CREST
CENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

**Série des Documents de Travail**

**n° 2017-08**

# Identification and Estimation in Non-Fundamental Structural VARMA Models

**C. GOURIEROUX**[1]
**A.MONFORT**[2]
**J-P.RENNE**[3]



[1] CREST, University of Toronto. E-mail : Christian.gourieroux@ensae.fr
[2] CREST, Banque de France E-mail : Alain.monfort@ensae.fr
[3] University of Lausanne E-mail : Jean-Paul.Renne@unil.ch

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

Christian Gouriéroux
CREST and University of Toronto

Alain Monfort
CREST and Banque de France

Jean-Paul Renne
University of Lausanne

(May 2017)

**Abstract**

The basic assumption of a structural VARMA model (SVARMA) is that it is driven by a white noise whose components are 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 kinds of 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 the SVARMA process may have a non invertible AR and/or MA matrix polynomial but, still, has the same second-order properties as a VARMA process in which both the AR and MA matrix polynomials are invertible (the fundamental representation). Moreover the standard Box-Jenkins approach [Box and Jenkins (1970)] automatically estimates the fundamental representation and, therefore, may lead to misspecified Impulse Response Functions. 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. We develop new simple parametric and semi-parametric estimation methods when there is non-fundamentalness 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, Shocks, Impulse Response Function, Incomplete Likelihood, Composite Likelihood, Economic Scenario Generators.

*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 for helpful comments. 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 independent and are interpreted as economic shocks,[1] 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 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 the stationary SVARMA process may feature a non-invertible moving average (MA) matrix lag polynomial. The latter situation, called non-fundamentalness, may occur 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 [see e.g. Hansen and Sargent (1991), Smets and Wouters (2003), Christiano, Eichenbaum, and Vigfusson (2007), Leeper, Walker, and Yang (2013)]. Typically the matrix MA polynomial is not invertible and the shock vector is not simply linearly linked to the (linear) causal innovation of the process [see e.g. Lippi and Reichlin (1993, 1994)]. Moreover the non-fundamental SVARMA process has exactly the same second-order 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 exarcerbated 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)]– provides a consistent estimation of the fundamental representation and, therefore, may lead to misspecified Impulse Response Functions (IRFs).

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. We also introduce simple semi-parametric and parametric estima-

[1] Our paper will not consider the debate about how structural are the parameters and the shocks in SVARMA models [see e.g. Pesaran and Smith (2011) and the reference therein, as well as the application in Gouriéroux, Monfort, and Renne (2017)].

tion approaches when there is a potential 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 linear innovations. We review different examples of non-fundamental representations in the moving average dynamics given in the literature. Next we discuss the identification issue in the Gaussian case and explain why the standard Box-Jenkins approach based on Gaussian pseudo-likelihood suffers from these identification issues.

Section 3 is the core of the paper. We consider 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]. We explain that, in this context, the standard static and dynamic identification problems encountered in the Gaussian SVARMA analysis disappear and we also discuss the identification of the structural shocks and of the Impulse Response Functions (IRFs) when the shocks are deterministic or stochastic. In Section 4 we suggest new parametric and semi-parametric estimation methods to improve the standard SVAR methodology. We first consider a semi-parametric SVARMA with non-fundamental representation in the moving average dynamics only and introduce a two-step moment approach to estimate the autoregressive and moving average parameters as well as the distributions of the errors. When the distribution of the error term is parametrically specified, it is easily seen that the maximum likelihood approach is computationally demanding, since the different regimes of possible non-fundamentalness can have to be considered. To circumvent this difficulty, we introduce estimators based on appropriate incomplete likelihood functions and composite likelihood functions. Applications are provided in Section 5. First, we conduct a Monte-Carlo analysis aimed at illustrating the performances of the parametric approaches in the context of a univariate ARMA(1) process. Second, the maximum likelihood estimation procedure is employed to estimate the processes followed by GDP growth rates of different countries. Third, following Blanchard and Quah (1989), Lippi and Reichlin (1993, 1994), we study the joint dynamics of U.S. GDP growth and unemployment rates; our results suggest that the data call for non-fundamental bivariate VARMA models. Section 6 concludes.

The special case of a one-dimensional MA(1) process is completely analysed in Appendix A. Appendix B provides a proof of the key proposition and Appendix C shows how to recover the structural shocks.

# 2 Dynamic Linear Model and Non-Fundamentalness

## 2.1 The dynamic 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:

$$\Phi(L)Y_t = \Theta(L)\varepsilon_t, \tag{2.1}$$

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

$$\Phi(L) = I - \Phi_1 L - \ldots - \Phi_p L^p, \Theta(L) = I - \Theta_1 L - \ldots - \Theta_q L^q, \tag{2.2}$$

and the matrix autoregressive and moving average lag-polynomials are of degree $p$ and $q$, respectively.[2]

Let us now introduce the following assumptions on model (2.1)-(2.2):

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

*i) The process $\varepsilon_t$ is a square-integrable strong white noise, i.e. the errors $\varepsilon_t$ are independently, identically distributed and such that $E(\varepsilon_t) = 0$ and $E(\|\varepsilon_t\|^2) < \infty$.*

*ii) The errors can be written as $\varepsilon_t = C\eta_t \Leftrightarrow \eta_t = C^{-1}\varepsilon_t$, where the components $\eta_{j,t}$ of $\eta_t$ are mutually independent, and satisfy $V(\eta_{j,t}) = 1$.*

Assumption A.1 i) on the errors is standard in the literature. Assumption A.1 ii) is required for defining separate shocks on the system when deriving the impulse response functions (see the discussion in Section 3.2). The random variables $\eta_{j,t}$ are usually called "structural shocks", and the representation $\Phi(L)Y_t = \Theta(L)C\eta_t$ is called a "structural" VARMA or SVARMA representation.

**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 $L$.*

[2]The underlying structural model may include state variables, which are not necessarily observable. This explains why the number of shocks $m$, corresponding to the number of state variables might be larger than the number $n$ of observed variables $Y$, even if models considered in practice are often such that $n = m$ [see e.g. Hansen and Sargent (1991), p83, Lippi and Reichlin (1994), Giannone and Reichlin (2006), p457, Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson (2007) Section C, for this assumption on the dimensions].

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 will greatly simplify 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 clearly out of the scope of this paper.[3]

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

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

Under Assumptions A.1–A.3, the linear dynamic system (2.1)-(2.2) 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 RHS of (2.1) is $\mu + \Theta(L)\varepsilon_t$, the process $Y_t - [\Phi(1)]^{-1}\mu$ satisfies (2.1) without intercept $\mu$; we can therefore assume $\mu = 0$ without loss of generality.

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

*The observable process is the stationary solution of model (2.1) associated with the true values of* $\Phi$, $\Theta$, $C$ *and of the 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 polynomial operator $\Phi(L)$ as a convergent one-sided series in the lag operator $L$:

$$\begin{aligned} \Phi(L)Y_t &= \Theta(L)\varepsilon_t \\ \iff Y_t &= \Phi(L)^{-1}\Theta(L)\varepsilon_t \equiv \Psi(L)\varepsilon_t = \sum_{k=0}^{\infty}\Psi_k L^k \varepsilon_t = \sum_{k=0}^{\infty}\Psi_k \varepsilon_{t-k} \\ &= \sum_{k=0}^{\infty}\Psi_k C\eta_{t-k} = \sum_{k=0}^{\infty} A_k \eta_{t-k}, \end{aligned} \tag{2.3}$$

with $A_k = \Psi_k C$.

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

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

[3] 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.

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 will say that the operator is invertible.

From the macroeconomic literature we know that SVARMA models do not always have roots of the moving average located outside the unit circle (see Section 2.2). If $\det(\Theta(z))$ has no roots on the unit circle, but some roots inside the unit circle, we get a two-sided autoregressive representation (see Appendix C):

$$\sum_{k=-\infty}^{\infty} B_k Y_{t-k} = \varepsilon_t,$$

and

$$\eta_t = C^{-1} \sum_{k=-\infty}^{\infty} B_k Y_{t-k}.$$

Here $B(L) = \sum_{k=-\infty}^{\infty} B_k L^k = \Theta^{-1}(L)\Phi(L)$ where $\Theta^{-1}(L)$ is the two-sided series operator satisfying $\Theta^{-1}(L)\Theta(L) = I$. In this case, we will say that $\Theta(L)$ is invertible in a general sense.

Let us now study the consequences of ill-located roots of $\det(\Theta(z))$. For expository purpose, we consider a one-dimensional ARMA(1,1) process:

$$(1-\varphi L)y_t = (1-\theta L)\varepsilon_t, \tag{2.4}$$

where $|\varphi| < 1$ and $|\theta| > 1$. Thus the root of $\det(\Theta(z))$ is ill-located, that is inside the unit circle. To get the (infinite) pure autoregressive representation of process $(y_t)$, we have to invert $(1-\theta L)$. We get:

$$\begin{aligned} (1-\varphi L)y_t &= \left(1-\frac{1}{\theta}L^{-1}\right)(-\theta L\varepsilon_t) \\ \Leftrightarrow \quad \left(1-\frac{1}{\theta}L^{-1}\right)^{-1}&(1-\varphi L)y_t = -\theta L\varepsilon_t. \end{aligned} \tag{2.5}$$

Formula (2.4) implies that

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

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

Formula (2.5) implies that $\varepsilon_t$ is function of the present and future values of $y_t$ and the information $\{y_{t-1}, y_{t-2}, \dots\}$ is strictly included in the information $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$; therefore, $\varepsilon_t$ is not the innovation of $y_t$, defined by $y_t - E(y_t|y_{t-1}, y_{t-2}, \dots)$.

To summarize, under Assumptions A.1-A.3, 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, we say that process $Y_t$ has a fundamental VARMA representation [see e.g. Hansen and Sargent (1980), p18, (1991), p79, and Lippi and Reichlin (1994) for the introduction of this terminology in the macroeconometric literature].[4,5] Otherwise, $\varepsilon_t$ is not equal to the causal innovation, and the VARMA representation is non-fundamental.

## 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.

**i) Lagged impact.** A well-known example appears in the comment of the Blanchard, Quah model [Blanchard and Quah (1989)] by Lippi and Reichlin (1993). The productivity, $y_t$, can be written as:

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

where $\varepsilon_t$ denotes the shock on productivity. 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. that $\theta > 1$. The MA(1) process is then non-fundamental (or non-invertible).

**ii) Non-observability.** Non-fundamentalness can also 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:

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

where $c_t$ (resp. $y_t$) denotes the consumption (resp. the surplus), $R > 1$ a constant gross interest rate on financial assets, and $\varepsilon_t$ is an i.i.d. labor income process. From the first

[4]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. At any date $t$, the information contained in the current and past values of the fundamental process coincides with the information contained in the current and past values of the observations. In Box and Jenkins (1970), a fundamental representation is also called invertible.

[5]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.

equation, we deduce:

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

and by substituting in the second equation, we get the dynamics of $y_t$ as:

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

Thus the root of the moving-average lag-polynomial is equal to $1/aR$. It is smaller than one when $aR > 1$.[6]

**iii) Rational expectation.** Other sources of non-fundamentalness are the rational expectations introduced in the models. In the simple example of Hansen and Sargent (1991) the economic 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 at date $t$ is $I_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$, 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 $\theta$, this root is larger or smaller than 1. When the root is strictly smaller than 1, the information contained in the observations, i.e. $(y_t, y_{t-1}, \ldots)$, is strictly included in $I_t$. In other words the information of the econometrician differs from the information of the economic agent.

**iv) Rational expectation and lagged impact.** Non-fundamentalness may also occur when the economic agent and econometrician information sets are not aligned. The literature on information flows applied to fiscal foresight or productivity belongs to this category [see e.g. Fève, Matheron, and Sahuc (2009), Fève and Jihoud (2012), Forni and Gambetti (2010), Leeper, Walker, and Yang (2013)]. A stylized model is [see Fève, Matheron, and Sahuc

[6] 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.

(2009)]:

$$\begin{aligned} y_t &= aE_t y_{t+1} + x_t, \\ x_t &= \varepsilon_{t-q}, \end{aligned}$$

where $\varepsilon_t$ is a white noise, and $E_t$ is the conditional expectation given $\varepsilon_t$, $\varepsilon_{t-1}$,...

If $|a| < 1$, the forward solution is easily seen to be:

$$y_t = \sum_{i=0}^{q} a^{q-i} \varepsilon_{t-i}. \tag{2.6}$$

The roots of $\Theta(L) = a^q \sum_{i=0}^{q} a^{-i} L^i = a^q \dfrac{1-(a^{-1}L)^{q+1}}{1-a^{-1}L}$ are equal to: $a\exp(2ik\pi/(q+1)), k = 1,\ldots,q$, with common modulus $|a| < 1$. Therefore the moving-average lag polynomial $\Theta(L)$ is noninvertible and the MA process is non-fundamental.

**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) and, also as the forward price (if, moreover, the riskfree interest rates are zero). 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, or invertible, MA(2) process: $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} = \Theta(L)\varepsilon_t$, we get $s_t = \varepsilon_t + \theta_1\varepsilon_{t-1} = \Theta_1(L)\varepsilon_t$, which 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 $|\theta| > 1/2$.

## 2.3 The limits of the Gaussian approach

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

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

where the roots of $\det(\Phi(L))$ are outside the unit circle and where the process $\eta_t$ is a Gaussian white noise, with $E(\eta_t) = 0$ and $V(\eta_t) = I$.

It is well-known that, in this case, $\Phi(L)$ is identified but $C$ is not, since replacing $C$ by $CQ$, where $Q$ is an orthogonal matrix, leaves the distribution of the process $Y_t$ unchanged. It is the *static identification problem*.

In order to solve this identification problem, additional short-run, long-run or sign restrictions

have been imposed in the literature [see e.g. the references in the introduction].[7] It turns out that if at most one of the components of $\eta_t$ is Gaussian, the identification problem disappears since $C$ is then identified, up to a permutation and a sign change 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 by Pseudo Maximum Likelihood techniques [see Gouriéroux, Monfort, and Renne (2017)]. The (quasi) identifiability of $C$ in the non-Gaussian case implies that, for instance, 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. However, the lack of correlation is not sufficient to define impulse response functions (IRFs) and their confidence intervals.[8]

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

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

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

In the Gaussian case, the distribution of the stationary process $Y_t$ depends on parameters through the second-order moments of the process or, equivalently, through the spectral density matrix:

$$f(\omega) = \frac{1}{2\pi}\Phi^{-1}(\exp i\omega)\Theta(\exp i\omega)CC'\Theta(\exp -i\omega)'\Phi^{-1}(\exp -i\omega)'. \tag{2.7}$$

Using the equalities $\Gamma(h) - \Phi_1\Gamma(h-1) - \cdots - \Phi_p\Gamma(h-p) = 0, \forall h \geq q+1$, with $\Gamma(h) = cov(Y_t, Y_{t-h})$, it is readily seen that the coefficient matrices $\Phi_1$, ..., $\Phi_p$ are identifiable from the distribution of the process $Y_t$ (Gaussian or not), but several sets of coefficients $(\Theta_1, \ldots, \Theta_q, C)$ yield the same spectral density and, therefore, 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 $\Theta(L)$ outside the unit circle– by use of the Blaschke matrices.[9] The lack of identification of $\Theta(L)$ is called the

[7] 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)]. This extended framework allows to test the identification restrictions. In this paper we will stay in a pure SVARMA framework.

[8] In most studies using SVAR models, the shocks are implicitly assumed Gaussian, but the Gaussian hypothesis is never tested in practice. This assumption is even explicit in some papers [see e.g. Forni, Gambetti, Lippi, and Sala (2017)].

[9] A Blaschke matrix is a square matrix of the lag operator $B(L)$ such that $[B(L)]^{-1} = B^*(L^{-1})$, where $B^*(.)$ is obtained from $B(.)$ by transposing and taking conjugate coefficients. See Leeper, Walker, and Yang (2013), p1123-

*dynamic identification problem*. We will see below that this second identification problem also disappears in the non-Gaussian case.

A simple example of the dynamic identification problem can be seen even 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 the process:

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

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 distribution for the process $y_t$. By contrast, we will see that if the $\eta_t$ are non-Gaussian, the distributions of the processes $y_t$ and $y_t^*$ are different, although their spectral density matrices are the same.

Another equivalent way to illustrate this situation is to consider the process $\eta_t^*$ defined through $y_t = \sigma\theta\eta_t^* - \sigma\eta_{t-1}^*$. In the Gaussian case (i.e. if $\eta_t$ and, therefore, $y_t$ are Gaussian), $\eta_t^*$ is also a standard Gaussian white noise; by contrast, in the non-Gaussian case, $\eta_t^*$ will be a weak white noise, but not a strong white noise, i.e. the $\eta_t^*$'s will be uncorrelated, but not independent.

Another problem with the usual Box-Jenkins approach is that the estimation of the parameters $\Phi_1$, ..., $\Phi_p$, $\Theta_1$, ..., $\Theta_q$, $\Sigma = CC'$ is based on a truncated VAR approximation relying on the assumption that $\Theta(L)$ is invertible (i.e. the roots of $\det(\Theta(L))$ are outside the unit circle), namely a truncation of $\Theta(L)^{-1}\Phi(L)Y_t = \varepsilon_t$, with $V(\varepsilon_t) = \Sigma = CC'$. In other words, a fundamental representation is a priori imposed without test.

# 3 Identification and Impulse Response Functions (IRFs) in the non-Gaussian SVARMA

## 3.1 Identification of the parameters

Let us consider again the SVARMA process:

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

1124 for a practical example of the use of Blaschke matrices.

with

$$\begin{aligned}\Phi(L) &= I - \Phi_1 L - \cdots - \Phi_p L^p,\\ \Theta(L) &= I - \Theta_1 L - \cdots - \Theta_q L^q,\end{aligned}$$

where the roots of $\Phi(L)$ are outside the unit circle, $\eta_t$ is a strong white noise, the components $\eta_{j,t}$ of $\eta_t$ are independent and such that $E(\eta_{j,t}) = 0$ and $V(\eta_{j,t}) = 1$.

We also make the following assumption on $\Theta(L)$:

**Assumption A.5.**

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

The previous assumption can also be written:

$$\det(\Theta(e^{i\omega})) \neq 0,\ \forall \omega \in ]-\pi, \pi].$$

In the simple univariate MA(1) case, $y_t = (1-\theta L)\sigma\eta_t$, this excludes the cases $\theta = \pm 1$. In the multivariate MA(1) case $y_t = (I - \Theta L)C\eta_t$, this excludes eigenvalues of $\Theta$ on the unit circle. However, the roots of $\det(\Theta(L))$ can be inside or outside the unit circle, and $\Theta(L)$ is invertible in a general sense, since there exists a two-sided series $B(L) = \sum_{k=-\infty}^{\infty} B_k L^k$ such that $B(L)\Theta(L) = I$ (see Appendix C).

Since $\Phi(L)$ is invertible, we have:

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

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

As mentioned above, we know that $\Phi(L)$ is identifiable. The question is the identification of $\Theta(L)$ and $C$. The proposition provided below, deduced from Theorem 1 in Chan, Ho, and Tong (2006) [based on Theorem 4 in Chan and Ho (2004)], solves both the static and the dynamic identification problems in the non-Gaussian case.[10,11] Let us first introduce two assumptions:

**Assumption A.6.** *The components of* $\eta_t$ *have the same distribution.*

[10] See Findley (1986), Cheng (1992) for the one-dimensional case $n = m = 1$.

[11] A similar identification result has been recently derived when the components of $\eta_t$ have fat tails [see Gouriéroux and Zakoian (2015)]. Note that the identification result in Chen, Choi, and Escanciano (2017), Theorem 1, is much less powerful. Indeed, while their result provides conditions to check if the fundamental representation is the right one, it cannot be used to find the correct non-fundamental representation if this is not the case.

**Assumption A.7.** *Each component of $\eta_t$ has a non-zero $r^{th}$ cumulant, with $r \geq 3$, and a finite moment of order $s$, where $s$ is an even integer greater than $r$.*

**Proposition.** *Under Assumptions A.1 to A.5, if $Y_t$ is non-Gaussian and if we consider $Y_t^*$ defined by*

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

*then, if either Assumption A.6 or Assumption A.7 is satisfied, the processes $Y_t$ 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 B.

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

The proposition in the previous section shows that, in the non-Gaussian case, $\Phi(L)$ and $\Theta(L)$ are identified and, therefore, $C\eta_t = \Theta^{-1}(L)\Phi(L)Y_t$ is identified too.

Since $C$ is identified up to a permutation and a sign change of its columns, the structural shocks $\eta_t$ are identified up to a permutation and a sign change of their components (not depending on $t$). It is just a labelling problem and the naming of each component will depend on the economic interpretation of the Impulse Response Function defined from the sequences $A_h = \Psi_h C$. More precisely, for a given choice of the order and of the sign of the columns of $C$, the differential impact on $Y_{i,t+h}$ of a unit shock on $\eta_{j,t}$ will be equal to

$$E\left(Y_{i,t+h}|\eta_{j,t} = 1, \underline{\eta_{t-1}}\right) - E\left(Y_{i,t+h}|\underline{\eta_{t-1}}\right),$$

where $\underline{\eta_t} = \{\eta_t, \eta_{t-1}, \dots\}$. Due to the linearity of the process, this impact is equal to $\Psi_{i,h}C^j$, where $\Psi_{i,h}$ is the $i^{th}$ row of $\Psi_h$ and $C^j$ is the $j^{th}$ column of $C$. It is important to note that the past values $\underline{\eta_{t-1}}$ do not appear in this expression and, therefore, the fact that these variables may not be functions of $\underline{y_{t-1}}$ is not a problem. For some $i$ and $j$, the function of $h$ defined by $\Psi_{i,h}C^j$ is called an impulse response function (IRF).

It is also obvious that the differential impact of $Y_{i,t+h}$ of a shock $\delta$ on $\eta_{j,t}$ is $\Psi_{i,h}C^j\delta$ and that this differential impact can also be obtained by computing recursively $Y_t$ from:

$$\Phi(L)Y_t = \Theta(L)C\eta_t, \tag{3.2}$$

by starting at $t$, setting all the past values of $Y_t$ at zero, all the past and future values of $\eta_t$ at 0 and all the components of $\eta_t$ at zero except $\eta_{j,t}$ which is set at $\delta$.

If the components of $\eta_t$ were only assumed to be non-correlated, but not necessarily independent, it would be impossible to assume that one component is shocked while the other components are left unchanged. For instance, if $\eta_{2,t} = (\eta_{1,t}^2 - 1)/\sqrt{2}$, where $\eta_{1,t} \sim \mathcal{N}(0,1)$, then $\eta_{1,t}$ and $\eta_{2,t}$ are uncorrelated, zero-mean and with unit variance, but $\eta_{2,t}$ is a deterministic function of $\eta_{1,t}$ and a shock $\delta$ on $\eta_{1,t}$ necessarily implies a shock $\delta^2/\sqrt{2}$ on $\eta_{2,t}$.

## 3.3 Stochastic shocks

The previous Impulse Response Function approach gives the average shift of the components of $Y_{t+h}$ when the mean of the distribution of $\eta_{j,t}$ is shifted by the quantity $\delta$. This practice has two limits. First, the other features of the distribution of $\eta_{j,t}$ like its variance, its skewness and its kurtosis are not changed. Second, only the modification of the conditional mean of $Y_{t+h}$ is considered but not the modification of its conditional distribution and, in particular, of its conditional quantiles, which might be interpreted as Values-at-Risk.

Let us now consider another type of transitory shock taking the form of a change in the distribution $f_j$ of $\eta_{j,t}$ into $g_j$. We can simulate the errors without and with this stochastic shock.[12] Without shock, the simulated errors $\eta_t^s, \ldots, \eta_{t+h}^s$ are drawn independently in $\otimes_k \hat{f}_k$, where $\hat{f}_k$ is an estimate of $f_k$, and with shock, they are drawn independently in $\hat{f}_1 \otimes \cdots \otimes \hat{f}_{j-1} \otimes g_j \otimes \hat{f}_{j+1} \otimes \cdots \otimes \hat{f}_n$ for $\eta_t^s$ and in $\otimes_k \hat{f}_k$ for $\eta_{t+1}^s, \ldots, \eta_{t+h}^s$. Then we can deduce simulated paths of $Y_t, \ldots, Y_{t+h}$ with and without shock by using recursively (3.2) with and without shock.

The only problem is to fix values for the past $\eta_t$, namely $\eta_{t-1}, \ldots, \eta_{t-q}$, which may not be computable from the past values of $Y_t$. A natural solution is to fix these values at their best linear approximation in terms of $Y_{t-1}, \ldots, Y_1$, which can be obtained from the Kalman filter applied to the linear state-space representation of the process $Z_t = (Y_t', \ldots, Y_{t-p}', \eta_t', \ldots, \eta_{t-q}')'$.

[12] This is called Stochastic Economic Scenario Generation (ESG) in the terminology of the new regulation for financial stability.

# 4 Estimation of Models with Non-Fundamentalness

In this section, we discuss parametric and semi-parametric estimation of non-fundamental SVARMA models.

The proposition of Section 3.1 suggests better semi-parametric estimation methods than the unconsistent Gaussian pseudo-maximum likelihood used in the standard Box-Jenkins methodology. These alternative methods provide consistent estimators of the true one-sided moving average polynomial $A(L)$ or, equivalently, of the true $(\Phi(L),\Theta(L))$ in a VARMA representation where the roots of $\det(\Theta(L))$ may be inside or outside the unit circle. They also provide consistent non-parametric estimators of the distribution of the components of the error term which are required for the derivation of the impulse response functions (IRFs) associated with stochastic shocks (see Section 3.3). The idea is to introduce appropriate moment restrictions deduced from the independence assumption on the components of error $\eta$ and derive associated moment methods for estimation. These semi-parametric approaches are robust, but can be weakly efficient especially in finite samples. We also introduce parametric likelihood-based estimation methods.

## 4.1 Semi-parametric estimation of a non-fundamental SVARMA model

For expository purpose, let us consider a SVARMA(1,1) model:

$$Y_t = \Phi Y_{t-1} + C_0\eta_t + C_1\eta_{t-1}, \tag{4.1}$$

where the components of $\eta_t$ are both serially and cross-sectionally independent with $E(\eta_t)=0$, $V(\eta_t)=I$ (for the sake of notational simplicity, we replace $C$ by $C_0$ and $-\Theta_1 C$ by $C_1$). We assume that the roots of the determinant of the autoregressive polynomial are well-located, 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 $\eta_{j,t}$, $t=1,\ldots,T$. We have to consistently estimate the (true values of) parameters $\Phi$, $C_0$, $C_1$ as well as the (true) functional parameters $f_j$, $j=1,\ldots,n$.[13]

**i) Pure moving average process.**

Let us first consider the case $\Phi=0$, that is, a pure moving average process, and focus on the estimation of the moving average matrix coefficients $C_0$, $C_1$. We know that $C_0$ and $C_1$ are locally identifiable in the non-Gaussian case; more precisely, they are identifiable up to a same permutation and sign change of their columns, but $\Theta=-C_1C_0^{-1}$ is fully identifiable.

[13] For expository purpose, we do not distinguish the generic parameters from their true values in the notations.

The Laplace transform of $Y_t$, $Y_{t-1}$ is:

$$\begin{aligned}
& E[\exp(u'Y_t + v'Y_{t-1})] \\
= \; & E\{\exp[u'(C_0\eta_t + C_1\eta_{t-1}) + v'(C_0\eta_{t-1} + C_1\eta_{t-2})]\} \\
= \; & E[\exp(u'C_0\eta_t)] \times E\{\exp[(u'C_1 + v'C_0)\eta_{t-1}]\} \times E[\exp(v'C_1)\eta_{t-2}] \\
= \; & \Pi_{j=1}^n E[\exp(u'C_{0j}\eta_{j,t})] \times \Pi_{j=1}^n E\{\exp[(u'C_{1j} + v'C_{0j})\eta_{j,t-1}\} \times \Pi_{j=1}^n E[\exp(v'C_{1j}\eta_{j,t-2})],
\end{aligned}$$

by using the independence assumptions.

The expression of this joint Laplace transform can be used to compute the expressions of the first, second, third (fourth) cross-moments of $Y_t$ as functions of $C_0$, $C_1$ and of the first, second, third (fourth) marginal moments of the errors $\eta_{j,t}$.[14]

By focusing on the first and second moments order only, we know that the corresponding GMM estimators will not provide consistent results (see the discussion in Section 2.3). The identification of parameters and then the consistency of associated moment methods are generally achieved if we also consider higher-order moments.[15]

For instance, let us consider moments up to order 3, with data preliminary demeaned; the "observable" second and third order moments are:

$$\begin{array}{lll}
E(y_{j,t}y_{k,t}), & j,k = 1,\ldots,n, \; j \leq k \\
E(y_{j,t}^3), & j = 1,\ldots,n, \\
E(y_{j,t}^2 y_{k,t-1}), & j,k = 1,\ldots,n, \\
E(y_{j,t}y_{k,t-1}^2), & j,k = 1,\ldots,n.
\end{array}$$

Thus we have $n(n+1)/2 + n + 2n^2$ observable moments.

The number of unknown parameters to estimate are the elements of $C_0$, $C_1$ and the third-order moments of the $\eta_{j,t}$, $j = 1,\ldots,n$ (since their first and second-order moments are already known and their cross third-order moments are equal to zero because of the cross-sectional independence). This number of unknown parameters is therefore $n + 2n^2$ and the order condition for identification, $n(n+1)/2 + n + 2n^2 > n + 2n^2$, is satisfied. The rank condition is more difficult to analyze [see this discussion in Gospodinov and Ng (2015), Section 2 in the one-dimensional case].

[14] When their moments exist. If the errors have fat tails, the expression of the Laplace transform can be used for pure imaginary arguments $u$ and $v$.

[15] See Giannakis and Mendel (1989), Friedlander and Porat (1990), Na, Kim, Song, and Kim (1995) for the development of this approach in signal processing in the one-dimensional case and the discussion in Appendix A.

**ii) The general case**

Let us now consider the general specification (4.1). Since the process is causal –the eigenvalues of $\Phi$ are inside the unit circle–, $\eta_t$ and $\eta_{t-1}$ are independent from $Y_{t-2}$. Thus, under the stationarity assumption A.4, we can estimate the autoregressive matrix coefficient by regressing $Y_t$ on $Y_{t-1}$, with the instruments $Y_{t-2}$. The corresponding instrumental variable (IV) estimator of $\Phi$ is:

$$\hat{\Phi} = (\Sigma_t Y_t Y_{t-2}')(\Sigma_t Y_{t-1} Y_{t-2}')^{-1}. \tag{4.2}$$

Once $\Phi$ has been estimated, the associated IV residuals:

$$\hat{V}_t \equiv Y_t - \hat{\Phi} Y_{t-1}, \tag{4.3}$$

are consistent approximations of $V_t = C_0 \eta_t + C_1 \eta_{t-1}$.

Then in a second-step we can apply to observations $\hat{V}_t$ the estimation method for pure MA process introduced in the subsection above and deduce consistent estimates of $C_0$ and $C_1$.[16]

**Example:** In the one-dimensional case:

$$y_t = \varphi y_{t-1} + c\eta_t - c\theta\eta_{t-1}, \text{ say.}$$

These moment estimators of the parameters are:

$$\begin{aligned}
\hat{\varphi} &= (\Sigma_t y_t y_{t-2})/(\Sigma_t y_{t-1} y_{t-2}), \\
\hat{\theta} &= -\Sigma_t[(y_t - \hat{\varphi} y_{t-1})^2 (y_{t-1} - \hat{\varphi} y_{t-2})]/\Sigma_t[(y_t - \hat{\varphi} y_{t-1})(y_{t-1} - \hat{\varphi} y_{t-2})^2] \\
&\quad \text{[see Appendix A.3]}, \\
\hat{c} &= \frac{1}{T}\Sigma_t \hat{\eta}_t^{*2}, \text{ where } \hat{\eta}_t^* = \frac{1 - \hat{\varphi} L}{1 - \hat{\theta} L} y_t,
\end{aligned}$$

and the inverse $(1 - \hat{\theta} L)^{-1}$ is computed by a backward expansion if $|\hat{\theta}| < 1$, by a forward expansion otherwise.[17]

[16] If $C_1 = 0$, $C_0$ can be directly estimated by ICA [see e.g. Chen, Choi, and Escanciano (2017), or Gouriéroux, Monfort, and Renne (2017)].

[17] In practice, the expansions are truncated to account for the finite number of observed values of $y$.

**iii) Nonparametric estimation of the error distribution**

Once $\Phi$, $C_0$ and $C_1$ have been estimated, we deduce consistent approximations of the errors:

$$\hat{\eta}_t = (\hat{C}_0 + \hat{C}_1 L)^{-1}(I - \hat{\Phi}L)Y_t, \tag{4.4}$$

where the inverse (in a general sense) $(\hat{C}_0 + \hat{C}_1 L)^{-1}$ is a two-sided series which has to be computed carefully, with backward expansions for roots larger than one, forward expansions otherwise (see Appendix C).

Then the p.d.f. $f_j$ can be estimated by a kernel density estimator applied to residuals $\hat{\eta}_{j,t}$, $t = 1, \ldots, T$.

## 4.2 Maximum Likelihood, Composite Likelihood and Simulated Composite Likelihood Approaches for Parametric SVARMA Models

To introduce these parametric estimation approaches, let us first discuss the case of a one-dimensional MA(1) process before considering the general framework of a SVARMA process.

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

We consider the MA(1) process:

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1}, \tag{4.5}$$

where the $\varepsilon_t's$ are independent.

Suppose that we observe $\{y_1, \ldots, y_T\}$. If the common distribution of the $\varepsilon_t's$ is $N(0, \sigma^2)$, we have seen in Section 2.3 that the model is not identifiable. If $\varepsilon_t$ is not Gaussian, the proposition of Section 3.1 shows that the model is identifiable.[18] 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 (4.5) in the standard way in order to get $\varepsilon_t$ as a function of current and lagged values of process $Y$ as:

$$\varepsilon_t = \sum_{h=0}^{\infty} \theta^h y_{t-h}. \tag{4.6}$$

[18] See Appendix A for a more detailed discussion of non-identifiability of a MA(1) process and the links with invertibility.

Then the log-likelihood function is approximated by:

$$L_1^a(\theta,\gamma) = \sum_{t=1}^{T} \log g\left(\sum_{h=0}^{t-1} \theta^h y_{t-h};\gamma\right), \tag{4.7}$$

where the infinite sums are truncated to be compatible with the observed $y_1,\ldots,y_T$.

**ii)** When $|\theta|>1$, equation (4.5) can still be inverted, but in reverse time. We get:

$$\begin{aligned} y_t &= \varepsilon_t - \theta\varepsilon_{t-1} \\ \Leftrightarrow \quad -\frac{y_{t+1}}{\theta} &= \varepsilon_t - \frac{1}{\theta}\varepsilon_{t+1} \\ \Leftrightarrow \quad \varepsilon_t &= -\sum_{h=0}^{\infty} \frac{1}{\theta^{h+1}} y_{t+h+1}. \end{aligned} \tag{4.8}$$

The log-likelihood function is approximated by:

$$L_2^a(\theta,\gamma) = \sum_{t=1}^{T} \log\left\{\frac{1}{|\theta|} g\left(-\sum_{h=0}^{T-t-1} \frac{1}{\theta^{h+1}} y_{t+h+1};\gamma\right)\right\}, \tag{4.9}$$

where the sums are now truncated to account for the most recent observations and the factor $1/|\theta|$ comes from the Jacobian formula.

**iii)** Let us now discuss the case $\theta=1$. Focussing on the regimes when approximating the log-likelihood function gives the misleading impression of a lack of continuity of the exact log-likelihood function w.r.t. $\theta$ at $|\theta|=1$. This exact log-likelihood is however continuous.[19] Indeed, we have:

$$\begin{aligned} \varepsilon_1 &= y_1 + \theta\varepsilon_0, \\ \varepsilon_2 &= y_2 + \theta y_1 + \theta^2\varepsilon_0, \\ &\ldots \\ \varepsilon_T &= y_T + \theta y_{T-1} + \ldots + \theta^{T-1} y_1 + \theta^T \varepsilon_0. \end{aligned}$$

Thus the joint p.d.f. of $\{y_1,\ldots,y_T\}$ given $\varepsilon_0$ is:

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

[19] Such 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.

and the exact log-likelihood is:

$$L(\theta,\gamma) = \log\left\{\int \Pi_{t=1}^{T} g\left(\sum_{h=0}^{t-1}\theta^h y_{t-h} + \theta^t \varepsilon;\gamma\right) g(\varepsilon;\gamma)d\varepsilon\right\}. \tag{4.10}$$

Hence, the exact log-likelihood is generally a differentiable function of $\theta$. By contrast, an approximated value of the log-likelihood function, given by:[20]

$$L^a(\theta,\gamma) = L_1^a(\theta,\gamma)\mathbb{1}_{|\theta|<1} + L_2^a(\theta,\gamma)\mathbb{1}_{|\theta|\geq 1}, \tag{4.11}$$

is only right-differentiable at $\theta = 1$.[21] In practice, however, using the approximated log-likelihood (4.11) is easier because it does not involve the computation of an integral as in the case of the exact log-likelihood (4.10).

To conclude, in the simple MA(1) case, maximum likelihood estimation can be conducted by maximising the approximated log-likelihood function (4.11). If $|\theta| \neq 1$, the standard asymptotic theory applies. Nevertheless, this is not the case if $|\theta| = 1$. In Subsection 4.2.3, we will discuss alternative approaches based on optimization criteria that are regular in the neighbourhood of $|\theta| = 1$.

### 4.2.2 The Maximum Likelihood approach in the VARMA context

For expository purpose, let us consider the VARMA(1,1) model:

$$Y_t - \Phi Y_{t-1} = \varepsilon_t - \Theta\varepsilon_{t-1}, \tag{4.12}$$

where the errors $\varepsilon_t$ are given by:

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

where the $\eta_t$ are serially and mutually independent, with $E(\eta_t) = 0$ and $V(\eta_t) = I$. The distribution of the $\eta_t$ is non-Gaussian and parameterized with $\gamma$, say. Therefore, the p.d.f. of the errors $\varepsilon_t$ is of the form $g(\varepsilon,\Gamma)$, with $\Gamma = (C,\gamma)$.[22] The errors $\eta_t$ are supposed to satisfy Assumption A.6 or Assumption A.7.

[20] The approximation is due to the truncations of lags and leads.

[21] An other approximated value of the log-likelihood function is $L_1^a(\theta,\gamma)\mathbb{1}_{|\theta|\leq 1} + L_2^a(\theta,\gamma)\mathbb{1}_{|\theta|>1}$. The latter function is left-differentiable at $\theta = 1$.

[22] Such a specification for the distribution of the errors $\varepsilon_t$ notably excludes that it belongs to the standard multivariate Student family. This assumption is indeed not appropriate for the analysis of impulse responses since this family does not include the case of independent components. We can assume, instead, that the $i^{th}$ component of $\eta_t$ follows a normalized univariate Student distribution with $\nu(i)$ degrees of freedom. Additional identification restrictions can be introduced to fix the denomination of the errors, i.e. to solve the problem of multiplicity by change of scale and permutations. For instance, permutation is ruled out if we impose either that the degrees of freedom $\nu(i)$ are in an

The roots of $\det(\Phi(z))$ are assumed to be outside the unit circle, but the roots of $\det(\Theta(z))$ can be anywhere: inside, outside, or even on the unit circle.

In order to get an approximated value of the likelihood function, one needs to recover, for a given parameterization of the model, some estimated values of the errors $\varepsilon_t$. To do this, as explained in Appendix C, one can exploit the real Schur decomposition of $\Theta$:[23]

$$\Theta = AUA' = A \begin{bmatrix} U_1 & U_{1,2} & \dots & & U_{1,K} \\ 0 & U_2 & U_{2,3} & \dots & U_{2,K} \\ \vdots & \ddots & \ddots & & \vdots \\ & & 0 & U_{K-1} & U_{K-1,K} \\ 0 & \dots & & 0 & U_K \end{bmatrix} A', \tag{4.13}$$

where $A$ is orthogonal, and $U$ is upper block-triangular, where the diagonal blocks ($U_k$, $k \in \{1,\dots,K\}$) are either $1\times 1$ or $2\times 2$ blocks, the $2\times 2$ blocks corresponding to complex conjugate complex eigenvalues of $\Theta$.

Left-multiplying $Y_t - \Phi Y_{t-1} = \varepsilon_t - \Theta\varepsilon_{t-1}$ by $A^{-1} = A'$, we get:

$$W_t = \varepsilon_t^* - U\varepsilon_{t-1}^*, \tag{4.14}$$

where $W_t = A'(I-\Phi L)Y_t$ and $\varepsilon_t^* = A'\varepsilon_t$.

Appendix C shows that the different components of $\varepsilon_t^*$ can be recovered from the $V_t$s –and therefore from the $Y_t$s– by means of infinite backward or forward expansions of the elements of $W_t$. The use of backward versus forward expansions depends on on the position of the eigenvalues of the $U_k$ matrices with respect to the unit circle. When we observe $\{y_1,\dots,y_T\}$ only, the $\varepsilon_t^*$s can only be approximated by truncated backward or forward expansions (as in equations (4.7) and (4.9)). Denoting by $\hat{\varepsilon}_t$ the resulting estimate of $\varepsilon_t$, Appendix D shows that we can approximate the log-likelihood with:

$$L^a(\Phi,\Theta,\Gamma) = -T\sum_{k=1}^{K}\log|\det(U_k)|\mathbb{1}_{|\det(U_k)|\geq 1} + \sum_{t=1}^{T-1}\log g(\hat{\varepsilon}_t,\Gamma). \tag{4.15}$$

Note that the first term appearing on the right-hand side of the previous equation is equal to $T$ times the opposite of the sum of the logarithms of the moduli of the eigenvalues of $\Theta$ whose modulus is

increasing order, or that the elements of the first row of matrix $C$ are in an increasing order. The second condition is preferable since it is compatible with the limiting case of equal degrees of freedom.

[23] 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.

larger than one. This term therefore does not depend on the (Schur) decomposition of matrix $\Theta$. The extension of the computation of the approximated log-likelihood function in the case where the order of $\Phi(L)$ is larger than 1 is straightforward. Besides, Appendix C explains how to recover the shocks $\varepsilon_t$ if the order of $\Theta(L)$ is larger than 1. Hence, the computation of the approximated log-likelihood function can be performed in the general VARMA($p$,$q$) case.

**Remark 1:** As in the MA(1) case (Subsection 4.2.1), the asymptotic theories apply in the context of the maximisation of (4.15) only if all the roots of $\Theta(L)$ are not on the unit circle.

**Remark 2:** Because of the discontinuity of the approximated likelihood function when the roots of $\det(\Theta(L))$ are on the unit circle, the numerical maximisation of the approximated likelihood may tend to result in (local) optima with parameters corresponding to the same fundamentalness/non-fundamentalness regime as the one used to initialize the numerical optimization procedure. To address this potential problem, one should launch the numerical optimization from initial conditions reflecting different possible fundamentalness regimes. Another complementary approach consists in running additional numerical optimizations with starting values corresponding to models featuring the same spectral density as a the one resulting from a preliminary-estimated model, but with different fundamentalness regimes; such models can be obtained by applying Blaschke-based transformations to the preliminary-estimated model (see Lippi and Reichlin (1994)).

The likelihood-based approaches proposed in the next subsection are based on optimization criteria decomposed into sums which are appropriate for applying standard asymptotic theory even if the roots of $\Theta(L)$ lie on the unit circle. These criteria do not suffer from discontinuity on the unit circle. However, as illustrated in the application section –in particular in Subsection 5.1–, these approaches may be substantially less efficient than the maximum likelihood approach.

### 4.2.3 Incomplete and Composite Maximum Likelihood approaches

The principle of Incomplete Maximum Likelihood (IML) is easily explained for the MA(1) process discussed above. Let us separate the observations by omitting one out of three observations. The set of observations becomes:

$$y_1,\ y_2,\ y_4,\ y_5,\ \ldots,\ y_{3j-1},\ y_{3j-2},\ \ldots$$

There is a loss of information since observations $y_3$, $y_6$, ..., $y_{3j}$, ... are not taken into account with the advantage that the pairs of observations $(y_{3j-1}, y_{3j-2})$, $j$ varying, are i.i.d.. Thus the

exact log-likelihood function corresponding to these incomplete observations is easily computed, naturally expressed as a sum and the standard asymptotic theory applies.

Let us now describe how the incomplete likelihood approach can be adapted to the multidimensional framework. Let us come back to the VARMA(1,1) model defined by equation (4.12). This model is a Seemingly Unrelated Regression (SUR) model:

$$Y_t = \Phi Y_{t-1} + v_t, \tag{4.16}$$

and the autoregressive matrix $\Phi$ can be estimated through instrumental variables (IV), using as instruments $Y_{t-2}$ ($Y_{t-3}, Y_{t-4}, \ldots$), which are uncorrelated with $v_t$. Let us denote by $\hat{\Phi}$ the corresponding IV estimator.

Next, let us consider the joint p.d.f. of $(v_t, v_{t-1})$, where $v_t = \varepsilon_t - \Theta\varepsilon_{t-1}$ and $v_{t-1} = \varepsilon_{t-1} - \Theta\varepsilon_{t-2}$. Because we have $\varepsilon_t = v_t + \Theta v_{t-1} + \Theta^2\varepsilon_{t-2}$ and $\varepsilon_{t-1} = v_{t-1} + \Theta\varepsilon_{t-2}$, this joint distribution is:

$$h(v_t, v_{t-1}; \Theta, \Gamma) = \int g(v_t + \Theta v_{t-1} + \Theta^2\varepsilon; \Gamma) g(v_{t-1} + \Theta\varepsilon; \Gamma) g(\varepsilon; \Gamma) d\varepsilon. \tag{4.17}$$

Hence, the two-step IML estimator of $(\Theta, \Gamma)$ is the solution of:

$$(\hat{\Theta}, \hat{\Gamma}) = \arg\max_{\Theta,\Gamma} \sum_{j=1}^{J=(T/3)} \log h(Y_{3j-1} - \hat{\Phi}Y_{3j-2}, Y_{3j-2} - \hat{\Phi}Y_{3j-3}; \Theta, \Gamma). \tag{4.18}$$

The two-step IML estimator has standard asymptotic properties, irrespective of the location of the roots of $\det(\Theta(z))$. It is in particular consistent, asymptotically normal, and its asymptotic variance-covariance matrix can be derived. All the observations of $(y_t)$ are used if $\Phi \neq 0$, but in a non-optimal way.

Other consistent estimators as simple to implement and using observations in a more efficient way can be based on the same idea. We can in particular consider the estimator solution of:

$$(\tilde{\Theta}, \tilde{\Gamma}) = \arg\max_{\Theta,\Gamma} \sum_{t=2}^{T} \log h(Y_t - \hat{\Phi}Y_{t-1}, Y_{t-1} - \hat{\Phi}Y_{t-2}; \Theta, \Gamma). \tag{4.19}$$

This two-step Composite Maximum Likelihood (CML) estimator uses the information on all the $v_t$'s. It is also consistent, asymptotically normal, but the asymptotic variance-covariance matrix is now computed by a sandwich formula involving a general central limit theorem [Varin, Reid, and Firth (2011), Gouriéroux and Monfort (2016)].

The composite likelihood function appearing on the right-hand side of (4.19) depends on integrals of the same dimension as the VARMA system. We can approximate the integral in function

$h$ by simulation to get a two-step simulated composite likelihood. The numerical optimization of the approximative criterion is:

$$(\Theta^*,\Gamma^*) = \arg\max_{\Theta,\Gamma} \sum_{t=1}^{T} \log \hat{h}(Y_t - \hat{\Phi}Y_{t-1}, Y_{t-1} - \hat{\Phi}Y_{t-2}; \Theta, \Gamma), \tag{4.20}$$

where:

$$\hat{h}(v_t, v_{t-1}; \Theta, \Gamma) = \frac{1}{S}\sum_{s=1}^{S} \{g(v_t + \Theta v_{t-1} + \Theta^2 \varepsilon^s(\Gamma), g(v_{t-1} + \Theta\varepsilon^s(\Gamma); \Gamma)\}, \tag{4.21}$$

and the $\varepsilon^s(\Gamma)$ are drawn independently in distribution $g(\varepsilon;\Gamma)$.[24]

When the number $S$ of simulations tends to infinity sufficiently fast w.r.t. the number $T$ of observations, this simulation-based estimator has the same asymptotic properties as the CML estimator itself [see Gouriéroux and Monfort (1996) for a general presentation of simulation-based estimation methods].

**Remark 3:** The composite maximum likelihood estimation approaches provide subefficient consistent estimators. As mentioned above, an advantage of the CML approaches is however that they are not affected by discontinuity problems when the roots of $\det(\Theta(L))$ are on the unit circle (contrary to the ML approach presented in Subsection 4.2.2). This may in particular make the numerical optimization of the CML criteria less sensitive to the choice of the initial conditions. The IV-CML approach presented above can therefore be used as a preliminary step, providing a consistent estimator $(\hat{\Phi},\hat{\Theta},\hat{\Gamma})$ of $(\Phi,\Theta,\Gamma)$ and of the true regime of (potential) non-fundamentalness. The IV-CML estimators $(\hat{\Phi},\hat{\Theta},\hat{\Gamma})$ can then be used as starting values in the numerical optimization of the approximated log-likelihood $L_s^a(\Phi,\Theta,\Gamma)$ given in equation (4.15).

**Remark 4:** Why not consider a standard simulated maximum likelihood (SML) approach? The SML estimators would be defined as:

$$(\hat{\hat{\Phi}},\hat{\hat{\Theta}},\hat{\hat{\Gamma}}) = \arg\max_{\Phi,\Theta,\Gamma} \log \left\{ \frac{1}{S}\sum_{s=1}^{S} \Pi_{t=1}^{T} g\left( \sum_{h=0}^{t-1} \Theta^h (Y_{t-h} - \Phi Y_{t-h-1}) + \Theta^t \varepsilon^s(\Gamma); \Gamma \right) \right\}.$$

However, the expression of the objective function has been derived backward, which implies terms like $\Theta^t\varepsilon^s(\Gamma)$. If the true representation is non-fundamental and $T$ is rather large, $\Theta^t$ will have exploding components when $t$ is increasing. Thus the SML estimator will be very sensitive to drawings of simulated $\varepsilon$ in the tail and not robust. The IML and CML approaches do not have

[24] As usual the same basic drawings must be kept when $\Gamma$ is modified in the optimization algorithm.

this drawback.

**Remark 5:** Simulated Method of Moment have also been suggested [see Gospodinov and Ng (2015) in the one dimensional case]. In the parametric framework they are in general less efficient then the composite likelihood approach.

# 5 Applications

## 5.1 Monte Carlo exercises

This subsection illustrates the performances of estimation approaches by means of Monte-Carlo experiments. We focus on the maximum likelihood (ML) and on the composite maximum likelihood (CML) approaches presented in Subsections 4.2.2 and 4.2.3, respectively.

For the sake of simplicity, we focus on a univariate MA(1) processes:

$$y_t = \varepsilon_t - \theta\varepsilon_{t-1}, \tag{4.22}$$

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

We consider different sample sizes ($T = 100$, 300 and 1000) and different types of (true) distributions of the errors $\varepsilon_t$. Four distributions are used: the Gaussian distribution, a Gaussian mixture distribution and two Student distributions with respective degrees of freedom of 5 and 10. In all simulations, we use $\theta = -2$. Hence, the data generating processes are non-fundamental.

In order to get the intuition behind the approach, it is instructive to look at the joint distributions of $y_t$ and $y_{t-1}$. Figure 1 displays contour plots associated with these distributions in the context of the four different types of distribution 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 $V(\varepsilon_t) = \theta^2 = 1/4$. In the purely Gaussian case (Panel (a)), the two distributions coincide, reflecting the fact that 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 $\mathcal{N}(0,\sigma_1^2)$ with probability $p$ and from $\mathcal{N}(0,\sigma_2^2)$ with probability $1-p$. In order to to have $V(\varepsilon_t) = 1$, we set $\sigma_1 = 5 = 10 \times \sigma_2$, which implies $p \approx 3\%$. This distribution depicts a situation in which very large shocks $\varepsilon_t$ may occur at each period, but with a relatively small probability (3%). Assume that, at date $t-1$, $\varepsilon_{t-1}$ is drawn from the large-variance distribution $\mathcal{N}(0,\sigma_1^2)$. Because $y_t = \varepsilon_t - \theta\varepsilon_{t-1}$, both $y_{t-1}$ and $y_t$ are then likely to take particularly large absolute values. If $\theta = -2$ (respectively $\theta = -1/2$) then, with a high probability, we will have $|y_t| > |y_{t-1}|$ (respectively $|y_t| < |y_{t-1}|$). Such

large drawings of $\varepsilon_t$ account for the distribution tails indicated with the letters A and B in Panel b of Figure 1. The fact that the tails associated with the fundamental and the non-fundamental processes are not located at the same place reflects the non-equivalence of the two processes.

Figure 2 shows the distributions of the estimators of $\theta$ resulting from both the CML and the ML approaches. For each simulated sample, we estimate the two model parameters, i.e. $\theta$ and the variance of $\varepsilon_t$. We however focus our discussion on the estimates of $\theta$. Each of the four rows of plots corresponds to one of the four considered distributions for $\varepsilon_t$. On each panel, the three curves correspond to the three considered sample sizes ($T = 100$, 300 and 1000). The distributions are often bimodal; one mode being generally close to the true value of $\theta$ –indicated by a vertical bar on each panel– and the other being close to $1/\theta$. In the Gaussian case (Panels a.1 and a.2), the first mode is closer to the fundamental value $1/\theta$ ($= -1/2$). In most of the other cases, the first mode is closer to the true value of $\theta$. In the Gaussian mixture case (Panels b.1 and b.2), and for longer samples, the estimator distributions are unimodal; this is consistent with the fact that, in the Gaussian mixture case, the differences between the distributions of $(y_{t-1}, y_t)$ in the fundamental and non-fundamental regimes are more marked than for the other distributions (as illustrated by Figure 1).

Table 1 reports summary statistics associated with the different estimators. Importantly, the results show that the CML approach is substantially less efficient than the ML one. Indeed, Root Mean Squared Errors (RMSEs) are systematically lower in the ML case. These lower RMSEs reflect both lower biases and lower standard deviations of the estimator distributions for the ML approach.

The last three columns of Table 1 are aimed at assessing the validity of the asymptotic distribution of the $\theta$ estimator. Specifically, they 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 $P(-\phi_\alpha < X < \phi_\alpha) = \alpha$, if $X \sim \mathcal{N}(0,1)$. The closer to $\alpha$ the reported fractions, the more satisfying is the approximation of the estimator distribution based on asymptotic theory. The results indicate that the inference based on the estimated asymptotic distribution is less adequate in the CML context than in the ML one. Both the biases and the underestimation of the estimator standard deviation (compare column $\overline{\sigma_{asy}}$ with column $S.D.$) explain the poor performance of the asymptotic formula in the CML case.

## 5.2 Univariate real-data example: per capita GDP growth rates

In this subsection, we use the Maximum Likelihood approach to estimate the parameterizations of ARMA(1) processes assumed to be followed by per capita real GDP growth. We consider long

historical samples taken from Bolt and van Zanden (2014).[25] As indicated in the second column of Table 2, the data, which are at the annual frequency, start as soon as 1800 for several countries. For most samples, the non-Gaussianity of the data is confirmed by the application of the normality test of Bai and Ng (2005), which is a generalization of the Bera and Jarque (1981) test to time series data. The p-values of this test are reported in the third column of the table. For the growth rates of all countries but two of them –France and Ireland–, the null hypothesis of normality is rejected at the 10% level.

Denoting by $y_t$ the demeaned per capita GDP growth rate, the model is as follows:

$$y_t = \phi y_{t-1} + c\eta_t - \theta c\eta_{t-1},$$

where the distribution of $\eta_t$ is a Gaussian mixture. More precisely, we assume that $\eta_t$ is drawn from $\mathcal{N}(\mu_1, \sigma_1^2)$ with a probability $p$ and from $\mathcal{N}(\mu_2, \sigma_2^2)$ with a probability $1-p$. Therefore, if we have $E(\eta_t) = 0$ and $V(\eta_t) = 1$, then the distribution of $\eta_t$ is completely defined by $\gamma = [\mu_1, \sigma_1, p]'$.

The results are reported in Table 2. For 10 countries out of 17, the absolute value of the estimate of $\theta$ is larger than one. That is, for these countries, the estimated MA process is non-fundamental.

## 5.3 Bivariate real-data example: GDP growth and unemployment

In this subsection, we consider the two-variable model of Blanchard and Quah (1989), referred to as BQ hereafter. The two endogenous variables are the U.S. real GDP growth and the unemployment rate.[26] BQ fit an 8-lag VAR model to these data for the period from 1948Q2 to 1987Q4 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 GDP. 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 analysing 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 lag polynomial associated with the 8-lag VAR 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 (non-fundamental ones and the fundamental one). Nevertheless, as explained in Section 3.1, if the

[25] The data are available at http://www.ggdc.net/maddison/maddison-project/home.htm.

[26] Our data are extracted from the FRED economic database of the Federal Reserve Bank of St Louis. GDP growth rates are computed as the first differences of the natural logarithms of real GDP. Following Blanchard and Quah (1989), we remove a linear trend from the unemployment rate series.

underlying structural shocks are non-Gaussian and independent, then the data-generating VARMA process, be it fundamental or not, is identifiable.

In our empirical analysis, we consider two sample periods: 1948Q2-1982Q3 and 1982Q4-2016Q4. These two sample periods are of equal sizes, but correspond to two substantially different periods.[27] For each of the two samples, we estimate a VARMA(1,1) model where the two endogenous variables are the log real GDP growth and the unemployment rate, as in Blanchard and Quah (1989). The estimation is carried out by maximizing the approximated likelihood function (see Subsection 4.2.2). The (true) distributions of the independent shocks $\eta_{j,t}$, for $j \in \{1,2\}$, are assumed to be Gaussian mixtures. Specifically, we assume that $\eta_{j,t}$ is drawn from $\mathcal{N}(\mu_{j,1}, \sigma^2_{j,1})$ with a probability $p = 0.5$ and from $\mathcal{N}(\mu_{j,2}, \sigma^2_{j,2})$ with a probability $1-p = 0.5$. Therefore, if we have $E(\eta_{j,t}) = 0$ and $V(\eta_{j,t}) = 1$, then the distribution of the two-dimensional vector $\eta_t$ is completely defined by $\gamma = [\mu_{1,1}, \mu_{2,1}, \sigma_{1,1}, \sigma_{2,1}]'$. All in all, 16 parameters have to be estimated.[28,29]

Parameter estimates are reported in Panel (a) of Table 3. Panel (b) shows the absolute values of the eigenvalues of $\Theta$, that are the inverses of the roots of $\det(I - \Theta L)$. For both samples, the two roots of $\det(I - \Theta L)$ lie on each side of the unit circle. Hence, the estimated processes are non-fundamental.

Figure 3 shows the estimated parametric distributions of the structural shocks (Gaussian mixtures, represented by dotted lines). Though each of them depends on two parameters only, these estimated distributions are fairly close to kernel density estimates associated with recovered $\eta_{j,t}$'s (black solid lines).

Figure 4 displays the impulse response functions resulting from our approach and compares them with those obtained with long-run restrictions *à la* Blanchard and Quah (1989).[30] Because the ML approach does not rely on restrictions based on economic theory, the two estimated structural shocks have no a priori economic interpretation. However, for the sake of comparing the two approaches, we will bring the BQ supply shock closer to the ML-estimated shock that accounts for the largest part of the GDP long-run variance. We will call this shock the "long-run shock". Figure 4 shows substantial differences between the two types of approaches (BQ and ML). More precisely, while the patterns of the different IRFs resulting from the two approaches show similarities, the scales of the responses are significantly different. Moreover, in the ML approach, both structural shocks –Shock 1 (first row of charts) and Shock 2 (second row of charts)– have

[27] For instance, the standard deviation of quarterly GDP growth rates is far lower in the more recent period than for the older (0.63% against 1.12%).

[28] Setting the mixing parameter $p$ to 0.5 facilitates the convergence of the numerical algorithm while allowing for a large variety of distribution shapes, as illustrated by Figure 3.

[29] Four parameters for $\Phi(L)$, four for $\Theta(L)$, four for $C$ and four for $\gamma$. Because the data are demeaned, no intercept is included in the VARMA specification.

[30] We consider here the IRFs that correspond to deterministic shocks on the components of the vector of errors (see Subsection 3.2).

a long-run impact on GDP. This is also illustrated by variance decompositions whose results are reported in Table 4. This table gives the share of variance accounted for by Shock 1, that is the demand shock in the BQ approach: while this share mechanically goes to 0 with the BQ approach and for GDP, such a restriction is not imposed in the ML case. For instance, the share of variance accounted for by Shock 1 is of 12% at the 100-year horizon for the older sample and for the ML approach. The fact that none of the two structural shocks identified within a bivariate VAR has a zero impact on GDP in the long-run was also obtained by, e.g., Cochrane (1994) or Gali (1999). For both samples, the long-run shock –i.e. Shock 2– explains a larger share of unemployment fluctuations with the BQ than with the ML approach.

# 6 Concluding remarks

We have shown in this paper that the static and dynamic identification issues encountered in the analysis of SVARMA models are due to the poor performance of the estimation method used in the Box-Jenkins methodology, namely the Gaussian pseudo maximum likelihood approach. This approach suffers from the lack of identification existing in the Gaussian SVARMA. Whenever the shocks are not Gaussian, the SVARMA becomes identified up to a permutation and sign change of the structural shocks. This paper further proposes simple estimation methods able to consistently estimate non-fundamental representation in the moving average dynamics.

A dynamic model constructed to derive impulse response functions requires much more structural assumptions on the error terms (i.e. independence) than a pure forecast model for which uncorrelated errors may be sufficient. In this respect the conventional econometric toolboxes available for macroeconomists have been conceived for a forecast purpose and are not appropriate for the analysis of policy shocks.

Because it focuses on the second-order properties the SVARMA literature often introduces potentially misleading identification assumptions that entail misspecification and naive interpretations of VARMA residuals. To paraphrase Sims (1980): *"Nonlinear analysis is getting easier, both because of improved techniques and because of better computational hardware. This weakens the excuse that second-order analysis has to be followed just since it is simple"*. Nevertheless, as shown in the parametric and semi-parametric analysis developed in Section 4 and in the applications presented in Section 5, SVARMA can still be useful for economic policy, provided that the independence assumption is valid and the appropriate estimation methods are used.

The methods developed in this paper can be extended in several directions. First the asymptotic Gaussian distributions of the various estimators proposed can be derived and testing procedures, in particular tests of fundamentalness, can 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 mixed causal/noncausal MA process]. Third, the identification and estimation results might be extended to the case of more errors than observables. Indeed, if identification results exist when the errors are not Gaussian [see e.g. Th 3.1. in Eriksson and Koivunen (2004) in the static case, Gouriéroux and Zakoian (2015) for stable multivariate processes, or Gagliardini and Gouriéroux (2013) for a non-Gaussian factor model], this possibility to identify the dynamics when $m > n$ 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 Identifiability, Reversibility and Estimation in the Case of a MA(1) Process

The aim of this appendix is to illustrate some of the general results of the paper by considering the example of the one-dimensional MA(1) process: $y_t = \varepsilon_t - \theta\varepsilon_{t-1}$, where the $\varepsilon_t's$ are independent. We first consider the asymptotic behaviour of the approximated maximum likelihood approach. Then we illustrate the reason of identifiability in a non-Gaussian case, and consider a moment estimation method.

## A.1 Limit optimization problem in the approximate ML method

We assume that the p.d.f. of the $\varepsilon_t's$ belongs to the family $g(\varepsilon;\gamma)$. The approximate log-likelihood function is:

$$\begin{aligned} L_T(\theta,\gamma) &= \mathbb{1}_{|\theta|<1}\sum_{t=1}^{T}\log\left\{g\left(\sum_{h=0}^{t-1}\theta^h y_{t-h};\gamma\right)\right\} \\ &+ \mathbb{1}_{|\theta|>1}\sum_{t=1}^{T}\log\left\{\frac{1}{|\theta|}g\left(-\sum_{h=0}^{T-t-1}\frac{1}{\theta^{h+1}}y_{t+h+1};\gamma\right)\right\}. \end{aligned}$$

When $T$ goes to infinity $\frac{1}{T}L_T$ converges to the limit function:

$$\begin{aligned} L_\infty(\theta,\gamma) &= \mathbb{1}_{|\theta|<1}E_0\log g\left(\sum_{h=0}^{\infty}\theta^h y_{t-h};\gamma\right) \\ &+ \mathbb{1}_{|\theta|>1}E_0[\log\frac{1}{|\theta|}g\left(-\sum_{h=0}^{\infty}\frac{1}{\theta^{h+1}}y_{t+h+1};\gamma\right), \end{aligned}$$

where $E_0$ is the expectation with respect to the true distribution of the process. We also have:

$$\begin{aligned} L_\infty(\theta,\gamma) &= \mathbb{1}_{|\theta|<1}E_0\log g\left[y_t - E_\theta(y_t|y_{-\infty}^{t-1}),\gamma\right] \\ &+ \mathbb{1}_{|\theta|>1}E_0\left\{-\frac{1}{2}\log\theta^2 + \log g\left[-\frac{1}{\theta}(y_{t+1} - E_\theta(y_{t+1}|y_{t+2}^{\infty});\gamma\right]\right\} \end{aligned}$$

with $E_\theta(y_t|y_{-\infty}^{t-1}) = -\sum_{h=1}^{\infty}\theta^h y_{t-h}$ and $E_\theta(y_{t+1}|y_{t+2}^{\infty}) = -\sum_{h=1}^{\infty}\frac{1}{\theta^{h+1}}y_{t+h+1}$ (which does not depend on $\sigma$).

In the Gaussian case, where the distribution of $\varepsilon_t$ is $N(0,\sigma^2)$, we get:

$$\begin{aligned} L_\infty(\theta,\sigma^2) &= \mathbb{1}_{|\theta|<1} E_0\left[-\frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left(y_t - E_\theta(y_t|y_{-\infty}^{t-1})\right)^2\right] \\ &+ \mathbb{1}_{|\theta|>1} E_0\left\{-\frac{1}{2}\log(\theta^2\sigma^2) - \frac{1}{2\sigma^2}\left[\frac{1}{\theta^2}\left(y_{t+1} - E_\theta(y_{t+1}|y_{t+2}^{\infty})\right)\right]^2\right\}. \end{aligned}$$

The limit optimization problem is:

$$\min_{(\theta,\sigma^2)} \left[\mathbb{1}_{|\theta|<1} L_1^a(\theta,\sigma^2) + \mathbb{1}_{|\theta|>1} L_2^a(\theta,\sigma^2)\right]$$

$$\begin{aligned} \text{with} \quad L_1^a(\theta,\sigma^2) &= \log\sigma^2 + \frac{1}{\sigma^2} E_0\left[(y_t - E_\theta(y_t|y_{-\infty}^{t-1}))^2\right], \\ L_2^a(\theta,\sigma^2) &= \log(\theta^2\sigma^2) + \frac{1}{\theta^2\sigma^2} E_0\left[(y_{t+1} - E_\theta(y_{t+1}|y_{t+2}^{\infty}))^2\right]. \end{aligned}$$

Due to the reversibility of the Gaussian process, the true distribution of $y_{t+1} - E_\theta(y_{t+1}|y_{t+2}^{\infty})$, if $|\theta| > 1$, is the same as the true distribution of $y_{t+1} - E_{1/\theta}(y_{t+1}|y_{-\infty}^{t})$.

Let us first assume that $|\theta_0| < 1$ and let us consider the solutions of the limit optimization problem. In order to minimize $L_1^a(\theta,\sigma^2)$ on $|\theta| < 1$, we can consider the case where $\sigma^2$ is fixed and we get:

$$\min_\theta E_0\left[(y_t - E_\theta(y_t|y_{-\infty}^{t-1}))^2\right].$$

The minimum is reached for $\theta = \theta_0$, which does not depend on $\sigma^2$, and the minimum value of $L_1^a(\theta,\sigma^2)$ is then easily found to be $\log\sigma_0^2 + 1$. In order to minimize $L_2^a(\theta,\sigma^2)$ on $|\theta| > 1$, we can put $\theta^2\sigma^2 = \tilde{\sigma}^2$. Considering the case where $\tilde{\sigma}^2$ is fixed, we obtain the problem:

$$\min_\theta E_0\left[(y_{t+1} - E_{1/\theta}(y_{t+1}|y_{-\infty}^{t}))^2\right],$$

whose minimum is reached for $\theta = 1/\theta_0$, which does not depend on $\tilde{\sigma}^2$. Hence the minimum of $L_2^a(\theta,\sigma^2)$ is again $\log\sigma_0^2 + 1$.

When $|\theta_0|$ is larger than 1, we can see that $L_1^a(\theta,\sigma^2)$ is optimal for $1/\theta_0$ and $L_2^a(\theta,\sigma^2)$ for $\theta_0$ and we still have two inverse values of $\theta$ giving the same optimum, namely $\log(\theta_0^2\sigma_0^2) + 1$. The model is not asymptotically identifiable.

However in finite sample the optimal values of $L_1^a$ and $L_2^a$ are different, even in the Gaussian case. Thus the approximated ML approach will provide a unique solution, not necessarily well-located.

## A.2 Identification in the non-Gaussian case

Let us consider the joint distribution of $(y_t, y_{t-1})$. The characteristic function of this distribution is:

$$\begin{aligned}\psi(u,\upsilon) &= E\exp[i(uy_t+\upsilon y_{t-1})] \\ &= E\exp(iu\varepsilon_t)E\exp[i(\upsilon-u\theta)\varepsilon_{t-1}]E[\exp(-i\upsilon\theta\varepsilon_{t-2})].\end{aligned}$$

Let us for instance assume that $\varepsilon_t$ follows a stable distribution, we get:

$$\psi(u,\upsilon)=\exp[-c(|u|^\alpha+|\upsilon-u\theta|^\alpha+|\upsilon\theta|^\alpha)].$$

Is this function of $(c,\theta)$ injective? If $\alpha=2$, i.e. in the Gaussian case, we verify that

$$c[u^2+(\upsilon-u\theta)^2+\upsilon^2\theta^2]=c[(u^2+\upsilon^2)(1+\theta^2)-2u\upsilon\theta]$$

takes the same value for $(c,\theta)$ and $(c\theta^2,1/\theta)$ and we do not have identifiability. On the contrary for $\alpha\neq 2$, we see, for instance, that $\psi(u,\upsilon)$ is not differentiable on the lines $u=0$, $\upsilon=0$ and $\upsilon-u\theta=0$. The latter condition implies the identifiability of $\theta$.

## A.3 Moment method

If we do not want to make a parametric assumption about the distribution of $\varepsilon_t$, we can use a moment method based on higher-order cross moments (see Section 4.1).

Let us consider again the one-dimensional MA(1) process. We have:

$$E(y_t y_{t-1}^2)=-\theta E\varepsilon_t^3, E(y_t^2 y_{t-1})=\theta^2 E\varepsilon_t^3,$$

and, therefore:

$$\theta=-\frac{E(y_t^2 y_{t-1})}{E(y_t y_{t-1}^2)},$$

whenever $\varepsilon_t$ has a skewed distribution, i.e. $E(\varepsilon_t^3)\neq 0$. Thus the location of $|\theta|$ w.r.t. 1 is identified from the lack of time reversibility of the process.

# B Proof of the Proposition in Section 3.1

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*

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

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

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

*C1 The components of $\varepsilon_t$ (resp. $\varepsilon_t^*$) are identically distributed.*

*C2 The components of $\varepsilon_t$ (resp. $\varepsilon_t^*$) have non-zero $r^{th}$ cumulant, with $r \geq 3$ and a finite even moment of order $s$ greater than $r$.*

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

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

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 $A_k$ are identified up to a permutation and a sign change of the columns. That is, there exist a permutation matrix $P$ and a diagonal matrix $D$, whose diagonal elements are either $-1$ or 1, that are such that:

$$\Psi_k C = \Psi_k^* C^* PD, \qquad \forall\, k.$$

For $k = 0$, this gives $C = C^* PD$, 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.

# C Recovering Structural Shocks

This appendix deals with the problem of recovering the errors $\varepsilon_t$, $t = 1, \ldots, T$, when one observes $Y_t$, $t = 1, \ldots, T$, when the (true) dynamics of $Y_t$ is defined through equation (2.1), that is:

$$\Phi(L) Y_t = \Theta(L)\varepsilon_t, \tag{a.3}$$

with the roots of $\det(\Theta(L))$ inside or outside the unit circle.

To begin with, let us consider the case where $\Theta(L) = I - \Theta L$ and $\Theta$ is of dimension $2 \times 2$ (i.e. $n = 2$). Assume further that $\Theta$ is diagonalizable, we have:

$$\Theta = A \Lambda A^{-1}, \quad \text{with} \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}. \tag{a.4}$$

Introducing $V_t = A^{-1}\Phi(L) Y_t$, we have:

$$V_t = (I - \Lambda L)\varepsilon_t^*,$$

where $\varepsilon_t^* = A^{-1}\varepsilon_t$. Let us consider the case where $|\lambda_1| < 1$ and $|\lambda_2| > 1$. We have:

$$\begin{aligned} V_{1,t} &= (1 - \lambda_1 L)\varepsilon_{1,t}^* \\ V_{2,t} &= (1 - \lambda_2 L)\varepsilon_{2,t}^* = -\lambda_2 L \left(1 - \frac{1}{\lambda_2} L^{-1}\right) \varepsilon_{2,t}^*, \end{aligned}$$

which gives (using $|\lambda_1| < 1$ and $|\lambda_2| > 1$):

$$\varepsilon_t = A\varepsilon_t^* = A \begin{bmatrix} (1 - \lambda_1 L)^{-1} & 0 \\ 0 & -\frac{1}{\lambda_2} L^{-1} \left(1 - \frac{1}{\lambda_2} L^{-1}\right)^{-1} \end{bmatrix} V_t.$$

Suppose now that $\Theta$ is $(n \times n)$, not necessarily diagonalizable, and with eigenvalues inside or outside the unit circle. In this case, one can exploit the Schur decomposition of $\Theta$, that is:

$$\Theta = AUA',$$

where $A$ is orthogonal, and $U$ is upper block-triangular, where the diagonal blocks ($U_k$, $k \in \{1,\dots,K\}$) are either $1 \times 1$ or $2 \times 2$ blocks, the $2 \times 2$ blocks corresponding to complex conjugate complex eigenvalues of $\Theta$ (see equation (4.13)).

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

$$W_t = \varepsilon^*_t - U\varepsilon^*_{t-1}, \tag{a.5}$$

where $W_t = A'\Phi(L)Y_t$ and $\varepsilon^*_t = A'\varepsilon_t$. In particular, we have:

$$W_{K,t} = \varepsilon^*_{K,t} - U_K\varepsilon^*_{K,t-1},$$

where $W_{K,t}$ is of dimension 1 or 2, depending on the dimension of $U_K$. The eigenvalues of $U_K$ have the same modulus. If this modulus is strictly inside the unit circle, we have:

$$\varepsilon^*_{K,t} = \sum_{h=0}^{\infty} U_K^h W_{K,t-h}, \tag{a.6}$$

otherwise:

$$\varepsilon^*_{K,t} = -\sum_{h=1}^{\infty} (U_K^{-1})^h W_{K,t+h}. \tag{a.7}$$

Let us now introduce $W^*_{K-1,t}$, defined by:

$$W^*_{K-1,t} = W_{K-1,t} + U_{K-1,K}\varepsilon^*_{K,t-1}.$$

From equation (a.5), we have:

$$W^*_{K-1,t} = \varepsilon^*_{K-1,t} - U_{K-1}\varepsilon^*_{K-1,t-1}.$$

Depending on the modulus of the eigenvalues of $U_{K-1}$, one can then obtain $\varepsilon^*_{K-1,t}$ as in equation (a.11) or equation (a.10), replacing $W$ by $W^*$. Iterating on this provides us with $\varepsilon^*_t = [\varepsilon^{*\prime}_{1,t}, \varepsilon^{*\prime}_{2,t}, \dots, \varepsilon^{*\prime}_t]'$.

In practice, to get estimates of the $\varepsilon^*_t$'s –and further of the $\varepsilon_t$'s (using $\varepsilon_t = A\varepsilon^*_t$)– the infinite sums in equations of the type of (a.11) or of (a.10) are truncated. More precisely, the estimates of

$\hat{\varepsilon}^*_{k,t}$ (for $K \in \{K,\dots,1\}$) are obtained by using either:

$$\hat{\varepsilon}^*_{k,t} = \sum_{h=0}^{t-1} U_k^h \hat{W}^*_{k,t-h}, \tag{a.8}$$

or

$$\hat{\varepsilon}^*_{k,t} = -\sum_{h=1}^{T-t} (U_k^{-1})^h \hat{W}^*_{k,t+h}, \tag{a.9}$$

where, for all $t$, $\hat{W}^*_{K,t} = W_{K,t}$ and where the $\hat{W}^*_{k,t}$'s are recursively defined by (for $k \in \{K-1,\dots,1\}$):

$$\hat{W}^*_{k,t} = W_{k,t} + \sum_{j=k}^{K-1} U_{k,j+1} \hat{\varepsilon}^*_{j+1,t-1}.$$

If $\Theta(L)$ is of order $q > 1$, on can go back to the previous case. For this, define $\tilde{\varepsilon}_t = [\varepsilon_t', \dots, \varepsilon_{t-q+1}']'$ and $\tilde{V}_t = [V_t', \dots, V_{t-q+1}']'$, where $V_t = \Phi(L)Y_t$. Using obvious notations, we then have: $\tilde{V}_t = (I - \tilde{\Theta}L)\tilde{\varepsilon}_t$. Note that the eigenvalues of $\tilde{\Theta}$ are the roots of $\det\Theta(L)$ [see Davis and Song (2012) and Gouriéroux and Jasiak (2017) where the problem is completely treated in the dual case, where the roots of $\det\Phi(L)$ can be inside or outside the unit circle].

# D Approximation of the log-likelihood

In this appendix, for ease of exposition, we focus on the computation of an approximate log-likelihood in the VMA(1) case:

$$Y_t = \varepsilon_t - \Theta\varepsilon_{t-1},$$

where the errors $\varepsilon_t$ are serially independent, with a p.d.f. of the form $g(\varepsilon,\Gamma)$. This case is easily extended to the VARMA($p$,1) case by replacing $Y_t$ by $\Phi(L)Y_t$ (sacrificing the first $p$ observations of $Y_t$). It can further be extended to the VARMA($p$,$q$) case ($q > 1$) by resorting to the approach sketched at the end of Appendix C.

Let's start from the equation (4.14):

$$W_t = \varepsilon^*_t - U\varepsilon^*_{t-1},$$

where $W_t = A'Y_t$ and $\varepsilon^*_t = A'\varepsilon_t$, and where $A$ and $U$ result from the real Schur decomposition of $\Theta$: $\Theta = A'UA$ (see equation (4.13)).

If the roots of $U$ are not on the unit circle, we can assume, without loss of generality, that $U_1,\dots U_q$ have eigenvalues with modulus strictly lower than 1 whereas $U_{q+1},\dots U_K$ have eigenvalues with modulus strictly larger than 1.

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 $\varepsilon_t^{*(1)}$ being equal to $m = n_1 + \cdots + n_q$, where $n_j$ is such that $U_j$ is of dimension $n_j \times n_j$ ($n_j \in \{1,2\}$). 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 $U$, equation (4.14) writes:

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

The previous equation leads to:

$$\begin{aligned} \varepsilon_t^{*(2)} &= \left\{-(U^{(2)})^{-1}\right\} W_{t+1}{}^{(2)} + \cdots + \left\{-(U^{(2)})^{-1}\right\}^{T-t-1} W_T{}^{(2)} + \\ &\quad \left\{-(U^{(2)})^{-1}\right\}^{T-t} \varepsilon_T^{*(2)}, \end{aligned} \tag{a.10}$$

and to

$$\begin{aligned} \varepsilon_t^{*(1)} &= W_t{}^{(1)} + U^{(1)} W_{t-1}{}^{(1)} + \cdots + U^{(1)^{t-1}} W_1{}^{(1)} + U^{(1)^t} \varepsilon_0^{*(1)} + \\ &\quad U^{(12)} \varepsilon_{t-1}^{*}{}^{(2)} + U^{(1)} U^{(12)} \varepsilon_{t-2}^{*}{}^{(2)} + \cdots + U^{(1)^{t-1}} U^{(12)} \varepsilon_0^{*(2)}. \end{aligned} \tag{a.11}$$

Replacing the $\varepsilon_t^{*(2)}$'s appearing in equation (a.11) by their expressions given in (a.10), one obtains:

$$\underbrace{\left[\begin{array}{c} \varepsilon_1^{*(1)} \\ \vdots \\ \varepsilon_T^{*(1)} \\ \hline \varepsilon_0^{*(2)} \\ \vdots \\ \varepsilon_{T-1}^{*}{}^{(2)} \end{array}\right]}_{=\varepsilon^*} = \underbrace{\left[\begin{array}{c|c} J_1 & J_{12} \\ \hline 0 & J_2 \end{array}\right]}_{=J} \underbrace{\left[\begin{array}{c} W_1{}^{(1)} \\ \vdots \\ W_T{}^{(1)} \\ \hline W_1{}^{(2)} \\ \vdots \\ W_T{}^{(2)} \end{array}\right]}_{=W} + \underbrace{\left[\begin{array}{c|c} U^{(1)} & M_1 \\ \vdots & \vdots \\ U^{(1)^T} & M_T \\ \hline 0 & \left\{-(U^{(2)})^{-1}\right\}^T \\ \vdots & \vdots \\ 0 & \left\{-(U^{(2)})^{-1}\right\} \end{array}\right]}_{=M} \begin{bmatrix} \varepsilon_0^{*(1)} \\ \varepsilon_T^{*(2)} \end{bmatrix}, \tag{a.12}$$

where $J_1$ is upper block triangular with identity matrices on its diagonal, where $J_2$ is an upper block triangular matrix with $\left\{-(U^{(2)})^{-1}\right\}$ matrices on its diagonal and where:

$$M_t = U^{(12)} \left\{-(U^{(2)})^{-1}\right\}^{T-(t-1)} + \cdots + U^{(1)^{t-1}} U^{(12)} \left\{-(U^{(2)})^{-1}\right\}^T.$$

Because the eigenvalues of $U^{(1)}$ and of $(U^{(2)})^{-1}$ are strictly inside the unit circle, the elements

of $M$ corresponding to $0 \ll t \ll T$ are extremely small. As a result, for $0 \ll t \ll T$, the $\varepsilon_t^*$'s are well approximated by the corresponding components of $JW$.

The likelihood associated with $\varepsilon^* = [{e^*}_1^{(1)'}, \ldots, {e^*}_T^{(1)'}, {e^*}_0^{(2)'}, \ldots, {e^*}_{T-1}^{(2)'}]$ is:

$$g^*(\varepsilon^*, \Gamma) = g_{\varepsilon^{*(1)}}\left({e^*}_T^{(1)}, \Gamma\right) g_{\varepsilon^{*(2)}}\left({e^*}_0^{(2)}, \Gamma\right) \prod_{t=1}^{T-1} g\left(A e_t^*, \Gamma\right),$$

with $e_t^* = [{e^*}_t^{(1)'}, {e^*}_t^{(2)'}]'$.

Let's consider the vector $Y = [Y_1', \ldots, Y_T']'$. If $A'Y_t = W_t$ for $t \in \{1, \ldots, T\}$ and if $\tilde{W} = [W_1', \ldots, W_T']'$, we have:

$$\tilde{W} = (I \otimes A')Y.$$

In addition, let's denote by $P$ the permutation that is such that $W = P\tilde{W}$, where $W$ is defined in equation (a.12). Then an approximation of $\varepsilon^*$ is given by $\mathscr{E}(Y) := JP(I \otimes A')Y$.

Because $A$ and $P$ are orthogonal matrices, we have:

$$|\det(JP[I \otimes A'])| = |\det(J)| = \frac{1}{|\det(U^{(2)})|^T}.$$

Therefore, the likelihood associated with $Y$ can be approximated by:

$$|\det(J)| g^*(\mathscr{E}(Y), \Gamma) = \frac{1}{|\det(U^{(2)})|^T} g^*(\mathscr{E}(Y), \Gamma),$$

or by:

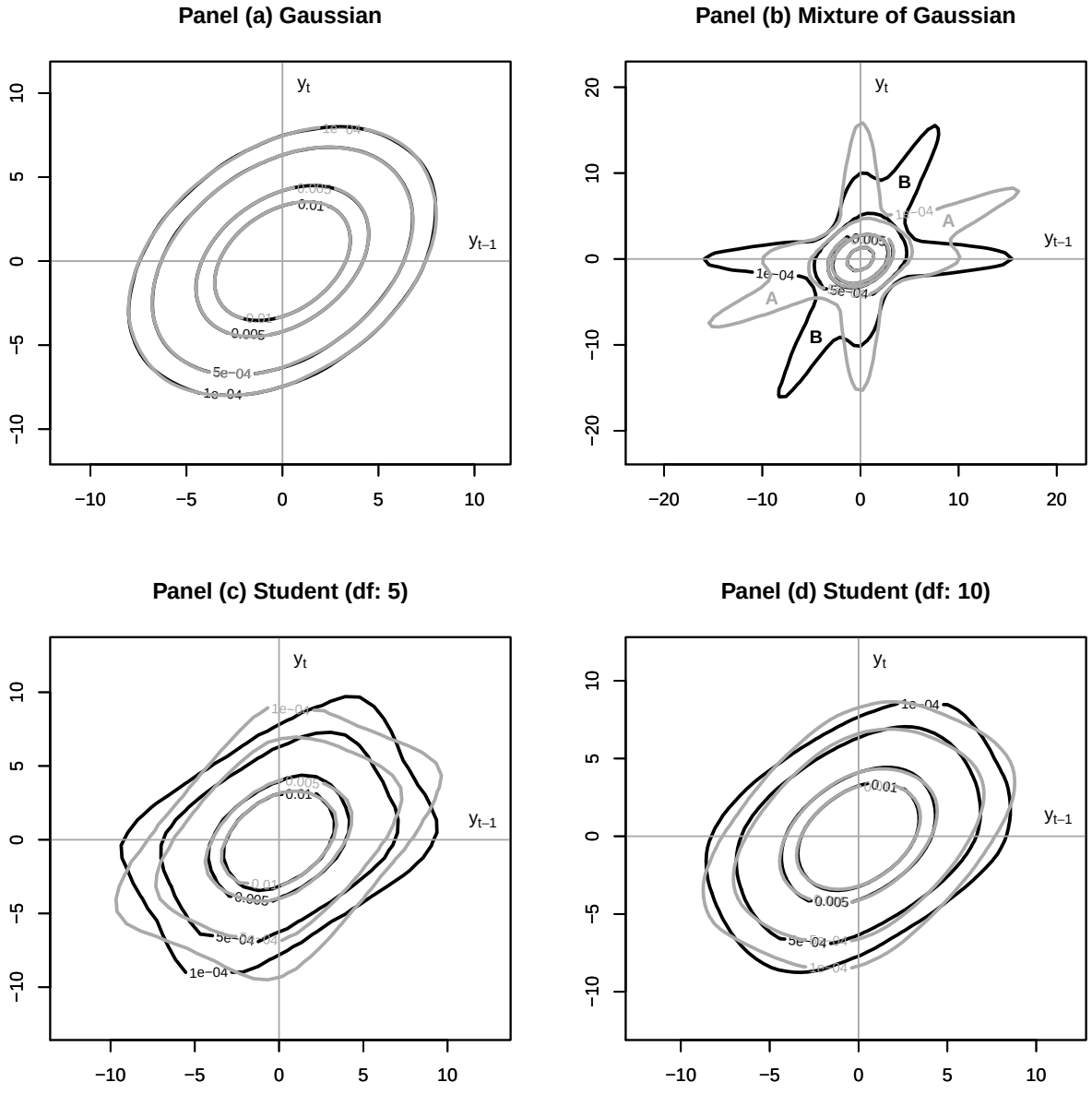
$$\frac{1}{|\det(U^{(2)})|^T} \tilde{g}(A\mathscr{E}(Y), \Gamma),$$

where

$$\tilde{g}(\varepsilon^*, \Gamma) = \prod_{t=1}^{T-1} g\left(A e_t^*, \Gamma\right),$$

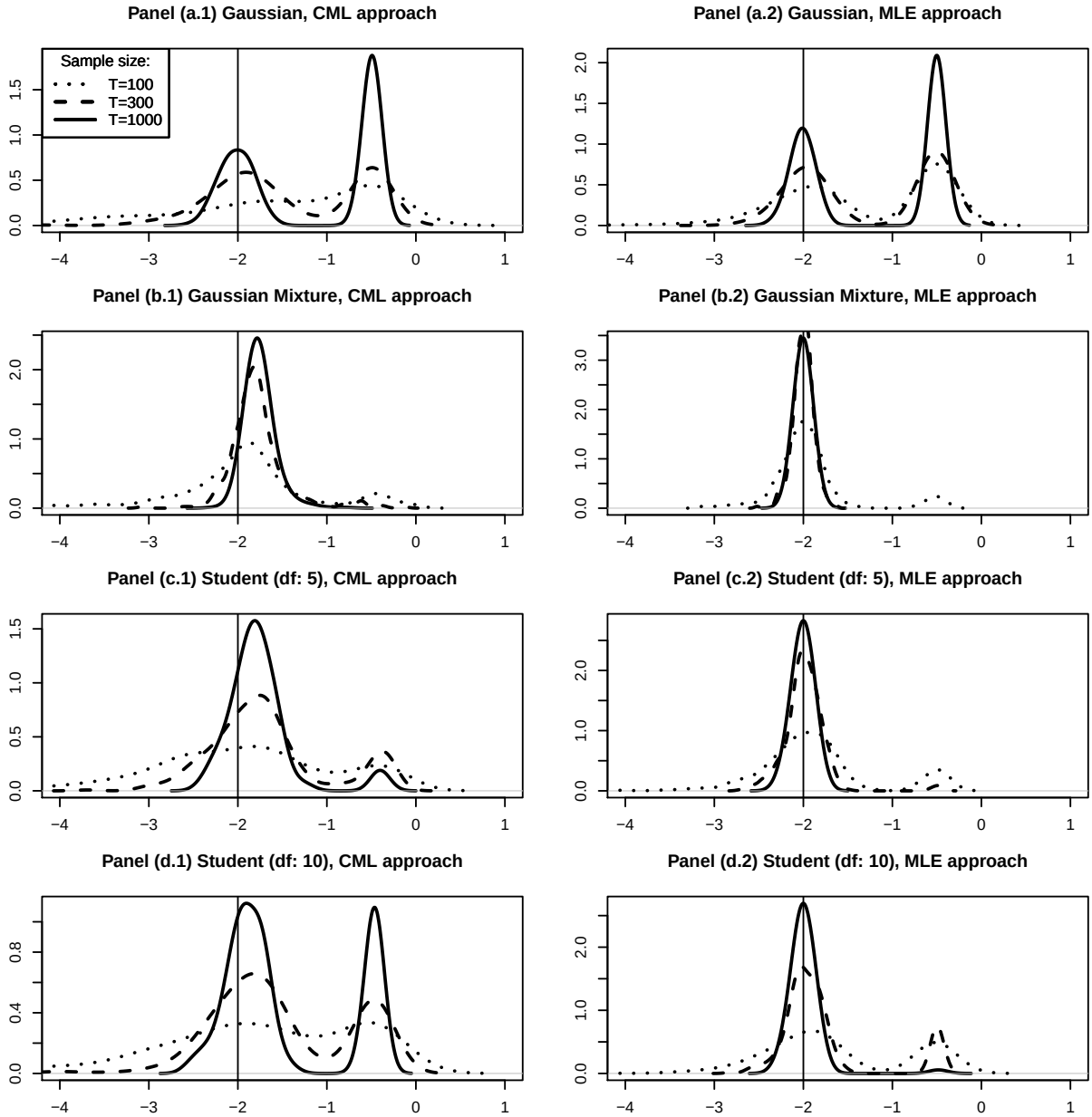
with $\varepsilon^* = [{e^*}_1^{(1)'}, \ldots, {e^*}_T^{(1)'}, {e^*}_0^{(2)'}, \ldots, {e^*}_{T-1}^{(2)'}]$ and $e_t^* = [{e^*}_t^{(1)'}, {e^*}_t^{(2)'}]'$.

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




*Notes:* Each of these 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 $1-p$, with $\sigma_1 = 5 = 10 \times \sigma_2$, which implies $p \approx 3\%$ (in order to have $V(\varepsilon_t) = 1$). See the text (Subsection 5.1) for explanations regarding the letters A and B appearing on Panel (b).

Figure 2: Distribution of the Composite Maximum Likelihood Estimator of $\theta$




*Notes:* These panels display the distributions of the estimates of $\theta$ obtained with the Composite Maximum Likelihood (CML) approach. 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: $T = 100$, 300 and 1000. For each distribution and each sample size, we simulate a large number $N = 500$ of $y_t$ samples of size $T$. For each simulated sample, we employ the CML approach to estimate $(\theta, V(\varepsilon_t))$. 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

| $\varepsilon_t$'s distribution: | Bias | RMSE | MAE | S.D. | $\overline{\sigma_{asy}}$ | $\alpha$ =75% | $\alpha$ =90% | $\alpha$ =95% |
|---|---|---|---|---|---|---|---|---|
| **Panel (a) Composite Maximum Likelihood approach** | | | | | | | | |
| Sample size: T=100 | | | | | | | | |
| Gaussian | 0.32 | 1.47 | 1.18 | 1.44 | 0.54 | 0.41 | 0.46 | 0.47 |
| Mixture of Gaussian | 0.06 | 0.85 | 0.57 | 0.85 | 0.30 | 0.42 | 0.55 | 0.63 |
| Student (df: 5) | 0.07 | 1.03 | 0.81 | 1.03 | 0.53 | 0.47 | 0.56 | 0.59 |
| Student (df: 10) | 0.19 | 1.27 | 0.99 | 1.26 | 0.56 | 0.45 | 0.49 | 0.53 |
| Sample size: T=300 | | | | | | | | |
| Gaussian | 0.58 | 0.99 | 0.77 | 0.80 | 0.29 | 0.47 | 0.55 | 0.57 |
| Mixture of Gaussian | 0.22 | 0.39 | 0.27 | 0.32 | 0.14 | 0.38 | 0.51 | 0.56 |
| Student (df: 5) | 0.33 | 0.74 | 0.53 | 0.66 | 0.30 | 0.47 | 0.60 | 0.66 |
| Student (df: 10) | 0.47 | 0.91 | 0.68 | 0.78 | 0.29 | 0.46 | 0.56 | 0.60 |
| Sample size: T=1000 | | | | | | | | |
| Gaussian | 0.80 | 1.11 | 0.87 | 0.77 | 0.12 | 0.37 | 0.43 | 0.45 |
| Mixture of Gaussian | 0.24 | 0.29 | 0.24 | 0.16 | 0.10 | 0.18 | 0.30 | 0.40 |
| Student (df: 5) | 0.24 | 0.45 | 0.30 | 0.38 | 0.20 | 0.51 | 0.62 | 0.70 |
| Student (df: 10) | 0.53 | 0.87 | 0.61 | 0.69 | 0.16 | 0.40 | 0.54 | 0.60 |
| **Panel (b) Maximum Likelihood approach** | | | | | | | | |
| Sample size: T=100 | | | | | | | | |
| Gaussian | 0.67 | 1.08 | 0.88 | 0.84 | 0.25 | 0.38 | 0.45 | 0.48 |
| Mixture of Gaussian | 0.06 | 0.46 | 0.27 | 0.46 | 0.21 | 0.71 | 0.82 | 0.87 |
| Student (df: 5) | 0.14 | 0.66 | 0.45 | 0.64 | 0.32 | 0.66 | 0.76 | 0.79 |
| Student (df: 10) | 0.38 | 0.88 | 0.66 | 0.79 | 0.30 | 0.53 | 0.64 | 0.66 |
| Sample size: T=300 | | | | | | | | |
| Gaussian | 0.73 | 1.05 | 0.81 | 0.75 | 0.12 | 0.37 | 0.45 | 0.48 |
| Mixture of Gaussian | 0.00 | 0.11 | 0.08 | 0.11 | 0.11 | 0.76 | 0.89 | 0.95 |
| Student (df: 5) | 0.02 | 0.26 | 0.16 | 0.26 | 0.17 | 0.72 | 0.86 | 0.89 |
| Student (df: 10) | 0.23 | 0.60 | 0.35 | 0.56 | 0.17 | 0.61 | 0.74 | 0.78 |
| Sample size: T=1000 | | | | | | | | |
| Gaussian | 0.80 | 1.10 | 0.85 | 0.75 | 0.06 | 0.34 | 0.39 | 0.43 |
| Mixture of Gaussian | 0.00 | 0.05 | 0.04 | 0.05 | 0.05 | 0.76 | 0.91 | 0.96 |
| Student (df: 5) | 0.00 | 0.09 | 0.07 | 0.09 | 0.09 | 0.75 | 0.89 | 0.96 |
| Student (df: 10) | 0.01 | 0.21 | 0.10 | 0.21 | 0.10 | 0.74 | 0.89 | 0.93 |

*Notes:* The model is $y_t = \varepsilon_t - \theta\varepsilon_{t-1}$, with $\theta = -2$ and $V(\varepsilon_t) = 1$. This table reports the results of a Monte-Carlo experiment based on the simulation of $N = 500$ samples for each of the four distributions considered for the errors $\varepsilon_t$ (see first column) and each of the three considered sample sizes ($T = 100$, 300 or 1000). For each simulated sample, we employ the Composite Maximum Likelihood (CML) approach (Panel (a)) and the Maximum Likelihood approach (Panel (b)) to estimate $(\theta, V(\varepsilon_t))$. Columns 2 to 5 give, respectively, the bias of the estimator of $\theta$, the root mean-squared error associated with this estimator, its mean absolute error and its standard deviation. The next column ($\overline{\sigma_{asy}}$) gives the mean (across the $N$ simulations) of the asymptotic standard deviations based on the estimated covariance matrix involving both the Hessian and the outer product of the first derivatives of the composite likelihood function. (In the sandwich formula, the computation of the matrix based on the outer product of the first derivatives makes use of the Newey-West approach to deal with the serial correlation of the different terms.) The last three 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 $P(-\phi_\alpha < X < \phi_\alpha) = \alpha$ if $X \sim \mathcal{N}(0,1)$ (i.e. $\phi_\alpha = 1.15$, 1.64 and 1.96 for the last three columns, respectively).

Table 2: Fitted MA(1) processes for per capita GDP growth rates

| Country | $1^{st}$ *year* | Bai-Ng p-value | $\phi$ | | $\theta$ | | $c$ | | $\mu_1$ | | $\sigma_1$ | | $p$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Austria | 1870 | 0.092 | 0.43 | *(0.08)* | −0.06 | *(0.05)* | 9.83 | *(2.18)* | −0.35 | *(0.31)* | 3.21 | *(0.53)* | 0.08 | *(0.03)* |
| Belgium | 1846 | 0.004 | 0.60 | *(0.04)* | 2.56 | *(0.27)* | 1.46 | *(0.39)* | 0.33 | *(0.89)* | 3.69 | *(0.44)* | 0.05 | *(0.00)* |
| Denmark | 1820 | 0.013 | −0.29 | *(0.31)* | −0.34 | *(0.31)* | 3.51 | *(0.32)* | −0.15 | *(0.17)* | 1.77 | *(0.20)* | 0.24 | *(0.07)* |
| Finland | 1860 | 0.015 | −0.29 | *(0.06)* | −1.66 | *(0.16)* | 2.60 | *(0.33)* | −0.59 | *(0.30)* | 1.56 | *(0.15)* | 0.27 | *(0.09)* |
| France | 1820 | 0.131 | −0.80 | *(0.06)* | −1.24 | *(0.10)* | 4.88 | *(0.66)* | −0.34 | *(0.24)* | 2.10 | *(0.27)* | 0.16 | *(0.05)* |
| Germany | 1850 | 0.045 | 0.06 | *(0.15)* | −0.33 | *(0.11)* | 7.50 | *(1.33)* | −0.63 | *(0.36)* | 2.91 | *(0.40)* | 0.09 | *(0.03)* |
| Italy | 1800 | 0.000 | 0.96 | *(0.00)* | 1.06 | *(0.02)* | 4.10 | *(0.39)* | −0.50 | *(0.22)* | 1.92 | *(0.21)* | 0.20 | *(0.06)* |
| Netherlands | 1815 | 0.000 | 0.54 | *(0.14)* | 0.37 | *(0.14)* | 5.95 | *(1.43)* | −0.09 | *(0.76)* | 3.89 | *(0.30)* | 0.05 | *(0.00)* |
| Norway | 1830 | 0.000 | 0.15 | *(0.17)* | −0.02 | *(0.18)* | 3.54 | *(0.34)* | −0.26 | *(0.15)* | 1.77 | *(0.19)* | 0.24 | *(0.06)* |
| Sweden | 1800 | 0.018 | 0.05 | *(0.04)* | −11.94 | *(7.83)* | 0.28 | *(0.18)* | −0.11 | *(0.04)* | 1.20 | *(0.04)* | 0.65 | *(0.06)* |
| Switzerland | 1851 | 0.029 | 0.44 | *(0.08)* | 1.76 | *(0.24)* | 3.51 | *(0.57)* | −0.14 | *(0.16)* | 1.65 | *(0.18)* | 0.31 | *(0.09)* |
| United Kingdom | 1800 | 0.061 | −0.11 | *(0.11)* | −6.42 | *(4.58)* | 0.51 | *(0.36)* | −0.64 | *(0.36)* | 1.51 | *(0.20)* | 0.23 | *(0.12)* |
| Ireland | 1921 | 0.297 | 0.95 | *(0.02)* | 0.17 | *(0.03)* | 2.68 | *(0.24)* | 0.00 | *(0.00)* | 1.16 | *(0.05)* | 0.74 | *(0.06)* |
| Portugal | 1865 | 0.073 | 0.96 | *(0.02)* | 1.13 | *(0.09)* | 3.47 | *(0.41)* | −0.25 | *(0.20)* | 1.65 | *(0.24)* | 0.25 | *(0.10)* |
| Spain | 1850 | 0.035 | 0.38 | *(0.24)* | 0.15 | *(0.25)* | 4.43 | *(0.45)* | −1.99 | *(0.79)* | 1.99 | *(1.00)* | 0.05 | *(0.00)* |
| Canada | 1870 | 0.015 | −0.24 | *(0.07)* | −2.21 | *(0.32)* | 2.20 | *(0.37)* | −0.85 | *(0.36)* | 1.52 | *(0.17)* | 0.22 | *(0.08)* |
| USA | 1800 | 0.057 | 0.04 | *(0.10)* | −9.11 | *(7.74)* | 0.50 | *(0.42)* | −0.05 | *(0.15)* | 1.65 | *(0.28)* | 0.27 | *(0.13)* |

*Notes:* This table reports the results of estimations of ARMA(1) processes for long historical series of per capita GDP annual growth rates (data from Bolt and van Zanden (2014)). The model is $y_t = \phi y_{t-1} + c\eta_t - \theta c\eta_{t-1}$, where the distribution of $\eta_t$ is a Gaussian mixture, i.e. $\eta_t$ is drawn from $\mathcal{N}(\mu_1, \sigma_1^2)$ with a probability $p$ and from $\mathcal{N}(\mu_2, \sigma_2^2)$ with a probability $1-p$. Because we impose that $E(\eta_t) = 0$ and $V(\eta_t) = 1$, the distribution of $\eta_t$ is completely defined by $\gamma = [\mu_1, \sigma_1, p]'$. The model is estimated by Maximum Likelihood. The first year of data is given in the second column and, for all countries, the last year is 2010. The third column reports the p-value of the Bai and Ng (2005) normality test, which is an extension of the Bera and Jarque (1981) test to time series.

Table 3: Estimation of bivariate VARMA(1,1) models

| | Sample: 1948-1982 | | Sample: 1982-2016 | |
|---|---|---|---|---|
| **Panel (a) Parameter estimates** | | | | |
| | param. | std dev. | param. | std dev. |
| $\Phi_{1,1}$ | 0.373 | *(0.086)* | 0.769 | *(0.048)* |
| $\Phi_{2,1}$ | −0.596 | *(0.063)* | −0.629 | *(0.082)* |
| $\Phi_{1,2}$ | 0.230 | *(0.056)* | 0.064 | *(0.014)* |
| $\Phi_{2,2}$ | 1.059 | *(0.025)* | 0.987 | *(0.016)* |
| $\Theta_{1,1}$ | 0.010 | *(0.110)* | 0.843 | *(0.172)* |
| $\Theta_{2,1}$ | −0.133 | *(0.142)* | −0.888 | *(0.215)* |
| $\Theta_{1,2}$ | −1.567 | *(0.872)* | −3.517 | *(1.749)* |
| $\Theta_{2,2}$ | 3.824 | *(1.232)* | 4.451 | *(2.024)* |
| $C_{1,1}$ | −0.724 | *(0.075)* | −0.142 | *(0.068)* |
| $C_{2,1}$ | 0.099 | *(0.020)* | −0.046 | *(0.015)* |
| $C_{1,2}$ | 0.753 | *(0.117)* | −0.479 | *(0.041)* |
| $C_{2,2}$ | −0.004 | *(0.026)* | −0.030 | *(0.011)* |
| $\mu_{1,1}$ | −0.048 | *(0.150)* | −0.555 | *(0.086)* |
| $\mu_{2,1}$ | 0.327 | *(0.093)* | −0.069 | *(0.234)* |
| $\sigma_{1,1}$ | 0.647 | *(0.127)* | 0.429 | *(0.067)* |
| $\sigma_{2,1}$ | 0.420 | *(0.068)* | 0.754 | *(0.151)* |
| **Panel (b) Inverses of the roots of** $\det\Theta(L)$ | | | | |
| | first | second | first | second |
| | 3.878 | 0.043 | 5.173 | 0.121 |

*Notes:* This table reports the results of the estimation of VARMA(1,1) models accounting for the dynamics of two endogenous variables: the real GDP growth (i.e. $\Delta gdp$, where $gdp = log(GDP)$) and the unemployment rate. The model is:

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

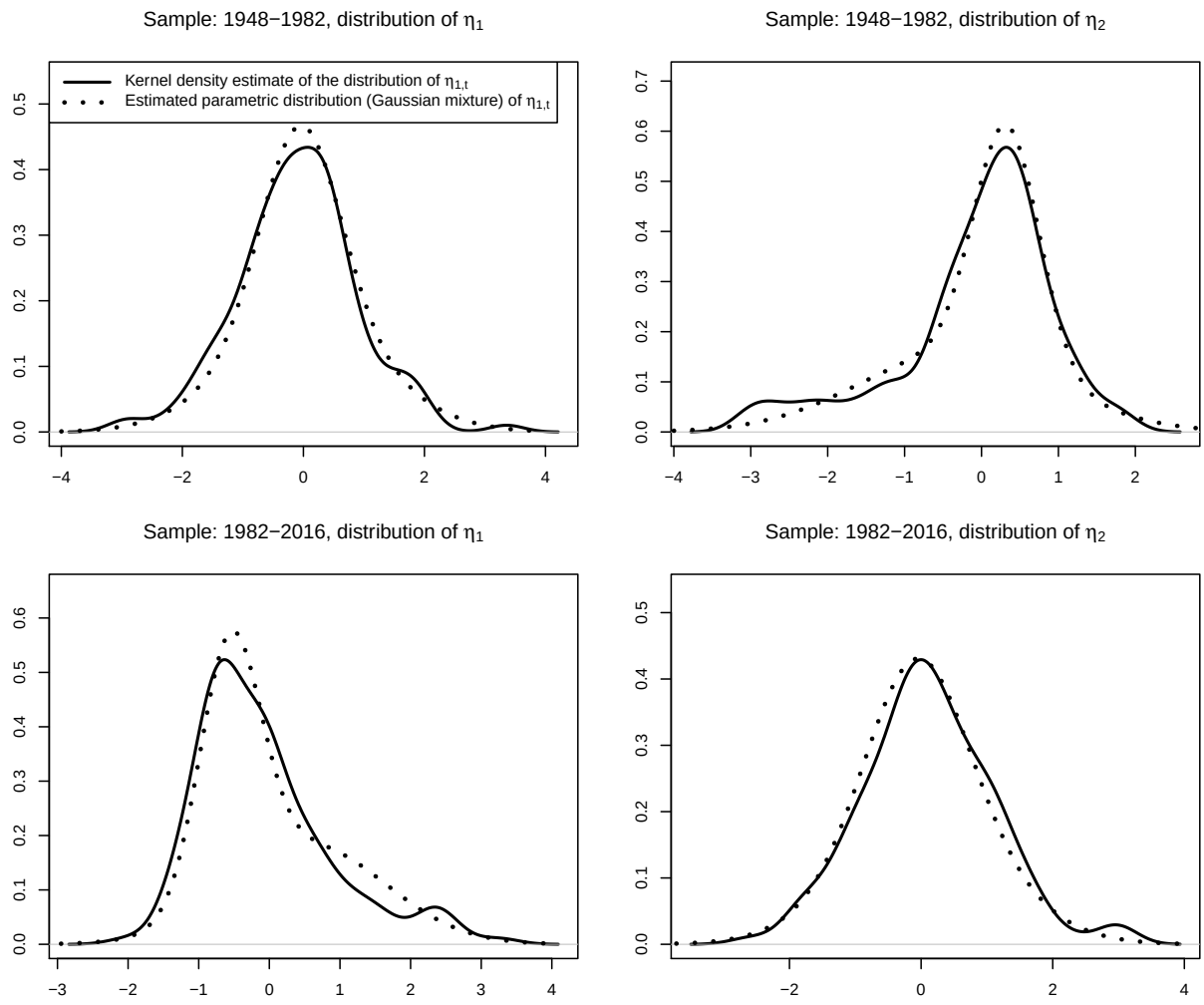
where, for $j \in \{1,2\}$, $\eta_{j,t}$ is drawn from $\mathcal{N}(\mu_{j,1}\sigma^2_{j,1})$ with probability $p = 0.5$ and from $\mathcal{N}(\mu_{j,2}\sigma^2_{j,2})$ with probability $1 - p = 0.5$. We impose $E(\eta_{j,t}) = 0$ and $V(\eta_{j,t}) = 1$, which implies that $\mu_{j,2}$ and $\sigma_{j,2}$ can be deduced from $\mu_{j,1}$ and $\sigma_{j,1}$. The (demeaned) data are quarterly and the model is estimated on two samples: 1948Q2-1982Q3 and 1982Q4-2016Q4. The Maximum Likelihood (ML) approach is employed to estimate the model. Panel (a) reports parameter estimates along with asymptotic standard deviations based on outer product of the first derivative of the likelihood function. Panel (b) indicates the inverses of the roots of $\det(I - \Theta L)$.

Table 4: Variance decompositions

| | Sample: 1948-1982 | | Sample: 1982-2016 | |
|---|---|---|---|---|
| **Share accounted for by Shock 1** | | | | |
| | $gdp$ | Unempl. | $gdp$ | Unempl. |
| horizon (quarters) | Long-Run Restrictions (Blanchard and Quah) | | | |
| 4 | 0.39 | 0.85 | 0.06 | 0.73 |
| 20 | 0.13 | 0.47 | 0.01 | 0.45 |
| 40 | 0.09 | 0.46 | 0.00 | 0.46 |
| 400 | 0.01 | 0.46 | 0.00 | 0.46 |
| horizon (quarters) | Maximum Likelihood | | | |
| 4 | 0.58 | 0.93 | 0.27 | 0.95 |
| 20 | 0.37 | 0.94 | 0.21 | 0.89 |
| 40 | 0.28 | 0.94 | 0.13 | 0.89 |
| 400 | 0.12 | 0.94 | 0.01 | 0.89 |

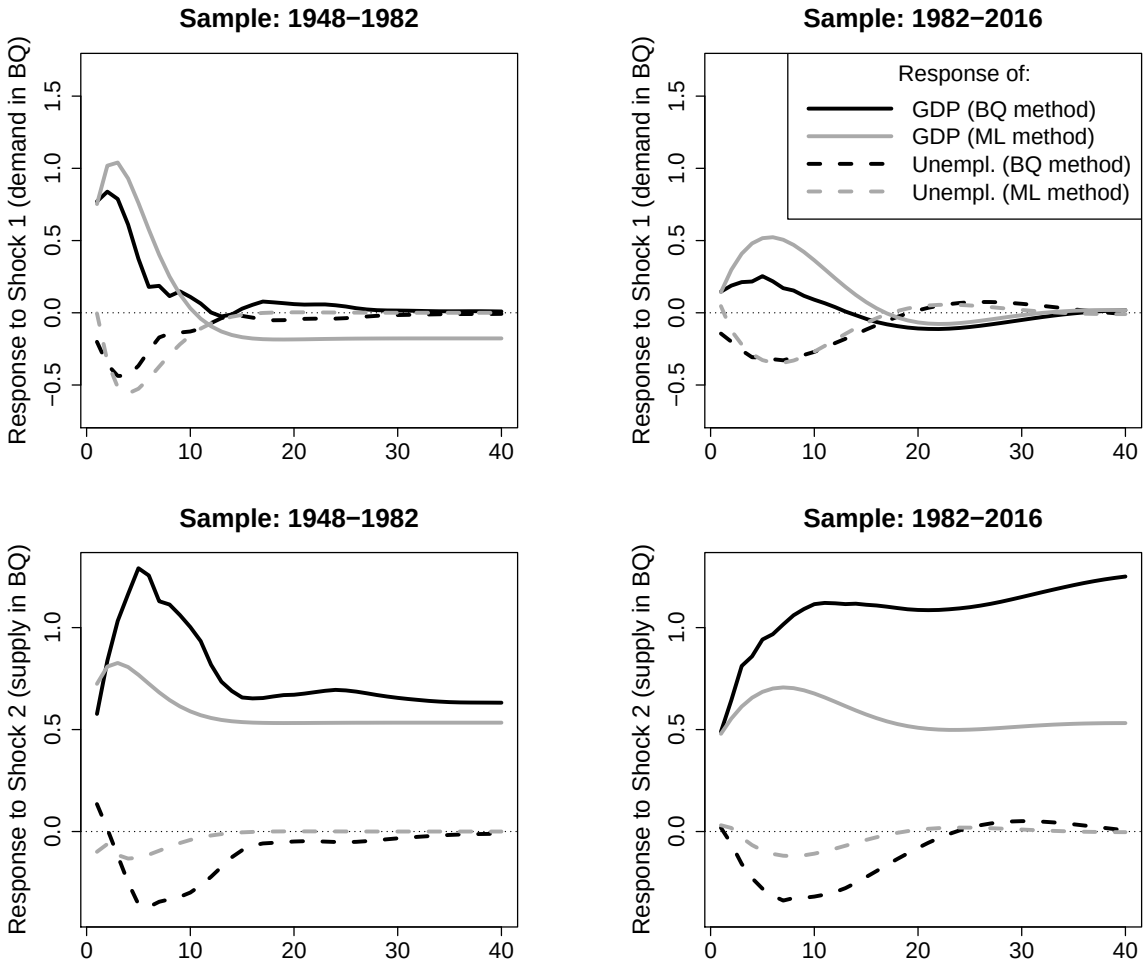
*Notes:* This table presents the results of variance decompositions associated with the two VARMA(1,1) models whose estimation results are documented in Table 3. It reports the shares of the variances of the endogenous variables accounted for by Shock 1 for different horizons; ML results are compared to those resulting from the implementation of the long-run restriction approach proposed by Blanchard and Quah (1989), where Shock 1 is the demand shock. In the ML approach, Shock 1 is defined as the shock having the lower influence on the long-run GDP variance. While the first endogenous variable of the VARMA model is $\Delta gdp$, the variance decomposition is conducted on $gdp$.

Figure 3: Estimated shock distributions (Gaussian mixtures)




*Notes:* This figure displays the estimated distributions of the structural shocks $\eta_{j,t}$, for $j \in \{1,2\}$ and for the two considered sample periods. The dotted lines correspond to the distributions resulting from the Maximum Likelihood approach. These distributions are Gaussian mixtures: Specifically, $\eta_{j,t}$ is drawn from $\mathcal{N}(\mu_{j,1}\sigma^2_{j,1})$ with probability $p = 0.5$ and from $\mathcal{N}(\mu_{j,2}\sigma^2_{j,2})$ with probability $1-p = 0.5$; $\mu_{j,2}$ and $\sigma_{j,2}$ are computed so as to have $E(\eta_{j,t}) = 0$ and $V(\eta_{j,t}) = 1$. Parameters are shown in Table 3. The black solid lines correspond to kernel density estimates of the distribution of the (estimated) structural shocks $\eta_{j,t}$. The latter are computed by applying the approach presented in Appendix C.

Figure 4: Impulse response functions




*Notes:* This figure compares impulse response functions obtained with the Maximum Likelihood (ML) approach and compares them with those resulting from long-run restrictions *à la* Blanchard and Quah (1989), referred to as BQ. While the ML approach is based on a VARMA(1,1) model, the BQ approach involves a 8-lag VAR model and structural shocks are identified by means of long-run restrictions. In the latter approach, whereas Shock 1 (first row of plots) is interpreted as a demand shock, Shock 2 (second row of plots) is interpreted as a supply shock. In BQ, by construction, the long-run impact of the demand shock (Shock 1) on GDP is null. In the ML approach, Shock 1 is defined as the shock having the lower influence on the long-run GDP variance. The three columns of plots correspond to three different estimation samples.

# References

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.

Bera, A. K., and C. M. Jarque, 1981, "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals: Monte Carlo Evidence," *Economics Letters*, 7(4), 313–318.

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., and D. Quah, 1989, "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79(4), 655–73.

Bolt, J., and J. L. van Zanden, 2014, "The Maddison Project: Collaborative Research on Historical National Accounts," *Economic History Review*, 67(3), 627–651.

Box, G. E. P., and G. Jenkins, 1970, *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.

Brockwell, P. J., and R. A. Davis, 1991, *Time Series: Theory and Methods*. 2nd ed., Springer-Verlag New York, Inc., New York, NY, USA.

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 Noninvertible 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.

Davis, R. A., and L. Song, 2012, "Noncausal Vector AR Processes with Application to Economic Time Series," working paper.

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.

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., and A. Jihoud, 2012, "News Shocks Information Flows and SVARs," TSE working paper 12-286, TSE.

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, 2010, "Fiscal Foresight and the Effects of Government Spending," Discussion Paper DP7840, CEPR.

Forni, M., L. Gambetti, M. Lippi, and L. Sala, 2017, "Noisy News in Business Cycle," *American Economic Journal: Macroeconomics*, forthcoming.

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.

Friedlander, B., and B. Porat, 1990, "Asymptotically Optimal Estimation of MA and ARMA Parameters of Non-Gaussian Processes of Higher-Order Moments," *IEEE Transactions on Automatic Control*, 35, 27–35.

Gagliardini, P., and C. Gouriéroux, 2013, "Identification by Laplace Transforms in Nonlinear Panel or Time Series Models with Unobserved Stochastic Dynamic Effects," Working paper, CREST.

Gali, J., 1999, "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," *American Economic Review*, 89(1), 249–271.

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.

Giannakis, G., and J. Mendel, 1989, "Identification of Nonminimum Phase Systems Using Higher Order Statistics," *IEEE Transactions, Acoustics, Speech and Signal Processing*, 37, 360–377.

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.

——— , 2017, "Noncausal Vector Autoregressive Process: Representation, Identification and Semi-Parametric Estimation," *Journal of Econometrics*, forthcoming.

Gouriéroux, C., and A. Monfort, 1996, *Simulation Based Econometric Methods*. Oxford University Press edn.

——— , 2016, "Composite Indirect Inference with Application to Corporate Risks," Working paper, CREST.

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.

Hannan, E., and M. Deistler, 1996, *The Statistical Theory of Linear Systems*. Wiley, New York edn.

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.

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.

Leeper, E., T. Walker, and S. Yang, 2013, "Fiscal Foresight and Information Flows," *Econometrica*, 81, 1115–1145.

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.

Lof, M., 2013, "Noncausality and Asset Pricing," *Studies in Nonlinear Dynamics and Econometrics*, 17, 211–220.

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.

Mountford, A., and H. Uhlig, 2009, "What Are the Effects of Fiscal Policy Shocks?," *Journal of Applied Econometrics*, 24, 960–992.

Na, Y., K. Kim, F. Song, and T. Kim, 1995, "Identification of Nonminimum Phase FIR Systems Using the Third and Fourth Order Cumulants," *IEEE Transactions on Signal Processing*, 43, 2018–2022.

Pesaran, H., and R. Smith, 2011, "Beyond the DSGE Straitjacket," Discussion Paper 1138, Cambridge University.

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.

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.

Skitovich, V., 1953, "On a Property of a Normal Distribution," *Doklady Academii 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.

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.

Varin, C., N. Reid, and D. Firth, 2011, "An Overview of Composite Likelihood Methods," *Statistica Sinica*, 21, 5–42.

Velinov, A., and W. Chen, 2014, "Are There Bubbles in Stock Prices? Testing for Fundamental Shocks," Dp diw berlin, DIW.