate higher-order cumulants, of order 3 and/or 4, are not zero.

Figure 1: Joint distributions of  $y_t$  and  $y_{t-1}$  in the fundamental and non-fundamental cases, for different distributions of the errors  $\varepsilon_t$



*Note:* Each of the four panels displays contour plots associated with the joint distributions of  $y_t$  and  $y_{t-1}$ , where  $y_t$  follows an MA(1) process:  $y_t = \varepsilon_t - \theta\varepsilon_{t-1}$ , where the  $\varepsilon_t$  are i.i.d.. Whereas the black lines correspond to the case  $\theta = -2$  and  $V(\varepsilon_t) = 1$  (non-fundamental process), the grey lines correspond to  $\theta = -1/2$  and  $V(\varepsilon_t) = \theta^2$  (fundamental process with same spectral density). The titles of the panels indicate the distribution types of the  $\varepsilon_t$ 's. For Panel b (mixture of Gaussian distributions),  $\varepsilon_t$  is drawn from the Gaussian distribution  $\mathcal{N}(0, \sigma_1^2)$  with probability  $p$  and from  $\mathcal{N}(0, \sigma_2^2)$  with probability  $1 - p$ ; specifically, we set:  $\mu_1 = -0.7$ ,  $\mu_2 = 0.7$ ,  $\sigma_1 = 0.32$ ,  $\sigma_2 = 0.95$ ,  $p = 0.5$ , which results in a zero-mean unit-variance distribution with order-3 and order-4 cumulants of 0.85 and 0, respectively.

Figure 2: Distribution of  $\theta$  estimators (univariate case)

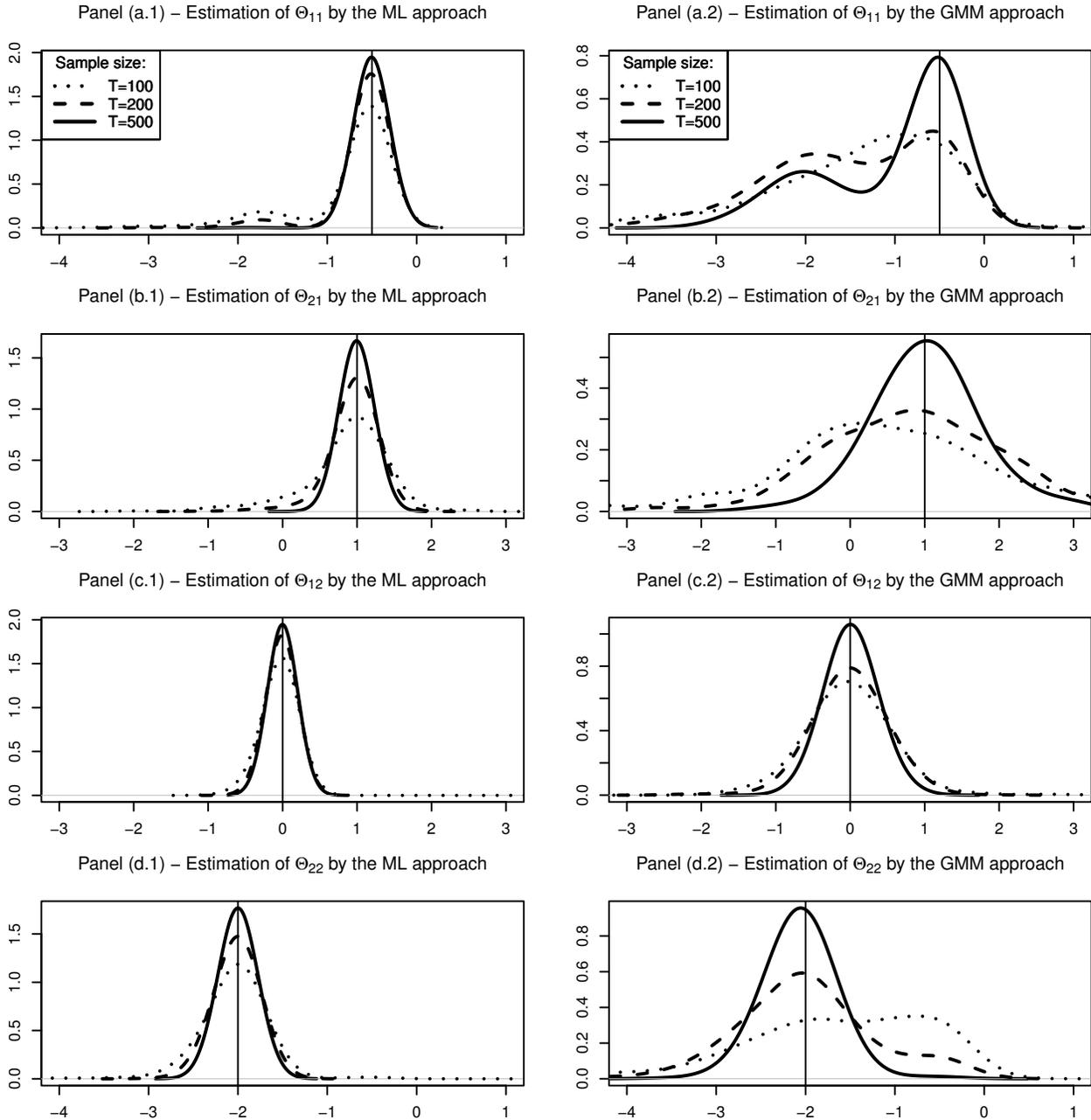
*Note:* The plots display the distributions of the estimates of  $\theta$  obtained by applying the Maximum Likelihood approach (Subsection 4.1) and the GMM approach (Subsection 4.2). The model is  $y_t = \varepsilon_t - \theta\varepsilon_{t-1}$ , with  $\theta = -2$  and  $V(\varepsilon_t) = 1$ . On each panel, the three distributions correspond to three sample sizes:  $T = 100, 200$  and  $500$ . For each distribution of the shocks (see Figure 1) and each sample size, we simulate a large number  $N = 1000$  of  $y_t$  samples of size  $T$ . For each simulated sample, we employ the two approaches to estimate  $(\theta, V(\varepsilon_t), \gamma)$ , where  $\gamma$  characterizes the distribution of  $\varepsilon_t$ . For the (P)ML approach,  $\gamma$  is a vector of three parameters specifying a Gaussian mixture distribution of mean zero and unit variance. When the true distribution is Student, the true and pseudo distributions are different, hence the ML approach is, more precisely, a Pseudo Maximum Likelihood (PML) approach. For the GMM approach,  $\gamma$  contains the order-3 and order-4 cumulants of  $\varepsilon_t$ . The displayed distributions are obtained by applying Gaussian kernel on the  $N$  estimates of  $\theta$ . The vertical dotted bar indicates the true value of  $\theta$ .

Table 1: Results of the Monte-Carlo experiment, univariate case

$\varepsilon_t$ 's distribution:	Bias	RMSE	MAE	S.D.	right reg.	$\alpha = 75\%$	$\alpha = 95\%$
<b>Panel (a) Maximum Likelihood approach</b>							
Sample size: T=100							
Gaussian	0.72	1.14	0.95	0.88	49%	24% / 49%	35% / 72%
Mixture of Gaussian	-0.02	0.22	0.16	0.22	100%	72% / 72%	92% / 92%
Student (df: 5)	0.31	0.88	0.63	0.83	74%	48% / 66%	62% / 84%
Student (df: 10)	0.53	1.04	0.82	0.90	60%	36% / 60%	47% / 78%
Sample size: T=200							
Gaussian	0.71	1.07	0.84	0.79	51%	31% / 60%	43% / 83%
Mixture of Gaussian	0.00	0.13	0.10	0.13	100%	74% / 74%	94% / 94%
Student (df: 5)	0.15	0.57	0.35	0.55	88%	64% / 73%	79% / 90%
Student (df: 10)	0.48	0.90	0.64	0.76	66%	44% / 66%	57% / 87%
Sample size: T=500							
Gaussian	0.72	1.05	0.79	0.76	52%	34% / 65%	44% / 86%
Mixture of Gaussian	0.00	0.08	0.06	0.08	100%	73% / 73%	94% / 94%
Student (df: 5)	0.01	0.23	0.13	0.23	98%	73% / 74%	92% / 94%
Student (df: 10)	0.23	0.61	0.33	0.56	84%	63% / 75%	78% / 92%
<b>Panel (b) Generalized Method of Moments approach</b>							
Sample size: T=100							
Gaussian	0.29	1.55	1.17	1.52	60%	40% / 65%	58% / 95%
Mixture of Gaussian	-0.12	1.31	0.94	1.30	81%	51% / 63%	74% / 91%
Student (df: 5)	0.31	1.47	1.13	1.44	61%	40% / 65%	59% / 94%
Student (df: 10)	0.40	1.93	1.20	1.89	58%	39% / 67%	57% / 96%
Sample size: T=200							
Gaussian	0.47	1.11	0.91	1.00	61%	40% / 65%	57% / 94%
Mixture of Gaussian	-0.01	0.82	0.61	0.82	88%	56% / 64%	80% / 90%
Student (df: 5)	0.28	0.99	0.78	0.95	71%	45% / 63%	66% / 92%
Student (df: 10)	0.49	1.10	0.91	0.98	60%	39% / 64%	57% / 93%
Sample size: T=500							
Gaussian	0.53	0.97	0.74	0.81	62%	43% / 69%	58% / 94%
Mixture of Gaussian	-0.03	0.38	0.27	0.38	98%	70% / 72%	92% / 94%
Student (df: 5)	0.15	0.67	0.46	0.65	84%	55% / 66%	78% / 93%
Student (df: 10)	0.42	0.90	0.67	0.79	68%	46% / 68%	65% / 95%

*Note:* The model is  $y_t = \varepsilon_t - \theta \varepsilon_{t-1}$ , with  $\theta = -2$  and  $V(\varepsilon_t) = 1$ . Results are based on the simulation of  $N = 1000$  samples for each of the four distributions considered for the errors  $\varepsilon_t$  (see first column) and each of the three considered sample sizes ( $T = 100, 200$  or  $500$ ). We estimate  $(\theta, V(\varepsilon_t), \gamma)$ , where  $\gamma$  characterizes the distribution of the shocks. In the ML approach,  $\gamma$  contains three parameters defining a Gaussian mixture of mean zero and unit variance. When the true distribution is Student, the true and pseudo distributions are different, hence the ML approach is, strictly speaking, a Pseudo ML (PML) approach. In the GMM approach,  $\gamma$  contains the order-3 and order-4 cumulants of the shock distribution. Columns 2 to 5 give, respectively: the bias, the root mean-squared error, the mean absolute error and the standard deviation of the estimator of  $\theta$ . Columns 6 (“right reg.”) reports the fraction of cases (among the  $N$  replications) where the right fundamental regime is estimated. The last two columns indicate the fractions of times where the true value of  $\theta$  lies within the interval  $[\hat{\theta} - \phi_\alpha \sigma_{asy}, \hat{\theta} + \phi_\alpha \sigma_{asy}]$  where  $\sigma_{asy}$  denotes the estimate of the asymptotic standard deviation of the estimator  $\hat{\theta}$  (based on the Hessian matrix of the log-likelihood function in the ML case and on the formula given in Subsection 5.2 of the Online Appendix in the GMM case) and where  $\phi_\alpha$  is such that  $P(-\phi_\alpha < X < \phi_\alpha) = \alpha$  if  $X \sim \mathcal{N}(0, 1)$ . Two ratios are shown: contrary to the first one, the second ratio excludes those estimations that lead to the wrong fundamental regime. We discard the replications leading to estimates of  $\theta$  that are larger than 50; the largest number of occurrences is 2 (out of  $N = 1000$  replications), obtained for  $T = 100$  and the GMM approach.

Figure 3: Distribution of the estimators of  $\Theta$ 's components (bivariate case)



*Note:* The plots display the distributions of the estimators of the  $\Theta_{i,j}$ 's,  $i, j \in \{1, 2\}$ , obtained by applying the Maximum Likelihood approach (Subsection 4.1) and the GMM approach (Subsection 4.2). The model is  $Y_t = C\eta_t - \Theta C\eta_{t-1}$ , where the parameterizations of  $C$  and  $\Theta$  are given in (4.27). The true distribution of  $\eta_{1,t}$  is a mixture of Gaussian (see Footnote 33 for parameter values) and the true distribution of  $\eta_{2,t}$  is a Student's distribution with 6 degrees of freedom. On each panel, the three distributions correspond to three sample sizes:  $T = 100, 200$  and  $500$ . For each sample size  $T$ , we simulate a large number ( $N = 500$ ) of  $Y_t$  samples of size  $T$  and employ the ML and GMM approaches to estimate the model parameters. As regards the ML approach, it is assumed that the distribution of both  $\eta_{1,t}$  and  $\eta_{2,t}$  are mixtures of Gaussian distributions (although the true distribution of  $\eta_{2,t}$  is Student); accordingly our ML approach is, strictly speaking, a Pseudo Maximum Likelihood (PML) approach. The displayed distributions are obtained by applying Gaussian kernel on the  $N$  estimates of the  $\Theta_{i,j}$ 's. The vertical dotted bar indicates the true value of  $\Theta_{i,j}$ .

Table 2: Results of the Monte-Carlo experiment, bivariate case

	Bias	RMSE	MAE	S.D.	right side	$\alpha = 75\%$	$\alpha = 95\%$
<b>Panel (a) Maximum Likelihood approach</b>							
Sample size: T=100							
$\Theta_{1,1}$	-0.27	0.64	0.33	0.58	80%	49% / 61%	68% / 83%
$\Theta_{2,1}$	-0.07	1.47	0.46	1.47	80%	60% / 68%	83% / 91%
$\Theta_{1,2}$	-0.02	0.42	0.14	0.42	80%	50% / 55%	76% / 80%
$\Theta_{2,2}$	-0.02	0.41	0.25	0.41	80%	72% / 73%	90% / 90%
Sample size: T=200							
$\Theta_{1,1}$	-0.09	0.35	0.13	0.34	94%	67% / 71%	86% / 92%
$\Theta_{2,1}$	-0.04	0.31	0.21	0.31	94%	67% / 71%	91% / 94%
$\Theta_{1,2}$	-0.01	0.09	0.06	0.09	94%	68% / 71%	87% / 88%
$\Theta_{2,2}$	-0.01	0.18	0.14	0.18	94%	73% / 74%	94% / 94%
Sample size: T=500							
$\Theta_{1,1}$	0.00	0.07	0.03	0.07	100%	76% / 76%	94% / 94%
$\Theta_{2,1}$	0.00	0.13	0.10	0.13	100%	71% / 71%	95% / 96%
$\Theta_{1,2}$	0.00	0.04	0.03	0.04	100%	73% / 73%	95% / 95%
$\Theta_{2,2}$	0.00	0.10	0.08	0.10	100%	77% / 77%	95% / 95%
<b>Panel (b) Generalized Method of Moments approach</b>							
Sample size: T=100							
$\Theta_{1,1}$	-1.01	1.71	1.13	1.38	36%	38% / 59%	66% / 87%
$\Theta_{2,1}$	-0.31	2.16	1.44	2.14	36%	50% / 54%	73% / 76%
$\Theta_{1,2}$	-0.03	0.66	0.42	0.66	36%	51% / 46%	73% / 71%
$\Theta_{2,2}$	0.33	1.27	0.95	1.22	36%	44% / 63%	59% / 85%
Sample size: T=200							
$\Theta_{1,1}$	-0.96	1.33	1.00	0.93	42%	38% / 73%	54% / 90%
$\Theta_{2,1}$	-0.04	1.33	1.01	1.34	42%	59% / 66%	78% / 84%
$\Theta_{1,2}$	0.00	0.44	0.33	0.44	42%	56% / 52%	75% / 70%
$\Theta_{2,2}$	-0.01	0.80	0.57	0.80	42%	58% / 65%	78% / 87%
Sample size: T=500							
$\Theta_{1,1}$	-0.53	0.92	0.58	0.75	67%	53% / 78%	64% / 92%
$\Theta_{2,1}$	0.02	0.73	0.56	0.74	67%	69% / 76%	87% / 91%
$\Theta_{1,2}$	0.01	0.23	0.18	0.23	67%	68% / 71%	86% / 88%
$\Theta_{2,2}$	-0.06	0.34	0.25	0.33	67%	72% / 72%	93% / 94%

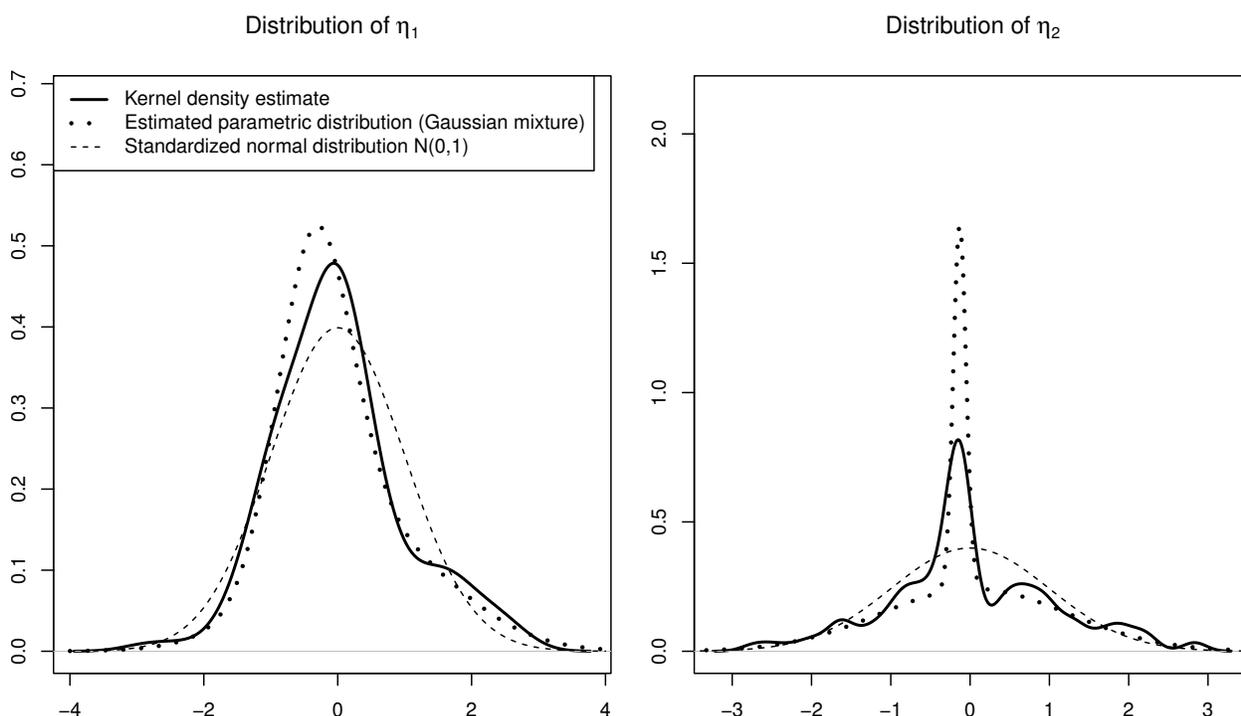
*Note:* This table documents the results of a Monte-Carlo study aimed at assessing the finite sample properties of the Maximum Likelihood approach (Subsection 4.1) and the GMM approach (Subsection 4.2) in the context of a Vectorial Moving-Average model of order 1. See the caption of Figure 3 for details regarding the model and the Monte-Carlo exercise. Columns 2 to 5 give, respectively: the bias, the root mean-squared error, the mean absolute error and the standard deviation of the estimator of the  $\Theta_{i,j}$ 's. Columns 6 ("right side") reports the fraction of cases (among the  $N = 500$  replications) where the right fundamental regime is estimated; specifically, we say that the right fundamental regime is estimated when the two eigenvalues of the estimate of  $\Theta$  lie on each side of the unit circle and when the absolute values of the estimates of  $\Theta_{1,1}$  and  $\Theta_{2,2}$  are respectively below and above 1 (as the true values). The last two columns indicate the fractions of times where the true value of  $\Theta_{i,j}$  lies within the interval  $[\hat{\Theta}_{i,j} - \phi_\alpha \sigma_{asy,i,j}, \hat{\Theta}_{i,j} + \phi_\alpha \sigma_{asy,i,j}]$  where  $\sigma_{asy,i,j}$  denotes the estimate of the asymptotic standard deviation of the estimator  $\hat{\Theta}_{i,j}$  (based on the Hessian matrix of the log-likelihood function in the ML case and on the formula given in Subsection 5.2 of the Online Appendix in the GMM case) and where  $\phi_\alpha$  is such that  $P(-\phi_\alpha < X < \phi_\alpha) = \alpha$  if  $X \sim \mathcal{N}(0, 1)$ . For each parameter and each value of  $\alpha$ , two ratios are shown: contrary to the first one, the second ratio excludes those estimations that lead to the wrong fundamental regime. We discard the replications leading to estimates of the  $\Theta_{i,j}$ 's that are larger than 50 (which, out of  $N = 500$  replications, happens at most once for each sample size and estimation approach).

Table 3: Estimated SVARMA models (GNP and Unemployment rate)

Approach:	GMM		ML	
	param.	std dev.	param.	std dev.
$\Phi_{1,[1,1]}$	0.883	(0.540)	0.641	(0.199)
$\Phi_{1,[2,1]}$	-0.425	(0.355)	-0.229	(0.096)
$\Phi_{1,[1,2]}$	0.064	(1.323)	-0.832	(0.367)
$\Phi_{1,[2,2]}$	1.153	(0.722)	1.744	(0.193)
$\Phi_{2,[1,1]}$	0.210	(0.166)	0.233	(0.082)
$\Phi_{2,[2,1]}$	-0.069	(0.088)	-0.029	(0.038)
$\Phi_{2,[1,2]}$	1.187	(1.661)	2.039	(0.511)
$\Phi_{2,[2,2]}$	-0.565	(0.904)	-1.271	(0.246)
$\Phi_{3,[1,1]}$	-0.165	(0.172)	-0.332	(0.077)
$\Phi_{3,[2,1]}$	0.019	(0.091)	0.035	(0.038)
$\Phi_{3,[1,2]}$	-1.763	(0.995)	-1.467	(0.434)
$\Phi_{3,[2,2]}$	0.600	(0.688)	0.609	(0.208)
$\Phi_{4,[1,1]}$	0.048	(0.107)	0.164	(0.059)
$\Phi_{4,[2,1]}$	0.016	(0.072)	-0.012	(0.031)
$\Phi_{4,[1,2]}$	0.604	(0.490)	0.321	(0.182)
$\Phi_{4,[2,2]}$	-0.198	(0.320)	-0.101	(0.098)
$\Theta_{1,1}$	0.814	(0.328)	0.388	(0.191)
$\Theta_{2,1}$	-0.468	(0.126)	0.005	(0.096)
$\Theta_{1,2}$	-2.733	(0.858)	-2.048	(0.305)
$\Theta_{2,2}$	1.496	(0.393)	1.775	(0.148)
$C_{1,1}$	0.298	(0.092)	0.555	(0.044)
$C_{2,1}$	0.172	(0.028)	-0.195	(0.017)
$C_{1,2}$	0.689	(0.082)	0.645	(0.059)
$C_{2,2}$	0.007	(0.072)	-0.006	(0.017)
$\kappa_{3,1}$	3.136	(0.126)	$\mu_{1,1}$	0.421 (0.209)
$\kappa_{3,2}$	0.071	(0.328)	$\mu_{2,1}$	-0.130 (0.020)
$\kappa_{4,1}$	19.828	(0.393)	$\sigma_{1,1}$	1.242 (0.099)
$\kappa_{4,2}$	1.459	(0.858)	$\sigma_{2,1}$	0.081 (0.012)
			$p_1$	0.448 (0.115)
			$p_2$	0.286 (0.040)

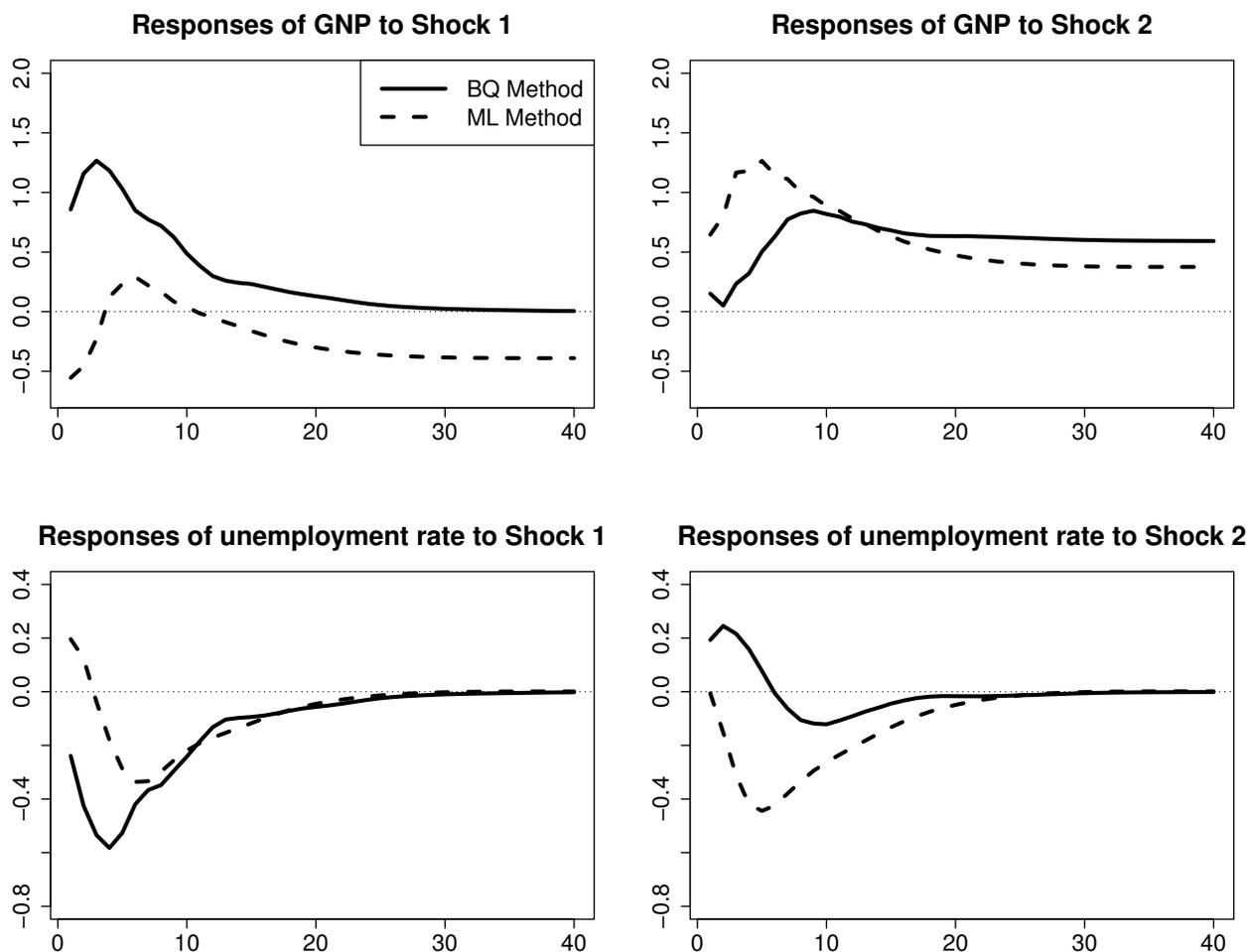
Note: This table reports the results of the estimation of bivariate VARMA(4,1) models. The dataset is the one used in Blanchard and Quah (1989); it covers the period from 1948Q1 to 1987Q4 at the quarterly frequency. The first endogenous variable is the U.S. real GNP growth (i.e.  $\Delta gnp$ , where  $gnp = \log(GNP)$ ) and the second endogenous variable is the unemployment rate. The model is:  $\Phi(L)Y_t = (I - \Theta L)C\eta_t$ , where, for  $j \in \{1, 2\}$ ,  $\eta_{j,t}$  is drawn from  $\mathcal{N}(\mu_{j,1}, \sigma_{j,1}^2)$  with probability  $p_j$  and from  $\mathcal{N}(\mu_{j,2}, \sigma_{j,2}^2)$  with probability  $1 - p_j$ . We impose  $E(\eta_{j,t}) = 0$  and  $V(\eta_{j,t}) = 1$ , which implies that  $\mu_{j,2}$  and  $\sigma_{j,2}$  can be deduced from  $\mu_{j,1}$ ,  $\sigma_{j,1}$  and  $p_j$ . Both the Maximum Likelihood (ML) and the 2SLS-GMM approaches are employed to estimate the model (see Subsections 4.1 and 4.2, respectively). Asymptotic standard deviations are reported in parentheses.  $\Phi_{k,[i,j]}$  is the  $(i, j)$  element of  $\Phi_k$ , with  $\Phi(L) = I - \Phi_1 L - \dots - \Phi_p L^p$ .

Figure 4: Estimated shock distributions

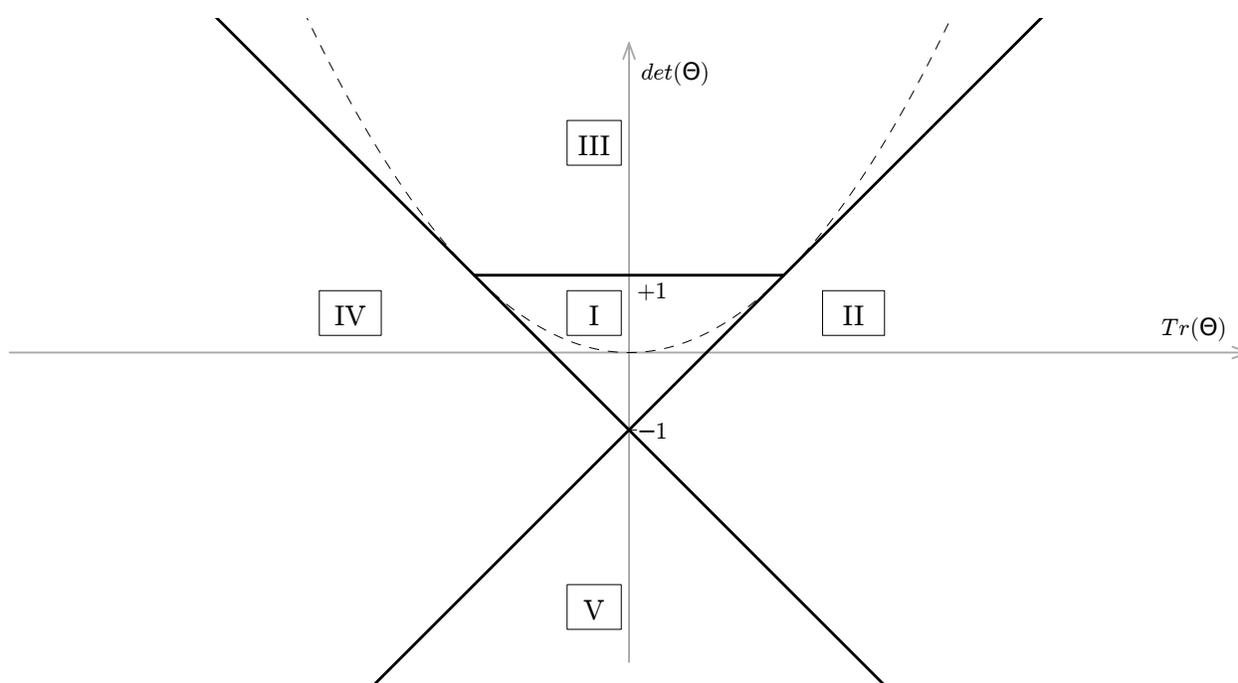


*Note:* This figure displays the estimated p.d.f. of the structural shocks  $\eta_{j,t}$ , for  $j \in \{1, 2\}$  resulting from the ML estimation of a SVARMA(4,1) model using [Blanchard and Quah \(1989\)](#)'s dataset. The dotted lines correspond to the estimated parametric distributions (mixture of Gaussian distributions):  $\eta_{j,t}$  is drawn from  $\mathcal{N}(\mu_{j,1}, \sigma_{j,1}^2)$  with probability  $p_j$  and from  $\mathcal{N}(\mu_{j,2}, \sigma_{j,2}^2)$  with probability  $1 - p_j$ ;  $\mu_{j,2}$  and  $\sigma_{j,2}$  are computed so as to have  $E(\eta_{j,t}) = 0$  and  $V(\eta_{j,t}) = 1$ . The model parameterization is detailed in the caption of [Table 3](#). The black solid lines correspond to kernel-based density estimates of the distribution of the (estimated) structural shocks  $\eta_{j,t}$ . The latter are computed by applying the backward/forward filtering algorithm underlying function  $a_t$  (see [Subsection 4.1.2](#) for details on this algorithm). For the sake of comparison, we also report the  $\mathcal{N}(0, 1)$  p.d.f. (dashed lines).

Figure 5: IRFs – GNP and unemployment rate



*Note:* This figure compares impulse response functions (IRFs) associated with models resulting from the BQ approach (long-run restrictions à la Blanchard and Quah, 1989, 8-lag VAR model) as well as the Maximum Likelihood (ML) approach (Subsection 4.1, VARMA(4,1) model). The dataset is the same as in Blanchard and Quah (1989). In the context of the BQ approach, whereas Shock 1 (first column of plots) is interpreted as a demand shock, Shock 2 (second column of plots) is interpreted as a supply shock. In BQ, by construction, the long-run impact of the demand shock (Shock 1) on real GNP is null. In the ML approach, Shock 1 is defined as the shock having the lower influence on the long-run GNP variance. For the GNP variable, we report the cumulated impacts of the shocks on the (log) GNP growth rate.

Figure 6: Nonfundamentalness regimes when  $n = 2$ 

*Note:* This figure shows the 5 regions corresponding to the different (non)fundamentalness regimes when  $Y_t$  is of dimension 2 (i.e. when  $\Theta$  is of dimension  $2 \times 2$ ). Region I is purely fundamental; Regions III and V are purely nonfundamental; Regions II and IV are mixed. The dashed line represents the pairs  $(Tr(\Theta), det(\Theta))$  for which the discriminant  $Tr(\Theta)^2 - 4det(\Theta)$  is zero; the area that is below (respectively above) this parabola corresponds to real (resp. complex) eigenvalues for  $\Theta$ .

## References

- Akaike, H., 1974, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Alessi, L., M. Barigozzi, and M. Capasso, 2011, "Non-Fundamentalness in Structural Econometric Models: A Review," *International Statistical Review*, 79(1), 16–47.
- Bai, J., and S. Ng, 2005, "Tests for Skewness, Kurtosis, and Normality for Time Series Data," *Journal of Business & Economic Statistics*, 23, 49–60.
- Barsky, R. B., and E. R. Sims, 2012, "Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence," *American Economic Review*, 102(4), 1343–1377.
- Bates, C., and H. White, 1985, "A Unified Theory of Consistent Estimation for Parametric Models," *Econometric Theory*, 1(02), 151–178.
- Beaudry, P., P. Fève, A. Guay, and F. Portier, 2015, "When is Nonfundamentalness in VARs a Real Problem? An Application to News Shocks," NBER Working Papers 21466, National Bureau of Economic Research.
- Bernanke, B. S., 1986, "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25(1), 49–99.
- Blanchard, O. J., J.-P. L’Huillier, and G. Lorenzoni, 2013, "News, Noise, and Fluctuations: An Empirical Exploration," *American Economic Review*, 103(7), 3045–3070.
- Blanchard, O. J., and D. Quah, 1989, "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79(4), 655–73.
- Bonhomme, S., and J.-M. Robin, 2009, "Consistent Noisy Independent Component Analysis," *Journal of Econometrics*, 149(1), 12–25.
- Boudt, K., D. Cornilly, and T. Verdonck, 2018, "Nearest Comoment Estimation with Unobserved Factors," working paper.
- Box, G. E. P., and G. Jenkins, 1970, *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.
- Box, G. E. P., and D. A. Pierce, 1970, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65(332), 1509–1526.
- Breidt, F. J., and R. A. Davis, 1992, "Time-Reversibility, Identifiability and Independence of Innovations for Stationary Time Series," *Journal of Time Series Analysis*, 13(5), 377–390.
- Brockwell, P. J., and R. A. Davis, 1991, *Time Series: Theory and Methods*. 2nd ed., Springer-Verlag New York., New York, NY, USA.
- Chahrour, R., and K. Jurado, 2018, "News or Noise? The Missing Link," *American Economic Review*, 108(7), 1702–1736.
- Chan, K.-S., and L.-H. Ho, 2004, "On the Unique Representation of non-Gaussian Multivariate Linear Processes," Working papers, Department of Statistics & Actuarial Science, University of Iowa.
- Chan, K.-S., L.-H. Ho, and H. Tong, 2006, "A Note on Time-Reversibility of Multivariate Linear Processes," *Biometrika*, 93(1), 221–227.
- Chari, V., P. J. Kehoe, and E. R. McGrattan, 2008, "Are Structural VARs with Long-run Restrictions Useful in Developing Business Cycle Theory?," *Journal of Monetary Economics*, 55(8), 1337–1352.
- Chen, B., J. Choi, and J. C. Escanciano, 2017, "Testing for Fundamental Vector Moving Average Representations," *Quantitative Economics*, 8(1), 149–180.
- Chen, M.-C., R. A. Davis, and L. Song, 2011, "Inference for Regression Models with Errors from a Non-invertible MA(1) Process," *Journal of Forecasting*, 30(1), 6–30.
- Cheng, Q., 1992, "On The Unique Representation of Non-Gaussian Linear Processes," *The Annals of Statistics*, 20, 1143–1145.

- Christiano, L. J., M. Eichenbaum, and R. Vigfusson, 2006, "Alternative Procedures for Estimating Vector Autoregressions Identified with Long-Run Restrictions," *Journal of the European Economic Association*, 4(2-3), 475–483.
- , 2007, "Assessing Structural VARs," in *NBER Macroeconomics Annual*. MIT Press, Cambridge, pp. 1–106.
- Cochrane, J. H., 1994, "Shocks," *Carnegie-Rochester Conference Series on Public Policy*, 41, 295–364.
- Comon, P., 1994, "Independent Component Analysis, A New Concept?," *Signal Processing*, 36, 287–314.
- Darmois, G., 1953, "Analyse Générale des Liaisons Stochastiques: Etude Particulière de l'Analyse Factorielle Linéaire," *Revue de l'Institut International de Statistique*, 21(1/2), 2–8.
- Deistler, M., and J. Schrader, 1979, "Linear Models with Autocorrelated Errors: Identifiability without Minimality Assumptions," *Econometrica*, 47, 495–504.
- Economic Sciences Prize Committee, 2011, "Thomas J. Sargent and Christopher A. Sims: Empirical Macroeconomics," Nobel Prize in Economics documents 2011-2, Nobel Prize Committee.
- Engle, R. F., and C. W. J. Granger, 1987, "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55(2), 251–276.
- Erceg, C., G. L., and C. Gust, 2005, "Can Long Run Restriction Identify Technology Shocks?," *Journal of the European Economic Association*, 3, 1237–1278.
- Eriksson, J., and V. Koivunen, 2004, "Identifiability, Separability and Uniqueness of Linear ICA Models," *IEEE Signal Processing Letters*, 11, 601–604.
- Evans, C., and D. Marshall, 2011, "Fundamental Economic Shocks and the Macroeconomy," Discussion Paper 351, Central Bank of Chile.
- Faust, J., and E. Leeper, 1997, "When Do Long Run Identifying Restrictions Give Reliable Results?," *Journal of Business and Economic Statistics*, 15, 345–353.
- Fernandez-Villaverde, J., J. Rubio-Ramirez, T. Sargent, and M. Watson, 2007, "A,B,C's (and D)'s for Understanding VAR's," *American Economic Review*, 97, 1021–1026.
- Fève, P., J. Matheron, and J.-G. Sahuc, 2009, "On the Dynamic Implications of News Shocks," *Economics Letters*, 102, 96–98.
- Findley, D., 1986, "The Uniqueness of Moving Average Representation with Independent and Identically Distributed Random Variables for Non-Gaussian Stationary Time Series," *Biometrika*, 73, 520–521.
- Forni, M., and L. Gambetti, 2014, "Sufficient Information in Structural VARs," *Journal of Monetary Economics*, 66(C), 124–136.
- Forni, M., L. Gambetti, M. Lippi, and L. Sala, 2017a, "Noise Bubbles," *Economic Journal*, 127(604), 1940–1976.
- , 2017b, "Noisy News in Business Cycles," *American Economic Journal: Macroeconomics*, 9(4), 122–152.
- Francis, N., and V. Ramey, 2005, "Is Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited," *Journal of Monetary Economics*, 52, 1379–1399.
- Gagliardini, P., and C. Gouriéroux, 2018, "Identification by Laplace Transforms in Nonlinear Panel or Time Series Models with Unobserved Stochastic Dynamic Effects," *Journal of Econometrics*, forthcoming.
- Galbraith, J., A. Ullah, and V. Zinde-Walsh, 2002, "Estimation of the Vector Moving Average Model by Vector Autoregression," *Econometric Reviews*, 21(2), 205–219.
- Gali, J., and P. Rabanal, 2005, "Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit Postwar US Data?," in *NBER Macroeconomics Annual*. MIT Press, Cambridge, vol. 19.
- Ghurye, S., and I. Olkin, 1961, "A Characterization of the Multivariate Normal Distribution," *Annals of Mathematical Statistics*, 33, 533–541.
- Giannone, D., and L. Reichlin, 2006, "Does Information Help Recovering Structural Shocks From Past Observations?," *Journal of European Economic Association*, 4, 455–465.

- Gospodinov, N., and S. Ng, 2015, “Minimum Distance Estimation of Possibly Noninvertible MA Process,” *Journal of Business and Economic Statistics*, 33, 403–417.
- Gouriéroux, C., and J. Jasiak, 2016, “Filtering and Prediction in Noncausal Processes,” *Journal of Time Series Analysis*, 37, 405–430.
- Gouriéroux, C., A. Monfort, and E. Renault, 1989, “Testing for Common Roots,” *Econometrica*, 57, 171–185.
- Gouriéroux, C., A. Monfort, and J.-P. Renne, 2017, “Statistical Inference for Independent Component Analysis,” *Journal of Econometrics*, 196, 111–126.
- Gouriéroux, C., and J.-M. Zakoian, 2015, “On Uniqueness of Moving Average Representation of Heavy Tailed Stationary Processes,” *Journal of Time Series Analysis*, 36, 876–887.
- Grassi, S., F. Ferroni, and M. Leon-Ledesma, 2015, “Fundamental Shock Selection in DSGE Models,” working paper.
- Guay, A., and M. Normandin, 2018, “Identification of Structural Vector Autoregressions Through Higher Unconditional Moments,” Working paper, UQAM and HEC Montreal.
- Hannan, E., and M. Deistler, 1996, *The Statistical Theory of Linear Systems*. Wiley, New York edn.
- Hannan, E. J., and B. G. Quinn, 1979, “The Determination of the Order of an Autoregression,” *Journal of the Royal Statistical Society*, 41(2), 190–195.
- Hansen, L., and R. Hodrick, 1980, “Forward Exchange Rates as Optimal Predictors of Future Sport Rates: An Econometric Analysis,” *Journal of Political Economy*, 88, 829–853.
- Hansen, L., and T. Sargent, 1980, “Formulating and Estimating Dynamic Linear Rational Expectations Models,” *Journal of Economic Dynamics and Control*, 2, 7–46.
- , 1991, “Two Difficulties in Interpreting Vector Autoregressions,” in *Rational Expectations Econometrics*. L., Hansen and T., Sargent, editors, Westview Press, Boulder, vol. 19.
- Herwartz, H., and H. Lütkepohl, 2017, “Structural Vector Autoregressions with Markov Switching: Combining Conventional with Statistical Identification of Shocks,” *Journal of Econometrics*, forthcoming.
- Hyvärinen, A., J. Karhunen, and E. Oja, 2001, *Independent Component Analysis*. Wiley, New York edn.
- Jennrich, R., 1969, “Asymptotic Properties of Nonlinear Least Squares Estimation,” *The Annals of Mathematical Statistics*, 40, 633–643.
- Koop, G., M. H. Pesaran, and S. M. Potter, 1996, “Impulse Response Analysis in Nonlinear Multivariate Models,” *Journal of Econometrics*, 74(1), 119–147.
- Kydland, F., and E. Prescott, 1982, “Time to Build and Aggregate Fluctuations,” *Econometrica*, 50, 1345–1370.
- Lanne, M., H. Lütkepohl, and K. Maciejowska, 2010, “Structural Vector Autoregressions with Markov Switching,” *Journal of Economic Dynamics and Control*, 34, 121–131.
- Lanne, M., M. Meitz, and P. Saikkonen, 2017, “Identification and Estimation of Non-Gaussian Structural Vector Autoregressions,” *Journal of Econometrics*, 196(2), 288–304.
- Leeper, E., T. Walker, and S. Yang, 2013, “Fiscal Foresight and Information Flows,” *Econometrica*, 81, 1115–1145.
- Lii, K.-S., and M. Rosenblatt, 1992, “An Approximate Maximum Likelihood Estimation for Non-Gaussian Non-minimum Phase Moving Average Processes,” *Journal of Multivariate Analysis*, 43, 272–299.
- , 1996, “Maximum Likelihood Estimation for Non-Gaussian Nonminimum Phase ARMA Sequences,” *Statistica Sinica*, 6, 1–22.
- Lippi, M., and L. Reichlin, 1993, “The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment,” *American Economic Review*, 83, 644–652.
- , 1994, “VAR Analysis, Nonfundamental Representations, Blaschke Matrices,” *Journal of Econometrics*, 63, 307–325.
- Ljung, G. M., and G. E. P. Box, 1978, “On a Measure of Lack of Fit in Time Series Models,” *Biometrika*,

- 65(2), 297–303.
- Lobato, I. N., and C. Velasco, 2004, “A Simple Test of Normality for Time Series,” *Econometric Theory*, 20(4), 671–689.
- , 2018, “Frequency Domain Minimum Distance Inference for Possibly Noninvertible and Noncausal ARMA Models,” *Annals of Statistics*, 46(2), 555–579.
- Lof, M., 2013, “Noncausality and Asset Pricing,” *Studies in Nonlinear Dynamics and Econometrics*, 17, 211–220.
- Lorenzoni, G., 2009, “A Theory of Demand Shocks,” *American Economic Review*, 99(5), 2050–2084.
- Lütkepohl, H., 2013, “Identifying Structural Vector Autoregression via Changes in Volatility,” in *VAR Models in Macroeconomics, Essays in Honor of Christopher Sims, Advances in Econometrics*. Thomas B. Fomby, Lutz Kilian, Anthony Murphy, editors, Emerald Group Publishing Limited, vol. 32.
- , 2014, “Fundamental Problems with Nonfundamental Shocks,” in *Essays in Nonlinear Time Series Econometrics*. N. Haldrup, M. Meltz and P. Saikkonen, editors, Oxford University Press.
- Lütkepohl, H., and B. Theilen, 1991, “Measures of Multivariate Skewness and Kurtosis for Tests of Non-normality,” *Statistical Papers*, 32(1), 179–193.
- Mertens, K., and M. Ravn, 2010, “Measuring the Impact of Fiscal Policy in the Face of Anticipation: A Structural VAR Approach,” *Economic Journal*, 120(544), 393–413.
- Mountford, A., and H. Uhlig, 2009, “What Are the Effects of Fiscal Policy Shocks?,” *Journal of Applied Econometrics*, 24, 960–992.
- Pesaran, H. H., and Y. Shin, 1998, “Generalized Impulse Response Analysis in Linear Multivariate Models,” *Economics Letters*, 58(1), 17–29.
- Rosenblatt, M., 2000, *Gaussian and Non Gaussian Linear Time Series and Random Fields*. Springer Verlag, New York edn.
- Rozanov, Y., 1960, “Spectral Properties of Multivariate Stationary Processes and Boundary Properties of Analytic Matrices,” *Theory of Probability and its Applications*, 5, 362–376.
- , 1967, *Stationary Random Processes*. Holden Day edn.
- Rubio-Ramirez, J., D. Waggoner, and T. Zha, 2010, “Structural Vector Autoregression: Theory of Identification and Algorithms for Inference,” *Review of Economic Studies*, 77, 665–696.
- Schwarz, G., 1978, “Estimating the Dimension of a Model,” *The Annals of Statistics*, 6(2), 461–464.
- Shapiro, S. S., and M. B. Wilk, 1965, “An Analysis of Variance Test for Normality (Complete Samples),” *Biometrika*, 52(3-4), 591–611.
- Sims, C., 1980, “Macroeconomics and Reality,” *Econometrica*, 48, 1–48.
- , 1986, “Are Forecasting Models Usable for Policy Analysis?,” *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 1–16.
- , 1989, “Models and their Uses,” *American Journal of Agricultural Economics*, 71, 489–494.
- Sims, E. R., 2012, “News, Non-Invertibility, and Structural VARs,” Working Papers 013, University of Notre Dame, Department of Economics.
- Skitovich, V., 1953, “On a Property of a Normal Distribution,” *Doklady Akademii nauk SSSR, (in Russian)*, 89, 217–219.
- Smets, F., and R. Wouters, 2003, “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area,” *Journal of the European Economic Association*, 1, 1123–1175.
- Söderlind, P., 1999, “Solution and Estimation of RE Macromodels with Optimal Policy,” *European Economic Review*, 43(4-6), 813–823.
- Uhlig, H., 2005, “What Are the Effects of Monetary Policy on Output? Results From an Agnostic Identification Procedure,” *Journal of Monetary Economics*, 52, 381–419.
- Velinov, A., and W. Chen, 2014, “Are There Bubbles in Stock Prices? Testing for Fundamental Shocks,” Dp diw berlin, DIW.

- Wald, A., 1949, "Note on the Consistency of the Maximum Likelihood Estimate," *The Annals of Mathematical Statistics*, 20(4), 595–601.
- Weiss, G., 1975, "Time-Reversibility of Linear Stochastic Processes," *Journal of Applied Probability*, 12(4), 831–836.
- Wu, R., and R. A. Davis, 2010, "Least Absolute Deviation Estimation for General Autoregressive Moving Average Time-Series Models," *Journal of Time Series Analysis*, 31(2), 98–112.