Identification and Estimation in Non-Fundamental Structural VARMA Models*

Christian Gouriéroux
CREST, Uni. of Toronto and TSE

Alain Monfort
CREST and Banque de France

Jean-Paul Renne
Uni. of Lausanne

(Revised version, May 2018)

Abstract

The basic assumption of a structural vector autoregressive moving-average (SVARMA) model is that it is driven by a white noise whose components are independent and can be interpreted as economic shocks, called "structural" shocks. When the errors are Gaussian, independence is equivalent to non-correlation and these models face two identification issues. The first identification problem is "static" and is due to the fact that there is an infinite number of linear transformations of a given random vector making its components uncorrelated. The second identification problem is "dynamic" and is a consequence of the fact that, even if a SVARMA admits a non invertible moving average (MA) matrix polynomial, it may feature the same second-order dynamic properties as a VARMA process in which the MA matrix polynomials are invertible (the fundamental representation). 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 that accommodate 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, Structural Shocks, Impulse Response Function.

*The first author gratefully acknowledges financial support of the chair ACPR/Risk Foundation: Regulation and Systemic Risks. We thank M. Barrigozi, M. Deistler, M. Dungey, A. Guay, A. Hyvärinen, J. Jasiak, M. Lippi, H. Lütkepohl, N. Meddahi, A. Melino, H. Uhlig for helpful comments. We also thank participant to the 2017 European Meeting of the Econometric Society and to the 2017 annual congress of the Swiss Society of Economics and Statistics. 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, called "structural" shocks. When the errors are Gaussian, independence is equivalent to non-correlation and these models have to face two kinds of identification problems.

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

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

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 parametric and semi-parametric estimation approaches that accommodate non-fundamentalness in the multivariate moving-average dynamics.

In Section 2, we consider a vector autoregressive moving average process, with roots of the moving average polynomial that are not necessarily outside the unit circle. We stress that the economic shocks are not necessarily interpretable in terms of causal 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 point out that the standard Box-Jenkins approach based on Gaussian pseudo-likelihood suffers from the same 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; we also discuss the identification of the structural shocks and of the Impulse Response Functions (IRFs). In Section 4 we present new parametric and semi-parametric estimation methods to improve upon the standard SVAR methodology. A key element is an algorithm aimed at estimating the structural shocks from samples of endogenous variables; this procedure is shown to provide consistent estimates of the shocks, irrespective of the moving-average fundamentalness regime. The algorithm further makes it possible to compute truncated log-likelihood functions, opening the door to Maximum Likelihood (ML) estimation. We also propose a consistent 2-step semi-parametric approach that is less subject to the curse of dimensionality than the ML approach.

Applications are provided in Section 5. First, we conduct a Monte-Carlo analysis aimed at illustrating the performances of our estimation approaches in the context of a univariate MA(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. GNP 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. Proofs are gathered in Appendix B. Appendix C provides details about the semi-parametric approach and its asymptotic properties.

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 a n-dimensional vector of observations at date t, ε_t is a n-dimensional vector of errors, L the lag operator,

$$\Phi(L) = I - \Phi_1 L - \ldots - \Phi_p L^p, \qquad \Theta(L) = I - \Theta_1 L - \ldots - \Theta_q L^q,$$

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

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

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

i) The process (ε_t) is a square-integrable strong white noise, i.e. the errors ε_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 η_t are mutually independent, and have unit variance, i.e. $V(\eta_{j,t}) = 1$, $j = 1, \ldots, n$.

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 independent random variables $\eta_{j,t}$, $j = 1, \ldots, n$, are usually called "structural shocks" and the following representation is called a "structural" VARMA, or SVARMA, representation:

$$\Phi(L)Y_t = \Theta(L)C\eta_t. \tag{2.2}$$

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

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

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

The next assumption, on the roots *det* $\Phi(z)$, is also made to simplify our analysis.[2]

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

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

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

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

The observable process is the stationary solution of model (2.1) associated with the true values of Φ*,* Θ*,* C *and with the true distribution of* ε*.*

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

[1] 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.

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

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

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

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

and

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

where $\Theta^{-1}(L)$ is the one-sided series operator involving positive powers of L and that satisfies $\Theta^{-1}(L)\Theta(L) = I$. In this case, we say that the operator $\Theta(L)$ is invertible and that the SVARMA model (2.2), is fundamental.

However, 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 roots on both sides of the unit circle, we get a two-sided autoregressive representation:

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

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 (unique) two-sided series operator satisfying $\Theta^{-1}(L)\Theta(L) = I$. In this case, we say that $\Theta(L)$ is invertible in a general sense.

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

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

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

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

and, therefore, y_t is a function of the present and past values of ε_t.

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

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

Formula (2.5) implies that ε_t is a function of the present and future values of y_t. Therefore, ε_t is not the causal innovation of y_t, defined by $y_t - E(y_t|y_{t-1}, y_{t-2}, \dots)$, the latter being a function of present and past values of y_t only.

Since (2.4) implies that the knowledge of $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$ results in the knowledge of $\{y_t, y_{t-1}, \dots\}$, but because (2.5) shows that ε_t is not a function of $\{y_t, y_{t-1}, \dots\}$, it comes that the information set $\{y_t, y_{t-1}, \dots\}$ is strictly included in the information set $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$.

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

Consider a non-fundamental SVARMA process (Y_t) defined by (2.2). (Y_t) admits a fundamental VARMA representation of the form: $\Phi(L)Y_t = \Theta^*(L)\varepsilon_t^*$.[5] Process (ε_t^*) is a weak white noise, i.e.

[3] 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. 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.

[4] 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.

[5] This representation can be obtained from the non-fundamental SVARMA representation by making use of Blaschke matrices. Consider a square matrix of the lag operator denoted by $B(L)$. $B(L)$ is a Blaschke matrix if

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

2.2 Examples of non-fundamentalness

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

i) Lagged impact. A 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 ε_t denotes the productivity shock, reflecting for instance the introduction of an innovation. It may be realistic to assume that the maximal impact of the productivity shock is not instantaneous and is maximal with a lag, i.e. that $\theta > 1$ (see Appendix A.5). The MA(1) process is then non-fundamental (or non-invertible).

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

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

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

[6] This is easily illustrated in the context of a univariate MA(1) process, see Appendix A.5.

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

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

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

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

Thus the root of the moving-average lag-polynomial is equal to $1/aR$, and it is smaller than one when $aR > 1$.[7]

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 to the economic agent 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 β and θ, this root is larger or smaller than 1. When the root is strictly smaller than 1, the model is non-fundamental. In such rational expectation models, the information sets of the economic agent and of the econometrician are assumed to be aligned.

iv) Lagged information and news shocks. 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,

[7] 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.

Walker, and Yang, 2013). A stylized model is (see Fève, Matheron, and Sahuc, 2009):

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

where ε_t is a white noise, and E_t is the conditional expectation given ε_t, ε_{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.

A related situation is that of models that incorporate news shocks. In these models, the demeaned growth rate of technology typically follows a process of the form:

$$y_t = \varepsilon_t + u_{t-q},$$

where ε_t is a standard technology shock and u_t is a news shock, in the sense that agents in the economy see it in period t, but it has no effect on the level of technology until date $t+q$, with $q \geq 1$. When incorporated in general equilibrium models, Sims (2012) shows that such a specification for the technology process may imply that vectors of the form $[y_t, x_t]'$, where x_t is another model variable, say output, follow a non-fundamental process (see also Blanchard, L'Huillier, and Lorenzoni, 2013).

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

2.3 The limits of the Gaussian approach

2.3.1 Static identification issue

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

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

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

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

In order to solve this identification issue, additional short-run, long-run, or sign restrictions have been imposed in the literature [see e.g. the references in the introduction].[8] It turns out that, if at most one of the components of η_t is Gaussian, the identification problem disappears. This result, shown by Comon (1994) (Theorem 11) is a consequence of the Darmois-Skitovich characterization of the multivariate Gaussian distribution (see Darmois, 1953; Skitovich, 1953; Ghurye and Olkin, 1961). In this case, C can be estimated using Independent Component Analysis (ICA) algorithms (see Hyvärinen, Karhunen, and Oja, 2001), or Pseudo Maximum Likelihood techniques (see Gouriéroux, Monfort, and Renne, 2017). These estimators of C are consistent, up to sign change and permutation of its columns. This (quasi) identifiability of the static parameter C in the non-Gaussian case implies that the recursive approach proposed by Sims, imposing that C is lower-triangular, cannot be used in general to find independent shocks, but only uncorrelated shocks. The importance of the independence assumption for the computation of IRFs is discussed in Subsection 3.2.

2.3.2 Dynamic identification issue

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

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

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

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

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

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

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 Φ_1, ..., Φ_p are identifiable from the distribution of process (Y_t) (Gaussian or not), but several sets of coefficients $(\Theta_1, \ldots, \Theta_q, C)$ yield the same spectral density and the same distribution for the process (Y_t) in the Gaussian case; the different polynomials $\Theta(L)$ are obtained from the fundamental one –the one with the roots of $\det(\Theta(z))$ outside the unit circle– by use of Blaschke matrices.[9] The lack of identification of the dynamic parameters Θ is called the *dynamic identification problem*. We see in Section 3 that this second identification problem also disappears in the non-Gaussian case.

The dynamic identification problem is simply illustrated in the univariate MA(1) model: $y_t = \sigma\eta_t - \theta\sigma\eta_{t-1}$, where (η_t) is a Gaussian white noise with $E(\eta_t) = 0$, $V(\eta_t) = 1$ and, for instance, $0 < \theta < 1$. If we replace θ by $\theta^* = 1/\theta$ and σ 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 (θ, σ) and $(1/\theta, \sigma\theta)$ give the same distributions for processes (y_t) and (y_t^*). By contrast, we will see that, if η_t is non-Gaussian, the distributions of processes (y_t) and (y_t^*) are different, although their spectral density matrices are the same (see e.g. Weiss, 1975; Breidt and Davis, 1992; Lii and Rosenblatt, 1992).

Alternatively, the process (η_t^*) defined by $y_t = \sigma\theta\eta_t^* - \sigma\eta_{t-1}^* \equiv \sigma^*(\eta_t^* - \frac{1}{\theta}\eta_{t-1}^*)$, with $\sigma^* = \sigma\theta$ is another Gaussian white noise with zero mean and unit variance. More generally in the MA(q) case, $y_t = \theta(L)\sigma\eta_t$, with $\theta(0) = 1$, we get other representations by replacing $\theta(L)$ by $\theta^*(L)$ obtained from $\theta(L)$ by inverting some roots (the complex roots being inverted by pairs) and replacing

[9] See Footnote 5 for the definition of Blaschke matrices.

σ by σ^* giving the same variance to y_t. Then the processes (η_t^*) defined by $y_t = \theta^*(L)\sigma^*\eta_t^*$ are Gaussian white noises with zero mean and unit variance. Among all these equivalent representations, one is fundamental. In the non-Gaussian case, if one of these (η_t^*) processes is a strong white noise, i.e. a serially independent process, the others will be only weak white noises, i.e. serially uncorrelated. If the strong white noise process does not correspond to the fundamental representation, the weak white noise corresponding to the invertible polynomial $\theta^*(L)$ is the linear innovation process associated with the Wold representation.

In the usual Box-Jenkins approach, the estimation of the parameters Φ_1, ..., Φ_p, Θ_1, ..., Θ_q, $\Sigma = CC'$ is based on a truncated VAR approximation relying on the assumption that $\Theta(L)$ is invertible (i.e. the roots of $\det(\Theta(z))$ are outside the unit circle), namely a truncation of $\Theta(L)^{-1}\Phi(L)Y_t = \varepsilon_t$, with $V(\varepsilon_t) = \Sigma = CC'$ (see e.g. Galbraith, Ullah, and Zinde-Walsh, 2002). In other words, a fundamental representation is a priori imposed without being tested. The introduction of multivariate non-fundamentalness tests is actually very recent (see e.g. Forni and Gambetti, 2014; Chen, Choi, and Escanciano, 2017).[10]

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:

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

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

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

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

In the 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 Θ on the unit circle. However, under Assumption A.5, 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

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

$B(L)\Theta(L) = I$.

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

Using the same argument as in Subsection 2.3.1, namely that $\Phi_1, \ldots, \Phi_p$ are characterized by the Yule-Walker equations, we know that $\Phi(L)$ is identifiable. What about $\Theta(L)$ and C? The next proposition is deduced from Theorem 1 in Chan, Ho, and Tong (2006) (based on Theorem 4 in Chan and Ho, 2004), and solves the dynamic identification issue in the non-Gaussian case.[11,12] Let us first introduce the following assumption:

Assumption A.6. *Each component of η_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.*

Assumption A.6 is introduced to eliminate the Gaussian framework in which all cumulants of order $r \geq 3$ are zero. Assumption A.6 is satisfied if the distribution of $\eta_{j,t}$ is skewed and has a finite moment of order 4, or if it is symmetric, but has a kurtosis different from 3 and has a finite moment of order 6.

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

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

then the process (Y_t) defined in (2.2) 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.1. □

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

[12] A similar identification result has been recently derived when the components of η_t have fat tails (see Gouriéroux and Zakoian, 2015).

There is no other normalisation needed on matrix C because the components $\eta_{j,t}$ are assumed to have a unit variance.

Proposition 1 provides conditions under which the SVARMA parameterization is identified. This identification result is however not constructive and does not explain how to estimate the correct –possibly non-fundamental– SVARMA representation.[13] The latter task is the objective of the methods presented in Section 4 below.

3.2 Identification of the structural shocks and of the IRFs: The independence assumption

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 disturbances η_t are identified up to a permutation and a sign change of their components, not depending on t. In this context, the IRFs corresponding to shocks to the process (η_t) of structural disturbances are also identified; these IRFs are defined from the sequences $A_h = \Psi_h C$, where matrices Ψ_h depend on $\Phi(L)$ and $\Theta(L)$ (see eq. 2.3). The only remaining problem is the labelling of these shocks. The labelling will depend on the economic interpretation of the IRFs. This is completely different from standard structural approaches where the identification of the shocks themselves necessitates additional theory-based economic restrictions. In the latter case, labelling is tied to the choice of the restrictions. Proposition 1 states that, under Assumptions A.1 to A.6, such restrictions are over-identifying restrictions.

In the rest of this subsection, we highlight the importance of the independence Assumptions A.1.ii) for the derivation of the IRFs. We use the traditional definition of IRF (see e.g. Koop, Pesaran, and Potter, 1996):

Definition 1. *The Impulse Response Function (IRF) of $Y_{i,t}$ to $\eta_{j,t}$ is defined as the sequence of the differential impacts of a unit shock on $\eta_{j,t}$ on the $Y_{i,t+h}$s, $h = 0,1,\ldots$. Formally, this IRF is given by $\{IRF_{i,j,0}, IRF_{i,j,1}, \ldots\}$, with:*

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

where $\underline{Y_t} = \{Y_t, Y_{t-1}, \ldots\}$.

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

The following proposition shows that, under the serial and mutual independence assumptions, i.e. A.1, the IRFs are directly deduced from the AR and the (possibly non-invertible) MA coefficients.

Proposition 2. *Under A.1, the IRFs are given by:*

$$IRF_{i,j,h} = \Psi_{i,h} C^j,$$

where C^j is the j^{th} column of matrix C and where $\Psi_{i,h}$ is the i^{th} row of Ψ_h defined in equation (2.3).

Proof By the iterated expectation theorem, using that the information set $\underline{Y_t}$ is included in the information set $\underline{\eta_t} = \{\eta_t, \eta_{t-1}, \dots\}$, we have:

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

and, by using (2.3):

$$IRF_{i,j,h} = E\left[\Psi_{i,h} C \eta_t | \eta_{j,t} = 1, \underline{Y_{t-1}} \right] = \Psi_{i,h} C E\left[\eta_t | \eta_{j,t} = 1, \underline{Y_{t-1}} \right].$$

Because the η_ts are serially independent, we get

$$IRF_{i,j,h} = \Psi_{i,h} C E\left[\eta_t | \eta_{j,t} = 1 \right].$$

Since the components of η_t are independent and of mean zero, we have:

$$E\left[\eta_t | \eta_{j,t} = 1 \right] = e_j, \tag{3.3}$$

where e_j is the j^{th} selection vector. This leads to the standard impact formula:

$$IRF_{i,j,h} = \Psi_{i,h} C^j. \quad \square$$

The previous proof can easily be adjusted to show that the IRF defined above also corresponds to this alternative formulation:

$$IRF_{i,j,h} = E(Y_{i,t+h}|\eta_{j,t} = 1, \underline{Y_{t-1}}) - E(Y_{i,t+h}|\eta_{j,t} = 0, \underline{Y_{t-1}})$$

and, further, that we have:

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

for any η and δ.

The standard derivation of the impact of a shock on $\eta_{j,t}$ shows the importance of both the serial and cross-sectional independence assumptions. In particular, if the components of η_t were not independent, there would be an instantaneous impact of the shock $\eta_{j,t} = 1$ on the other components $\eta_{i,t}$, $i \neq j$. In other words, the quantity $E\left[\eta_{i,t}|\eta_{j,t} = 1\right]$, $i \neq j$, would not necessarily be zero, contradicting equation (3.3).[14,15]

Hence, the serial and cross-correlation assumptions on the noise are not sufficient for structural applications. A case where errors can be made uncorrelated but not independent notably arises in the recent literature proposing to decompose a series into its short- and long-run components in order to focus on the impact of long-run shocks (see e.g. Bandi, Perron, Tamoni, and Tebaldi, 2018). The decomposition of the series at different frequencies is based on the spectral decomposition of the series. Though uncorrelated, the resulting linear innovations associated with different frequencies depend on a unique underlying i.i.d. noise. It is then impossible to shock separately the short-run and the long-run components of the series.

Proposition 2 also highlights that the IRFs hinge on the fundamentalness regime of the SVARMA model. It shows indeed that the IRF depends on the (possibly non-invertible) MA coefficients through the Ψ_hs. Appendix A.5 shows for instance that, in the MA(1) case, if the IRF associated with the fundamental representation is monotonously decreasing, then the one associated with the non-fundamental representation is hump-shaped.

4 Estimation of Models with Non-Fundamentalness

In this section, we present parametric and semi-parametric estimation approaches of non-fundamental SVARMA models. For a parametric SVARMA model, we show that the parameters can be estimated by the Maximum Likelihood (ML) method. We first discuss the regularity of the likelihood function on the frontier between fundamentalness and non-fundamentalness. Next, we explain how to numerically compute, or approximate, the likelihood function.

[14] Such an adjustment has already been taken into account in previous studies (see e.g. Pesaran and Shin, 1998; Jardet, Monfort, and Pegoraro, 2013).

[15] To illustrate this point, let us introduce a stylized example. Consider the case where $\eta_{1,t}$ is drawn from a uniform distribution on $[-\sqrt{3}, \sqrt{3}]$, such that $E(\eta_{1,t}) = 0$ and $V(\eta_{1,t}) = 1$ and where $\eta_{2,t}$ is independently drawn from $\mathcal{N}(0,1)$. Now consider the vector $[\eta^*_{1,t}, \eta^*_{2,t}]'$ obtained by rotating η_t by an angle of $-\pi/4$. We have that $\eta^*_{1,t}$ and $\eta^*_{2,t}$ are uncorrelated, zero-mean and with unit variance, but not independent. In particular, we have $E(\eta^*_{2,t}|\eta^*_{1,t}) > 0$ for $\eta^*_{1,t} > \sqrt{3}$, which derives from the fact that the support of $[\eta^*_1, \eta^*_2]'$ is the area delineated by the two curves defined by $y = x - \sqrt{3}$ and $y = x + \sqrt{3}$.

When the distribution of the errors is left unspecified, Proposition 1 suggests better semi-parametric estimation methods than the inconsistent Gaussian pseudo-maximum likelihood used in the standard Box-Jenkins methodology. These alternative methods provide consistent estimators of the true values of $\Phi(L)$, $\Theta(L)$ and C in a SVARMA representation where the roots of $\det(\Theta(z))$ may be inside or outside the unit circle. The idea is to introduce appropriate moment restrictions deduced from the independence assumption on the components of error η and derive associated moment methods for estimation. This semi-parametric approach is robust to a misspecification of the error distribution.

When the dimension of the vector of endogenous variables Y_t increases, the number of parameters specifying the SVARMA representation (2.2) increases at a much faster rate, giving rise to a curse of dimensionality problem. As a result, in practice, there is a tradeoff between the dimension n and the degrees p and q in VARMA modelling. In most applications, q is equal to 0, 1 or 2. In the following, the estimation methods are presented for VARMA(p,1) models. They can be extended to VARMA(p,q) models for $q > 1$. Contrary to our ML approach, the numerical complexity of our semi-parametric approach does not depend on the autoregressive order p. Nevertheless, both methods are subject to the curse of dimensionality when augmenting the moving-average order q.

4.1 Maximum Likelihood (ML) estimation of parametric SVARMA models

For illustrative purpose, let us first discuss the case of a one-dimensional MA(1) process before considering the general framework of a SVARMA process. Derivations of truncated log-likelihood functions and associated asymptotic results of ML estimators in the context of possibly noninvertible univariate MA(q) and ARMA(p,q) processes can notably be found in Lii and Rosenblatt (1992, 1996), or Wu and Davis (2010). The approach exposed in Subsection 4.1.2 can be seen as an extension of these previous studies to the multivariate case.

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

We consider the MA(1) process:

$$y_t = \varepsilon_t - \theta\varepsilon_{t-1}, \tag{4.1}$$

where the ε_ts are independent.

Suppose that we observe $\{y_1, \ldots, y_T\}$. If the common distribution of the ε_ts is $N(0, \sigma^2)$, the model is not identifiable (see Section 2.3). If ε_t satisfies Assumptions A.1 to A.6, i.e. in particular if it is not Gaussian, then Proposition 1 states that the model is identifiable.[16] Let us denote by

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

$g(\varepsilon;\gamma)$ the common p.d.f. of the ε_ts, where γ is an unknown parameter, and let us consider three cases, depending on the position of $|\theta|$ with respect to 1:

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

$$\varepsilon_t = \sum_{h=0}^{\infty} \theta^h y_{t-h}. \tag{4.2}$$

Using the notation $y_1^T = \{y_1,\dots,y_T\}$, the truncated log-likelihood function is:

$$L_1(y_1^T;\theta,\gamma) = \sum_{t=1}^{T} \log g\left(\sum_{h=0}^{t-1} \theta^h y_{t-h};\gamma\right), \tag{4.3}$$

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

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

$$\begin{aligned} y_t &= \varepsilon_t - \theta\varepsilon_{t-1} \\ \Leftrightarrow \quad -\frac{y_{t+1}}{\theta} &= \varepsilon_t - \frac{1}{\theta}\varepsilon_{t+1} \\ \Leftrightarrow \quad \varepsilon_t &= -\sum_{h=0}^{\infty} \frac{1}{\theta^{h+1}} y_{t+h+1}. \end{aligned} \tag{4.4}$$

The truncated log-likelihood function is then:

$$L_2(y_1^T;\theta,\gamma) = \sum_{t=1}^{T} \log\left\{\frac{1}{|\theta|} g\left(-\sum_{h=0}^{T-t-1} \frac{1}{\theta^{h+1}} y_{t+h+1};\gamma\right)\right\}, \tag{4.5}$$

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

Focussing on the regimes when truncating the log-likelihood function gives the misleading impression of a lack of continuity of the exact log-likelihood function w.r.t. θ at $|\theta|=1$. This

exact log-likelihood is however continuous.[17] Indeed, we have:

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

Thus the joint p.d.f. of $\{y_1, \ldots, y_T\}$ given ε_0 is:

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

and the exact log-likelihood is:

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

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

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

is only right-differentiable at $\theta = 1$. In practice, however, using the truncated log-likelihood (4.7) is easier because it does not involve the computation of an integral as in the case of the exact log-likelihood (4.6), while providing estimators with the same asymptotic properties.

To conclude, in the simple MA(1) case, the maximum likelihood estimation can be conducted by maximizing the truncated log-likelihood function (4.7). If $|\theta| \neq 1$, the standard asymptotic theory applies (see e.g. Davidson, 1994). This is however not the case if $|\theta| = 1$.

4.1.2 The Maximum Likelihood approach in the VARMA context

Let us consider the VARMA(p,1) model:

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

where the errors ε_t are given by:

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

[17] 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.

where the η_t are serially and mutually independent, with $E(\eta_t) = 0$ and $V(\eta_t) = I$. The distribution of the η_ts is non-Gaussian and parameterized with γ, say. Therefore, the p.d.f. of the errors ε_t is of the form $g(\varepsilon, \Gamma)$, with $\Gamma = (C, \gamma)$. This model is supposed to satisfy Assumptions A.1 to A.6.

To get an approximation of the likelihood function, we need to compute the filtered values of the errors ε_t for a given parameterization of the model. To this purpose, we exploit the real Schur decomposition of matrix Θ:[18]

$$\Theta = AUA' = A \begin{bmatrix} U_1 & U_{1,2} & \dots & & U_{1,K} \\ 0 & U_2 & U_{2,3} & \dots & U_{2,K} \\ \vdots & \ddots & \ddots & & \vdots \\ & & 0 & U_{K-1} & U_{K-1,K} \\ 0 & \dots & & 0 & U_K \end{bmatrix} A', \tag{4.9}$$

where A is orthogonal, U is upper block-triangular, and the diagonal blocks (U_k, $k \in \{1, \dots, K\}$) are either 1×1 or 2×2 blocks, the 2×2 blocks corresponding to complex conjugate complex eigenvalues of Θ. Under Assumption A.5, the roots of U are not on the unit circle. We denote by n_k the dimension of U_k (with $n_k \in \{1, 2\}$). We assume, without loss of generality, that $U_1, \dots U_q$ have eigenvalues with modulus strictly larger than 1 whereas $U_{q+1}, \dots U_K$ have eigenvalues with modulus strictly lower than 1.[19]

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

$$W_t = \varepsilon_t^* - U\varepsilon_{t-1}^*, \tag{4.10}$$

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

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 $m = n_1 + \dots + n_q$. 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 (4.10) writes:

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

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

[19] In standard softwares, such as R or Matlab, eigenvalues of U are ordered w.r.t. their modulus.

which leads to:

$$\varepsilon_t^{*(2)} = W_t{}^{(2)} + U^{(2)} W_{t-1}{}^{(2)} + \cdots + U^{(2)^{t-1}} W_1{}^{(1)} + U^{(2)^t} \varepsilon_0^{*(2)} \tag{4.11}$$

$$\begin{aligned}\varepsilon_t^{*(1)} = {} & \left\{(U^{(1)})^{-1}\right\}^{T-t} \varepsilon_T^{*(1)} - (U^{(1)})^{-1} W_{t+1}{}^{(1)} - \cdots - \left\{(U^{(1)})^{-1}\right\}^{T-t} W_T{}^{(1)} \\ & - \left\{(U^{(1)})^{-1}\right\}^{2} U^{(12)} \varepsilon_{t+1}^{*}{}^{(2)} - \cdots - \left\{(U^{(1)})^{-1}\right\}^{T+1-t} U^{(12)} \varepsilon_T^{*(2)}.\end{aligned} \tag{4.12}$$

Equation (4.11) shows that, once $\varepsilon_0^{*(2)}$ is known, the $\varepsilon_t^{*(2)}$'s can be computed by forward recursions. Once all the $\varepsilon_t^{*(2)}$'s are computed, and if $\varepsilon_T^{*(1)}$ is known, the $\varepsilon_t^{*(1)}$ can then be obtained by backward recursions using (4.12). The following proposition directly derives from the observation that the eigenvalues of $(U^{(1)})^{-1}$ and $U^{(2)}$ lie inside the unit circle.

Proposition 3. *For periods t that are far enough from sample boundaries (0 and T), ε_t can be approximated by $\hat{\varepsilon}_t \equiv A e_t^*$, with $e_t^* = \left[e_t^{*(1)\prime}, e_t^{*(2)\prime}\right]'$, where $e_t^{*(1)}$ and $e_t^{*(2)}$ are the respective approximations of $\varepsilon_t^{*(1)}$ and $\varepsilon_t^{*(2)}$ obtained by applying (4.11) and (4.12) with $\varepsilon_0^{*(2)} = \varepsilon_T^{*(1)} = 0$.*

Proof See Appendix B.2. □

The previous proposition calls for three remarks.

First, because of the two-sided nature of the filtering algorithm underlying Proposition 3, the quality of the approximations of ε_t is relatively poor at both ends of the sample when there are roots of $\det(\Theta(z))$ on both sides of the unit circle.

Second, when the different roots of $\det(\Theta(z))$ are on each side of the unit circle, none of the components of ε_t corresponds, in general, to the components of the causal innovations. To see this, consider the bivariate case and a VMA(1) process:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} - \Theta \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix},$$

where Θ is of the form:

$$\Theta = \underbrace{\begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix}}_{=A} \begin{bmatrix} u^{(1)} & u^{(12)} \\ 0 & u^{(2)} \end{bmatrix} \begin{bmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{bmatrix}.$$

In this case, the fact that the roots of $\det(I - \Theta z)$ are on each side of the unit circle implies that $|u^{(1)}| > 1$ and $|u^{(2)}| < 1$. In the general case where $u^{(2)} \neq 0$, the asymptotic versions of (4.11) and (4.12) show that ε_t^* is function of both past and future values of W_t, and therefore of Y_t. Since

$\varepsilon_t = A\varepsilon_t^*$, and exluding the specific case where ω is a multiple of $\frac{\pi}{2}$, we then obtain that both $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ depend on past and future values of Y_t.

A third remark, that relates to the previous one, is that (4.11) and (4.12) can be exploited to shed light on the conditions under which a (structural) shock of interest can be approximated by means of a VAR model (see the recent studies by Sims, 2012; Forni and Gambetti, 2014; Beaudry, Fève, Guay, and Portier, 2015). Take for instance the previous bivariate case and suppose that the shock of interest is the first structural shock, that is $\eta_{1,t} = \gamma_1'\varepsilon_t$, where we denote by γ_1' the first row of matrix C^{-1}. A necessary condition for this shock to be well approximated by means of a VAR model is that $\eta_{1,t}$ is included in the information set $\{Y_t, Y_{t-1}, \dots\}$. Assuming that $|u^{(1)}| > 1$ and $|u^{(2)}| < 1$, this is the case if and only if $\gamma_1'A\varepsilon_t^*$ does not depend on $\varepsilon_t^{*(1)}$, that is if γ_1 is orthogonal to $[\cos(\omega), -\sin(\omega)]'$. The previous reasoning suggests that this is only for a restricted set of matrices C that one can approximate a structural shock of interest by means of VAR models.

Let us denote by $\mathscr{L}(Y_1^T;\Phi,\Theta,\Gamma)$ the log-likelihood associated with the observations $Y_1^T \equiv \{Y_1,\dots,Y_T\}$, where (Y_t) follows (4.8). Proposition 3 opens the door to the computation of a truncated log-likelihood function:

Proposition 4.

$$\frac{1}{T}\mathscr{L}(Y_1^T;\Phi,\Theta,\Gamma) \approx \frac{1}{T}L(Y_1^T;\Phi,\Theta,\Gamma) = -\sum_{k=1}^{K}\log|\det(U_k)|\mathbb{1}_{|\det(U_k)|\geq 1} + \frac{1}{T}\sum_{t=1}^{T-1}\log g(\hat{\varepsilon}_t,\Gamma), \tag{4.13}$$

where the $\hat{\varepsilon}_t$'s are the approximations of the ε_ts defined in Proposition 3.

Proof See Appendix B.2. □

The first term on the right-hand side of the previous equation is equal to the opposite of the sum of the logarithms of the moduli of the eigenvalues of Θ whose modulus is larger than one and, therefore, this term does not depend on the (Schur) decomposition of matrix Θ.

If $\Theta(L)$ is of order $q > 1$, one can go back to the previous case. For this, define $\tilde{\varepsilon}_t = [\varepsilon_t', \dots, \varepsilon_{t-q+1}']'$ and $\tilde{Z}_t = [Z_t', \dots, Z_{t-q+1}']'$, where $Z_t = \Phi(L)Y_t$. Using obvious notations, we then have: $\tilde{Z}_t = (I - \tilde{\Theta}L)\tilde{\varepsilon}_t$. Note that the eigenvalues of $\tilde{\Theta}$ are the roots of $\det\Theta(z)$ (see Davis and Song, 2012; Gouriéroux and Jasiak, 2017, where the problem is completely treated in the dual case, where the roots of $\det\Phi(z)$ can be inside or outside the unit circle). Hence, the computation of the approximated log-likelihood function can be performed in the general VARMA(p,q) case.

Because of the discontinuity of the truncated log-likelihood function when the roots of $\det(\Theta(z))$ are on the unit circle, the numerical maximization of the truncated log-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. An alternative, or complementary, approach consists in running additional numerical optimizations with starting values corresponding to models featuring the same spectral density as the one resulting from a preliminary-estimated model, but with different fundamentalness regimes; such starting values can be obtained by applying Blaschke-based transformations to a preliminary-estimated model (see e.g. Lippi and Reichlin, 1994). We employ this method in our empirical applications.

4.2 Semi-parametric estimation of a non-fundamental SVARMA model

Let us consider a SVARMA(p,1) model:

$$Y_t = \mu + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + C_0 \eta_t + C_1 \eta_{t-1}, \tag{4.14}$$

where the components of η_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 outside the unit circle, but the roots of the determinant of the moving average polynomial may be inside or outside the unit circle. We denote by f_j the common probability density function of the independent $\eta_{j,t}$'s, $t = 1,\ldots,T$. We have to consistently estimate the (true values of the) parameters μ, Φ_1, ..., Φ_p, C_0, C_1 as well as the (true) functional parameters f_j, $j = 1,\ldots,n$. We consider below a 2-step moment method.

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

Once μ and Φ are estimated, the associated residuals

$$\hat{Z}_t \equiv Y_t - \hat{\mu} - \hat{\Phi}_1 Y_{t-1} - \cdots - \hat{\Phi}_p Y_{t-p}, \tag{4.15}$$

are consistent approximations of $Z_t = C_0\eta_t + C_1\eta_{t-1}$. Then in a second-step we proceed with the estimation of C_0 and C_1 in a pure moving average framework.[20]

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

4.2.1 Moment restrictions for C_0, C_1

Let us consider the estimation of the moving average parameters as if the true Z_t's were observed (this is relaxed in the next subsection). To address the lack of second-order identifiability of C_0, C_1, the econometric literature has proposed to exploit moments of order 3 and/or 4 (see e.g. Bonhomme and Robin, 2009, in the special case $C_1 = 0$, Gospodinov and Ng, 2015, in the one-dimensional case, Lobato and Velasco, 2018, in the one-dimensional case through bi-spectral and tri-spectral density functions).

Loosely speaking, it is appropriate to consider (cross) moments of order 3 if some components of η_t are skewed (i.e. with non-zero third-order cumulants κ_{3j}) and it is appropriate to use (cross) moments of order 4 if η_t features kurtotic components (with non-zero fourth-order cumulants κ_{4j}). The description of the moment conditions associated with cumulants is provided in Appendix C.2.[21]

This approach leads to a set of moment restrictions:

$$E\left[h(Z_t, Z_{t-1}; \beta)\right] = 0, \tag{4.16}$$

where

$$\beta = [vec\,C_0{}', vec\,C_1{}', \kappa_{31}, \ldots, \kappa_{3n}, \kappa_{41}, \ldots, \kappa_{4n}]'. \tag{4.17}$$

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

$$E\left([u'Z_t + v'Z_{t-1}]^2\right), \quad E\left([u'Z_t + v'Z_{t-1}]^3\right), \quad \text{or} \quad E\left([u'Z_t + v'Z_{t-1}]^4\right), \tag{4.18}$$

for different pairs (u, v). We get an infinite set of moment restrictions but, in practice, a finite set of pairs has to be selected (see Section 5). Denoting by r the dimension of $h(Z_t, Z_{t-1}; \beta)$, the order condition is, in this case, $r \geq 2n^2 + 2n$. The rank condition is challenging to analyse in the multivariate case.[22] In practice, difficulties in inverting the asymptotic covariance matrix of the estimator –derived in Appendix C.3– constitutes a signal of non-identification.

In the spirit of the approach proposed by Lobato and Velasco (2018) who work in the frequency domain, additional types of moment restrictions might be introduced:[23]

i) We might consider similar moment restrictions based on a different lag order, i.e. $E([u'Z_t +$

[21] See Boudt, Cornilly, and Verdonck (2018) for a counting of up-to-order-4 comoments in the multivariate static case.

[22] Existing results are only for ARMA(p,q) processes (Gospodinov and Ng, 2015, Lemma 3, and Lobato and Velasco, 2018, Theorem 1). In Gospodinov and Ng (2015), the conditions pertain to specific cumulants of order $p+q$; in Lobato and Velasco (2018), the conditions are on multi-dimensional integrals of spectral densities.

[23] The approach of Lobato and Velasco (2018) based on bi-spectral and tri-spectral densities seems difficult to extend in the multivariate framework.

$v'Z_{t-k}]^2)$, $E([u'Z_t + v'Z_{t-k}]^3)$ or $E([u'Z_t + v'Z_{t-k}]^4)$ for some $k > 1$.

ii) We might also consider quantities such as: $E([u'Z_t + v'Z_{t-1} + w'Z_{t-2}]^2)$, $E([u'Z_t + v'Z_{t-1} + w'Z_{t-2}]^3)$ or $E([u'Z_t + v'Z_{t-1} + w'Z_{t-2}]^4)$.

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

4.2.2 2-step moment method

The first step of our approach, namely the 2SLS estimation, provides us with an estimator of $\alpha = [\mu', vec(\Phi)']'$. The second step of the approach involves the sample counterparts of (4.16), after replacement of Z_t by $\hat{Z}_t = Y_t - \hat{\mu} - \hat{\Phi}_1 Y_{t-1} - \cdots - \hat{\Phi}_p Y_{t-p}$. In terms of the observation themselves, the moment restrictions (4.16) become:

$$E\left[h_Y(Y_t, Y_{t-1}, \ldots, Y_{t-p-1}; \alpha, \beta)\right] = 0, \tag{4.19}$$

where

$$h_Y(y_t, y_{t-1}, \ldots, y_{t-p-1}; \alpha, \beta) = h(z_t(\alpha), z_{t-1}(\alpha); \beta), \tag{4.20}$$

with $z_t(\alpha) = y_t - \mu - \Phi_1 y_{t-1} - \cdots - \Phi_p y_{t-p}$.

The asymptotic accuracy of this 2-step moment estimator is derived in Appendix C.3.

4.2.3 Nonparametric estimation of the error distribution

Once $\mu, \Phi_1, \ldots, \Phi_p, C_0$ and C_1 have been estimated, consistent approximations of the errors η_t are obtained by employing the filtering approach of Proposition 3. The p.d.f. f_j can then be estimated by a kernel density estimator applied to the residuals $\hat{\eta}_{j,t}$, $t = 1, \ldots, T$.

5 Applications

5.1 Monte Carlo exercises

This subsection illustrates the performances of the estimation approaches by means of Monte-Carlo experiments. For the sake of simplicity, we focus on a univariate MA(1) processes:

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1}, \tag{4.21}$$

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

We consider different sample sizes ($T = 100$, 200 and 500) and different true distributions of the errors ε_t. Four distributions are used: the Gaussian distribution, a mixture of Gaussian

distributions and two Student distributions with respective degrees of freedom of 5 and 10. In all simulations, we use $\theta = -2$. Hence, data generating processes are non-fundamental.

In order to get the intuition behind our approach, it is instructive to look at the joint distribution of y_t and y_{t-1}. Figure 1 displays contour plots associated with these distributions in the context of the four different distributions used for ε_t. While the black solid lines correspond to the non-fundamental case, the grey lines represent the (pseudo) distribution that would prevail under the fundamental case, i.e. with $\theta = -1/2$ and $V(\varepsilon_t) = 4$. In the purely Gaussian case (Panel (a)), the two distributions coincide, since the two processes are observationally equivalent. By contrast, in the other three cases –Panels (b), (c) and (d)– the two distributions are different. The case of the mixture of Gaussian distributions is particularly illustrative. For this distribution, and in the non-fundamental case, the shock ε_t is drawn from $\mathcal{N}(\mu_1, \sigma_1^2)$ with probability p and from $\mathcal{N}(\mu_2, \sigma_2^2)$ with probability $1-p$. We set: $\mu_1 = -0.7$, $\mu_2 = 0.7$, $\sigma_1 = 0.32$, $\sigma_2 = 0.95$, $p = 0.5$, which results in a zero-mean unit-variance distribution with order-3 and order-4 cumulants of 0.85 and 0, respectively.[24]

For the ML approach (Subsections 4.1), and for each of the four distributions, we estimate five parameters: θ, the variance of ε_t and three parameters specifying a mixture of Gaussian distributions with zero mean and unit variance. As far as the Student distributions are concerned, we proceed under the assumption that we do not know the true distribution, which is generally the case in practice. Hence, when the true distributions are Student, the ML approach is, more precisely, a Pseudo Maximum Likelihood (PML) approach.[25] This exercise is in the spirit of Lii and Rosenblatt (1996) and Wu and Davis (2010), who study the performances of their ML estimators when using a misspecified distribution in the computation of the log-likelihood function.

In the context of our variant of the GMM approach (Subsection 4.2), we consider the order-2, order-3 and order-4 moments given in (4.18), with $(u,v) \in \{(1,0),(1,2),(2,1)\}$.[26] Hence, we use 9 restrictions to estimate, for each model, the four following parameters: θ, the variance of ε_t, κ_3 and κ_4.

In our discussion of the results, we focus on the estimates of θ. Figure 2 shows the distributions of the estimators of θ resulting from both approaches. Each of the four rows of plots corresponds to one of the four considered distributions for ε_t, that are those distributions represented on Figure 1. On each panel, the three curves correspond to the three considered sample sizes ($T = 100$, 200 and 500). The (finite sample) distributions are often bimodal; one mode being close to the true value

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

[25] By contrast, when the true distribution is Gaussian or a Gaussian mixture, the approach is not a PML, but a standard ML approach. (Note that the Gaussian distribution is a special case of Gaussian mixture.)

[26] The reason why we do not take $(u,v) \in \{(1,0),(1,1)\}$ is that, if $(u,v) = (1,1)$, the moments of $uy_t + vy_{t-1}$ are the same as those of $vy_t + uy_{t-1}$, thereby preventing the identification of the parameters. Indeed, if $y_t = \varepsilon_t - \theta\varepsilon_{t-1}$ and $y_t^* = -\theta\varepsilon_t + \varepsilon_{t-1}$, then $y_t + y_{t-1}$ and $y_t^* + y_{t-1}^*$ have the same distribution, and therefore the same moments.

of θ –indicated by a vertical bar on each panel– and the other being close to $1/\theta$.

The results illustrate the fact that, the closer the distribution of the generated shocks is to a normal one, the weaker the identification. Let us focus for instance on the fraction of estimated fundamental processes, i.e. with θ estimates that are lower than 1 in absolute value (whereas the data generating process is non-fundamental since $|\theta| > 1$). In the Gaussian case (Panels a.1 and a.2 of Figure 2), about half of the estimated processes are fundamental, irrespective of the sample size or of the estimation approach. Among the three remaining distributions, the one leading to the largest fractions of estimated fundamental processes is the Student t(10). For the latter distribution and in the ML case, even for a relatively large sample size ($T = 200$), about a third of the estimated θ's have a modulus lower than 1, whereas this proportion is of about 13% in the Student-t(5) case and close to 0% in the Gaussian mixture case. Identification is easier in the Student t(5) and in the Gaussian mixture cases (Panels b.1, b.2, c.1 and c.2) since, in these cases, the differences between the distributions of (y_{t-1}, y_t) in the fundamental versus non-fundamental regimes are more marked than for the other two distributions (as illustrated by Figure 1).

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

To the best of our knowledge, this study is the first one to compare GMM and ML estimates in the context of non-fundamental univariate MA processes. As regards the GMM approach, our results are comparable to the results in the literature, such as the ones reported by Gospodinov and Ng (2015) for $T = 500$, or by Lobato and Velasco (2018) with spectral-density-based estimator for $T = 100$, 200 and Student t(5) distribution.[27]

The last three columns of Table 1 aim at assessing the validity of the asymptotic distribution of the θ estimators. These columns indicate the fractions of times (among the N simulations) where the true value of θ lies within the interval $[\hat{\theta} - \phi_\alpha \sigma_{asy}, \hat{\theta} + \phi_\alpha \sigma_{asy}]$, where σ_{asy} denotes the estimate of the asymptotic standard deviation of the estimator $\hat{\theta}$ and where ϕ_α is such that $P(-\phi_\alpha < X < \phi_\alpha) = \alpha$, if $X \sim \mathcal{N}(0,1)$. The closer to α the reported fractions, the better the asymptotic approximation of the estimator distribution. The results indicate that the inference

[27] Our GMM estimator however appears to be less efficient than the spectral-density-based estimator of Lobato and Velasco (2018): For $T = 200$ and when the true distribution is t(5), Lobato and Velasco (2018) report that the absolute value of their θ estimator is on the right side of 1 in 93% of the cases (last column in their Table 1), while we get an equivalent percentage of 75% with our GMM estimator. This percentage is of 87% for our PML estimator. Our PML RMSEs are close to those reported by Lobato and Velasco (2018) in this case ($T = 200$, t(5) distribution, $\theta = -2$).

based on the estimated asymptotic distribution is less adequate in the GMM context than in the (P)ML one. The fact that the wrong fundamentalness regime is more often obtained in the GMM case than in the (P)ML case accounts for the lower adequacy of GMM-based confidence intervals.

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(1,1) processes assumed to be followed by per capita real GDP growth. We consider long historical samples taken from Bolt and van Zanden (2014).[28] 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 two time-series normality tests: the Bai and Ng (2005) and the Lobato and Velasco (2004) tests. The p-values of these tests are reported in the third column of the table. According to the latter test, the null hypothesis of normality is rejected at the 5% level for all countries.

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 η_t is assumed to be a Gaussian mixture. We assume that η_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, since $E(\eta_t) = 0$ and $V(\eta_t) = 1$, the distribution of η_t is completely defined by $\gamma = [\mu_1, \sigma_1, p]'$.

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

The results are reported in Table 2. For 9 countries out of 17, the absolute value of the estimate of θ is larger than one. For these countries, the estimated MA process is non-fundamental.

For an ARMA(1,1) model, the first values of the IRF are: c, $(\phi - \theta)c$, $(\phi - \theta)\phi c$, ... For instance, for the USA, the first three values of the IRF are 0.5, 4.6 and 0.2. Hence, the nonfundamentalness of the estimated US GDP growth process implies the existence of a bump effect one period after the shock. It is interesting to compare this IRF to the one associated with the fundamental ARMA(1,1) process featuring the same second-order properties as the previous one. The

[28] The data are available at http://www.ggdc.net/maddison/maddison-project/home.htm. For the 17 considered countries (see the list of countries in Table 2), the Phillips-Perron test rejects the null hypothesis of the presence of a unit root in the per capita real GDP growth at the 1% significance level.

dynamics of the latter process reads:

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

In the latter case, the first three terms of the IRF are: 4.6, 0.7, 0.03. In particular, for the latter process, the IRF does not feature a bump one period after the shock. This illustrates the importance of the estimation of the fundamentalness regime for the IRF analysis.

5.3 Bivariate real-data example: GNP growth and unemployment

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

Using the same dataset and 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 (the non-fundamental ones and the fundamental one). Nevertheless, as explained in Section 3.1, if the underlying structural shocks are non-Gaussian and independent, then the data-generating SVARMA process, be it fundamental or not, is identifiable. Preliminary indications regarding the non-normality of the system can be obtained by applying normality tests on the endogenous variable themselves. In this regard, the Lütkepohl and Theilen (1991)'s multivariate tests, based on standardized third and fourth moments of possibly correlated variables, point to the non-Gaussianity of the vectors of endogenous variables.[29]

We employ both the ML and the 2SLS-GMM approaches described in Subsections 4.1 and 4.2 to fit VARMA(p,1) models to BQ's dataset. As in Blanchard and Quah (1989) and Lippi and Reichlin (1993, 1994), the two endogenous variables follow a stationary process and are not cointegrated.[30] In the context of the ML approach, the distributions of the independent shocks

[29] The p-value of this test, whose null hypothesis is that the considered variables are Gaussian, is of 3.5%.

[30] As noted by Lippi and Reichlin (1994), Footnote 1 p. 310, assuming that the dynamics of the two endogenous

$\eta_{j,t}$, for $j \in \{1,2\}$, are assumed to be Gaussian mixtures. Specifically, we assume that $\eta_{j,t}$ is drawn from $\mathcal{N}(\mu_{j,1}, \sigma_{j,1}^2)$ with a probability p_j and from $\mathcal{N}(\mu_{j,2}, \sigma_{j,2}^2)$ with a probability $1-p_j$. Therefore, if we have $E(\eta_{j,t}) = 0$ and $V(\eta_{j,t}) = 1$, the distribution of the two-dimensional vector η_t is completely defined by $\gamma = [\mu_{1,1}, \mu_{2,1}, \sigma_{1,1}, \sigma_{2,1}, p_1, p_2]'$. As regards the 2SLS-GMM approach, in the second step of this approach, we consider the order-2, order-3 and order-4 moments given in (4.18), with:

$$\begin{aligned}(u_1,u_2,v_1,v_2) \quad \in \quad & \{(2,0,0,0),(0,2,0,0),(1,0,2,0),(2,0,1,0),(1,0,0,2),(2,0,0,1),\\ & (0,1,0,2),(0,2,0,1),(0,1,2,0),(0,2,1,0)\},\end{aligned}$$

which results in 30 restrictions.[31]

We consider different values of the order p of $\Phi(L)$. Specifically, for $p \in \{1,\ldots,6\}$, we proceed as follows. We first apply the 2SLS-GMM approach, benefitting from the fact that the numerical complexity of this approach does not depend on order p of $\Phi(L)$. We then use the 2SLS-GMM estimates as starting values for the numerical optimization of the truncated log-likelihood function.[32]

An advantage of the ML approach is to make it possible to use standard lag selection criteria. Our preferred model is the one minimizing the BIC criteria, leading to $p = 4$. Parameter estimates resulting from both the 2SLS-GMM and the ML approach are reported in Table 3. For both approaches, the roots of $\det(\Phi(z))$ are outside the unit circle; the estimated processes are therefore stationary. Besides, for both estimation methods, the two roots of $\det(I - \Theta z)$ are on each side of the unit circle. Hence, the estimated processes are nonfundamental. Figure 3 shows the ML estimated parametric distributions of the structural shocks (Gaussian mixtures, represented by dotted lines). These estimated distributions are fairly close to kernel-based densities associated with the $\eta_{j,t}$'s estimates derived from Proposition 3 (black solid lines).

It is difficult to check ex ante the reasonableness of the assumptions on the structural shocks, in particular regarding the independence assumption A.1. Tests can however be applied to the estimated structural shocks. In the present case, the application of Box and Pierce (1970) and Ljung and Box (1978) tests on the estimated series of structural shocks do not detect auto-correlation in the estimated series of $\eta_{j,t}$, $|\eta_{j,t}|$, nor in the $\eta_{j,t}^2$, which is a necessary condition for independence. The normality hypothesis of the $\eta_{j,t}$'s is rejected by the Shapiro and Wilk (1965)'s test, with p-

variables is described by a stationary VAR model (as done in Blanchard and Quah, 1989; Lippi and Reichlin, 1993, 1994) amounts to precluding cointegration.

[31] See Footnote 26 for the reason why the u_i's and v_i's are not only in $\{0,1\}$. In the 2SLS step, we use $k = p+3$ (i.e. we regress Y_{t-1} on $Y_{t-2},\ldots,Y_{t-7}$), see Appendix C.1.

[32] As regards starting values for the elements of γ, we do the following: after the 2SLS-GMM estimation, we apply the algorithm of Proposition 3 to obtain estimates of the structural shocks $\eta_{j,t}$. Then, for $j \in \{1,2\}$, we estimate parameters $(\mu_{1,i}, \sigma_{1,i}, p_i)$ by MLE, which is very fast.

values below 2%.[33] A limitation of these residual tests is however that they do not take into account the uncertainty regarding the estimation of the $\eta_{j,t}$'s; this is left for future research.

Figure 4 displays the impulse response functions resulting from the ML and 2SLS-GMM approaches and compares them with those obtained with long-run restrictions *à la* Blanchard and Quah (1989). Because the ML and 2SLS-GMM approaches do not rely on restrictions based on economic theory, the resulting structural shocks have no a priori economic interpretation. However, for comparing the different approaches, we bring the BQ supply shock closer to the ML/2SLS-GMM-estimated shocks that accounts for the largest part of the GNP long-run variance. We call this shock the "long-run shock".

Figure 4 shows that, overall, the different response functions show similarities across the three estimation approaches. In the ML and 2SLS-GMM approaches, both structural shocks –Shock 1 (first column of charts) and Shock 2 (second column of charts)– have a long-run impact on GNP. This is not the case with the BQ approach where, by construction, the effect of Shock 1 on GNP converges to zero. The fact that none of the two structural shocks identified within a bivariate VAR has a zero impact on output in the long-run was also obtained by, e.g. Cochrane (1994).[34] Interestingly, both the ML and 2SLS-GMM approaches result in unemployment IRFs that feature the same shapes, but they do not agree with the amplitudes of the responses (lower row of plots). In particular, the bottom-right plot shows that, for both the ML and 2SLS-GMM approaches, the instantaneous response of unemployment to the second long-run shock is close to zero, reflecting the fact that the estimated coefficients $C_{2,2}$ are small (and not statistically different from zero, see Table 3).[35] Besides, while the two approaches (ML and 2SLS-GMM) lead to close responses of GNP to the long-run shock (upper right plot), they suggest different responses of GNP to the other shock (upper left plot).

6 Concluding remarks

We have shown in this paper that the static and dynamic identification problems associated with Gaussian SVARMA processes disappear in the non-Gaussian case. Whenever the shocks are not

[33] Because no auto-correlation is detected in the residual, there is no need for using autocorrelation-robust normality tests such as those of Bai and Ng (2005) and of Lobato and Velasco (2004). By contrast, the latter two tests had to be used on the autocorrelated output growth series used in Subsection 5.2.

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

[35] Note that these results could not have been obtained by implementing an identification strategy *à la* Sims (imposing that one structural shock has no contemporaneous impact on the unemployment rate). Indeed, the standard short-run restriction methodology is based on VAR models and therefore necessarily leads to fundamental models. By contrast, our approach remains a priori agnostic with respect to the fundamentalness regime. In this regard, our estimated SVARMA model is non-fundamental.

Gaussian, the SVARMA becomes identified up to a permutation and sign change of the structural shocks. This paper further proposes parametric and semi-parametric estimation methods able to consistently estimate possibly non-fundamental representation in the moving average dynamics.

Our analysis highlights that a dynamic model constructed to derive impulse response functions requires much more structural assumptions on the error terms, namely serial independence, than a pure forecast model for which uncorrelated errors may be sufficient. In this respect the conventional econometric toolboxes available for macroeconomists have often been conceived for forecasting purposes and are not appropriate for the analysis of policy shocks. Moreover, the instantaneous independence of the structural shocks is also required.

Because it focuses on second-order properties, the standard SVARMA literature often introduces potentially misleading identification assumptions that may entail misspecification and naive interpretations of VARMA residuals. As shown in the parametric and semi-parametric analysis developed in Section 4 and in the applications presented in Section 5, such potential pitfalls can be addressed in the non-Gaussian case, provided that the independence assumptions are valid and the appropriate estimation methods are used.

The methods developed in this paper might be extended in several directions, which are left for future research. First testing procedures, in particular tests of fundamentalness, may be obtained. Second the properties of estimation methods can be analyzed in a neighbourhood of the Gaussian assumption (see e.g. Gouriéroux and Jasiak, 2016), or in a neighbourhood of unit roots (see e.g. Chen, Davis, and Song, 2011, for mixed causal/noncausal MA process). Third, the asymptotic properties of our ML estimate in the presence of cointegration may be derived. Fourth, the identification and estimation results might be extended to the case of more errors than observables. Indeed, while identification results exist when the errors are not Gaussian (see e.g. Th 3.1. in Eriksson and Koivunen, 2004, for the static case, Gouriéroux and Zakoian, 2015, for stable multivariate processes, or Gagliardini and Gouriéroux, 2018, for a non-Gaussian factor model), the possibility to identify the dynamics when the number of shocks is larger than the number of endogenous variables and the errors are not Gaussian would be important in the discussion of the effect of omitted variables (see e.g. Giannone and Reichlin, 2006; Lütkepohl, 2014).

A Identifiability, Reversibility, Estimation and Responses 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 based on (cross) moments up to order 3. Finally we discuss the bias on the IRF when a misspecified fundamental representation is used.

A.1 Limit optimization problem under the truncated ML method

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

$$\begin{aligned} L_T(y_1^T;\theta,\gamma) &= \mathbb{1}_{|\theta|<1}\sum_{t=1}^{T}\log\left\{g\left(\sum_{h=0}^{t-1}\theta^h y_{t-h};\gamma\right)\right\} \\ &+ \mathbb{1}_{|\theta|>1}\sum_{t=1}^{T}\log\left\{\frac{1}{|\theta|}g\left(-\sum_{h=0}^{T-t-1}\frac{1}{\theta^{h+1}}y_{t+h+1};\gamma\right)\right\}. \end{aligned}$$

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 denotes 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 $E_\theta(y_{t+1}|y_{t+2}^{\infty}) = -\sum_{h=1}^{\infty}\frac{1}{\theta^{h+1}}y_{t+h+1}$ (these expectations do not depend on γ).

A.2 Lack of identification in the Gaussian case

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

$$L_\infty(\theta,\sigma^2) = -\frac{1}{2}\left[\mathbb{1}_{|\theta|<1}\tilde{L}_1(\theta,\sigma^2) + \mathbb{1}_{|\theta|>1}\tilde{L}_2(\theta,\sigma^2)\right],$$

with

$$\begin{aligned}
\tilde{L}_1(\theta,\sigma^2) &= \log\sigma^2 + \frac{1}{\sigma^2}E_0\left[\left(y_t + \sum_{h=1}^{\infty}\theta^h y_{t-h}\right)^2\right], \\
\tilde{L}_2(\theta,\sigma^2) &= \log\theta^2\sigma^2 + \frac{1}{\theta^2\sigma^2}E_0\left[\left(y_{t+1} + \sum_{h=1}^{\infty}\frac{1}{\theta^h}y_{t+h+1}\right)^2\right].
\end{aligned}$$

We have to minimize $\mathbb{1}_{|\theta|<1}\tilde{L}_1(\theta,\sigma^2) + \mathbb{1}_{|\theta|>1}\tilde{L}_2(\theta,\sigma^2)$ w.r.t. (θ,σ^2). Let us denote by θ_0 and σ_0^2 the true values of θ and σ^2.

If $|\theta_0|<1$, we know that $E_0(y_t|y_{-\infty}^{t-1}) = \sum_{h=1}^{\infty}\theta_0^h y_{t-h}$ and, due to the time-reversibility of a Gaussian process, we have $E_0(y_{t+1}|y_{t+2}^{\infty}) = \sum_{h=1}^{\infty}\theta_0^h y_{t+h+1}$. Moreover $E_0[\{y_t - E_0(y_t|y_{-\infty}^{t-1})\}^2] = E_0[\{y_{t+1} - E_0(y_{t+1}|y_{t+2}^{\infty})\}^2] = \sigma_0^2$.

Therefore, $\tilde{L}_1(\theta,\sigma^2)$ is minimized on $|\theta|<1$ at $\tilde{\theta}_0 = \theta_0$ and $\tilde{\sigma}_0^2 = \sigma_0^2$, and the minimum is $\log\sigma_0^2 + 1$. Similarly, $\tilde{L}_1(\theta,\sigma^2)$ is minimized on $|\theta|>1$ at $\tilde{\theta}_0 = 1/\theta_0$ and $\tilde{\sigma}_0^2 = \theta_0^2\sigma_0^2$ and the minimum is, again, $\log\sigma_0^2 + 1$.

This means that, if $|\theta_0|<1$, the asymptotic log-likelihood reaches a maximum of $\log\sigma_0^2+1$ for two different values: (θ_0,σ_0^2) and $(1/\theta_0,\theta_0^2\sigma_0^2)$, and the model is not asymptotically identifiable.

Let us now consider the case $|\theta_0|>1$. We have:

$$\left(1 - \frac{1}{\theta_0}L^{-1}\right)^{-1} y_{t+1} = -\theta_0\varepsilon_t,$$

therefore

$$y_{t+1} = -\theta_0\varepsilon_t - \sum_{h=1}^{\infty}\frac{1}{\theta_0^h}y_{t+h+1},$$

and, as a result $E_0(y_{t+1}|y_{t+2}^{\infty}) = -\sum_{h=1}^{\infty}\frac{1}{\theta_0^h}y_{t+h+1}$ and $E_0(\{y_{t+1} - E_0(y_{t+1}|y_{t+2}^{\infty})\}^2) = \theta_0^2\sigma_0^2$. Due to the time-reversibility of a Gaussian process, we also have $E_0(y_t|y_{-\infty}^{t-1}) = -\sum_{h=1}^{\infty}\frac{1}{\theta_0^h}y_{t-h}$ and $E_0(\{y_{t+1} - E_0(y_{t+1}|y_{-\infty}^{t-1})\}^2) = \theta_0^2\sigma_0^2$.

Therefore, $\tilde{L}_1(\theta,\sigma^2)$ is minimized on $|\theta|<1$ at $\tilde{\theta}_0 = 1/\theta_0$ and $\tilde{\sigma}_0^2 = \theta_0^2\sigma_0^2$, and the minimum is $\log(\theta_0^2\sigma_0^2)+1$. Similarly, $\tilde{L}_1(\theta,\sigma^2)$ is minimized on $|\theta|>1$ at $\tilde{\theta}_0 = \theta_0$ and $\tilde{\sigma}_0^2 = \sigma_0^2$ and the minimum is, also, $\log(\theta_0^2\sigma_0^2)+1$.

Again, the asymptotic log-likelihood is maximized at two different points and the model is not asymptotically identified.

A.3 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:

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

Let us for instance assume that ε_t follows a stable distribution, with stability parameter α, we get:

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

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

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

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

A.4 Moment method

If we do not want to make a parametric assumption about the distribution of ε_t, we can use a moment method based on higher-order cross moments (see Section 4.2 and Appendix C).

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

A.5 Impulse Response Functions

The MA(1) case allows to easily illustrates the fact that fundamental and non-fundamental SVARMA models that feature the same second-order properties entail different IRFs. Consider for instance the MA(1) processes (y_t) and (y_t^*) defined by

$$y_t = \sigma\eta_t - \theta\sigma\eta_{t-1} \quad \text{and} \quad y_t^* = \sigma\theta\eta_t - \sigma\eta_{t-1}.$$

Although (y_t) and (y_t^*) have the same second-order properties, they react differently to shocks to η_t.

Consider a one-unit increase in η_t. Whereas this shock results in increases in y_t, y_{t+1} and y_{t+h}, $h>1$, by σ, $-\theta\sigma$ and 0, respectively, it implies increases in y_t^*, y_{t+1}^* and y_{t+h}^* by $\sigma\theta$, $-\sigma$ and 0. In particular, if one of these two IRFs is decreasing (in absolute values), the other is hump-shaped.

B Proofs

B.1 Proof of Proposition 1

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

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

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

where the processes ε_t and ε_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 π is a permutation and $A_{k,j}$ (respectively $A_{k,j}^*$) is the j^{th} column of A_k (respectively A_k^*) provided the following condition holds:

> *The components of ε_t (resp. ε_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 ε_t (resp. ε_t^*) is replaced by η_t (resp. η_t^*), where the components of η_t (resp. η_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 moving average matrix coefficients 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 Ψ_k are therefore identified and C is identified up to a permutation and a sign change of its columns. Since $\Phi(L)$ and $\Psi(L)$ are identified, $\Theta(L) = \Phi(L)\Psi(L)$ is also identified.

B.2 Proof of Propositions 3 and 4

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

$$\underbrace{\left[\begin{array}{c} \varepsilon_0^{*(1)} \\ \vdots \\ \varepsilon_{T-1}^{*(1)} \\ \hline \varepsilon_1^{*(2)} \\ \vdots \\ \varepsilon_T^{*(2)} \end{array}\right]}_{=\varepsilon^*} = \underbrace{\left[\begin{array}{c|c} J_1 & J_{12} \\ \hline 0 & J_2 \end{array}\right]}_{=J} \underbrace{\left[\begin{array}{c} W_1{}^{(1)} \\ \vdots \\ W_T{}^{(1)} \\ \hline W_1{}^{(2)} \\ \vdots \\ W_T{}^{(2)} \end{array}\right]}_{=W} + \underbrace{\left[\begin{array}{c|c} \left\{(U^{(1)})^{-1}\right\}^T & M_1 \\ \vdots & \vdots \\ (U^{(1)})^{-1} & M_T \\ \hline 0 & U^{(2)} \\ \vdots & \vdots \\ 0 & (U^{(2)})^T \end{array}\right]}_{=M} \left[\begin{array}{c} \varepsilon_T^{*(1)} \\ \varepsilon_0^{*(2)} \end{array}\right], \tag{a.3}$$

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

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

Because the eigenvalues of $U^{(2)}$ and of $(U^{(1)})^{-1}$ are strictly inside the unit circle, the elements of M corresponding to periods t that are far from sample boundaries (0 and T, assuming that T is large) are extremely small. As a result, for such periods t, the ε_t^*'s are well approximated by the corresponding components of JW. (These components correspond to the e_t^* defined in Proposition 3).

The likelihood associated with $\varepsilon^* = [e^{*(1)\prime}_0, \ldots, e^{*(1)\prime}_{T-1}, e^{*(2)\prime}_1, \ldots, e^{*(2)\prime}_T]$ is:

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

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

In the limit (for $T \to \infty$), $1/T \log g^*(\varepsilon^*, \Gamma)$ does not depend on a potential conditioning initial values $Y_0, Y_{-1}, \ldots, Y_{-p+1}$. Accordingly, in the following, we proceed under the assumption that $Y_0 = Y_{-1} = \cdots = Y_{-p+1} = 0$.

Let's consider the vector $V = [Z_1', \ldots, Z_T']'$ where $Z_t = \Phi(L)Y_t$. If $Y = [Y_1', \ldots, Y_T']'$, we have $V = \Lambda Y$ where Λ is a lower triangular matrix whose diagonal is filled with ones (and therefore $\det \Lambda = 1$).

By definition of W_t, we have $AW_t = \Phi(L)Y_t$ and, therefore, $A'Z_t = W_t$. Hence, with $\tilde{W} = [W_1', W_2', \ldots, W_T']'$, we have:

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

Let's denote by P the permutation that is such that $W = P\tilde{W}$, where W is defined in equation (a.3). An approximation of ε^* is given by $\mathscr{E}(Y) := JP(I \otimes A')\Lambda Y$ (this is Proposition 3).

Because A and P are orthogonal matrices, we have $|\det A| = |\det P| = 1$. We also have $\det \Lambda = 1$. This implies:

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

Therefore, when multiplied by $1/T$, the likelihood associated with Y can be approximated by:

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

or, neglecting $\frac{1}{T}\log g_{\varepsilon^{*(1)}}({e^*}_0^{(1)},\Gamma)$ and $\frac{1}{T}\log g_{\varepsilon^{*(2)}}({e^*}_T^{(2)},\Gamma)$), by:

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

where

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

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

C 2SLS-GMM: Estimation and asymptotic properties

C.1 Estimation of $\mu,\Phi_1,\ldots,\Phi_p$ by 2SLS

Consider the model:

$$Y_t = \mu + \Phi_1 Y_{t-1} + \cdots + \Phi_1 Y_{t-p} + Z_t,$$

where $Z_t = C_0\eta_t + \cdots + C_q\eta_{t-q}$ and where the components of η_t, that are the $\eta_{i,t}$'s, are zero-mean, unit-variance, serially and mutually independent shocks. Assuming that (Y_t) is covariance-stationary, we denote by m the unconditional expectation of Y_t, i.e. $m = E(Y_t) = (I - \Phi_1 - \cdots - \Phi_p)^{-1}\mu$, and by Γ_i its auto-covariance matrix of order i, i.e. $\Gamma_i = E(Y_tY_{t-i}') - E(Y_t)E(Y_{t-i})'$.

Consistent estimates of $\Pi = [\mu,\Phi_1,\ldots,\Phi_p]$ can be obtained by applying two-stage least squares (2SLS). For $j > 0$, Y_{t-q-j} is independent from Z_t. As a result, the Y_{t-q-j}'s, $j > 0$ can be used as instruments to estimate Π.

Take $k \geq p$ and let us introduce the notations $W_t = [1, Y_{t-q-1}', \ldots, Y_{t-q-k}']'$, $X_t = [1, Y_{t-1}', \ldots, Y_{t-p}']'$, $\mathbf{W} = [W_1,\ldots,W_T]'$, $\mathbf{X} = [X_1,\ldots,X_T]'$, $\mathbf{Y} = [Y_1,\ldots,Y_T]'$ and $\mathbf{Z} = [Z_1,\ldots,Z_T]'$. The first step of the

2SLS approach provides the following fitted value of $\mathbf{X}$:

$$\hat{\mathbf{X}} = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}.$$

We then regress the Y_t's on the $\hat{X}_t$'s by OLS and get the 2SLS estimate of Π:

$$\begin{aligned}
\hat{\Pi}' &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{Y} = [\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{Y} \\
&= \Pi' + \left[\mathbf{X}'\mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{Z} \\
&= \Pi' + \frac{1}{\sqrt{T}}\left[\frac{\mathbf{X}'\mathbf{W}}{T}\left(\frac{\mathbf{W}'\mathbf{W}}{T}\right)^{-1}\frac{\mathbf{W}'\mathbf{X}}{T}\right]^{-1}\frac{\mathbf{X}'\mathbf{W}}{T}\left(\frac{\mathbf{W}'\mathbf{W}}{T}\right)^{-1}\left(\sqrt{T}\frac{\mathbf{W}'\mathbf{Z}}{T}\right). \qquad \text{(a.4)}
\end{aligned}$$

We have $\frac{\mathbf{X}'\mathbf{W}}{T} \xrightarrow{p} Q_{XW}$ and $\frac{\mathbf{W}'\mathbf{W}}{T} \xrightarrow{p} Q_W$, where Q_{XW} and Q_W are given by:

$$Q_{XW} = \begin{bmatrix} 1 & m' & m' & \cdots & m' \\ m & \widetilde{\Gamma}_q & \widetilde{\Gamma}_{q+1} & \cdots & \widetilde{\Gamma}_{q+k-1} \\ m & \widetilde{\Gamma}_{q-1} & \widetilde{\Gamma}_q & \cdots & \widetilde{\Gamma}_{q+k-2} \\ \vdots & & & \vdots & \\ m & \widetilde{\Gamma}_{q-p+1} & & \cdots & \widetilde{\Gamma}_{q+k-p} \end{bmatrix} \text{ and } Q_W = \begin{bmatrix} 1 & m' & m' & \cdots & m' \\ m & \widetilde{\Gamma}_0 & \widetilde{\Gamma}_1 & \cdots & \widetilde{\Gamma}_{k-1} \\ m & \widetilde{\Gamma}_{-1} & \widetilde{\Gamma}_0 & \cdots & \widetilde{\Gamma}_{k-2} \\ \vdots & & & \vdots & \\ m & \widetilde{\Gamma}_{-k+1} & \widetilde{\Gamma}_{-k+2} & \cdots & \widetilde{\Gamma}_0 \end{bmatrix},$$

with $\widetilde{\Gamma}_i = E(Y_t Y_{t-i}') = \Gamma_i + mm'$.

Using the notations $\hat{\alpha}_T = vec(\hat{\Pi})$ and $\alpha = vec(\Pi)$, the expansion (a.4) leads to:

$$\sqrt{T}(\hat{\alpha}_T - \alpha) \approx Q_\pi \left(\sqrt{T} vec\left(\frac{\mathbf{Z}'\mathbf{W}}{T}\right)\right), \qquad \text{(a.5)}$$

where

$$Q_\pi = \left\{[Q_{XW}Q_W^{-1}Q_{XW}']^{-1}Q_{XW}Q_W^{-1}\right\} \otimes I_{n\times n}.$$

C.2 Cumulant-based identification

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

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

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

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

Let us denote by Ψ_ξ the log-Laplace transform of a given random variable ξ defined by:

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

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

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

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

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

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

Using (a.7) with $\xi = u'Z_t + v'Z_{t-1}$ and (a.8), and taking the expansions of both sides of (a.6) up to order 4, we get the following restrictions, holding for any pair (u, v):

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

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

C.3 Asymptotic accuracy of the 2SLS-GMM approach

In the first step of the 2SLS-GMM approach, $\alpha = vec(\Pi) = vec([\mu, \Phi_1, \ldots, \Phi_p])$ is estimated by implementing the 2SLS approach described in Appendix C.1. The remaining parameters, gathered in vector β, are then estimated by employing the GMM approach presented in Subsection 4.2.2. In the context of this second step, the moment restrictions write:

$$E_0\left[h(\underline{Y_t}; \alpha_0, \beta_0)\right] = 0, \tag{a.9}$$

where $\underline{Y_t} = (Y_t, Y_{t-1}, Y_{t-2})$ is a stationary process. These moment restrictions overidentify β for a given α if the size of h is larger than that of β.

For a given first-step 2SLS estimator $\hat{\alpha}_T$ of α, the estimator of β is defined as the solution of the minimization problem:

$$\hat{\beta}_T = \underset{\beta}{\operatorname{argmin}} \left(\frac{1}{T}\sum_{t=1}^{T} h(\underline{y_t}; \hat{\alpha}_T, \beta)\right)' \Omega \left(\frac{1}{T}\sum_{t=1}^{T} h(\underline{y_t}; \hat{\alpha}_T, \beta)\right), \tag{a.10}$$

where Ω is a given $(\dim h \times \dim h)$ positive symmetric matrix.

Under standard regularity conditions, this 2-step estimator is consistent and asymptotically normal. Its asymptotic covariance matrix has to account for the first-step estimation error. The derivation of this asymptotic covariance matrix is based on the first-order conditions (FOCs) associated with the second step of the approach. These FOCs are:

$$\left(\frac{1}{T}\sum_{t=1}^{T} h(\underline{y_t}; \hat{\alpha}_T, \hat{\beta}_T)\right)' \Omega \left(\frac{1}{T}\sum_{t=1}^{T} \frac{\partial h}{\partial \beta'}(\underline{y_t}; \hat{\alpha}_T, \hat{\beta}_T)\right) = 0.$$

The expansion of the FOCs leads to:

$$\begin{aligned}\sqrt{T}(\hat{\beta}_T - \beta_0) \;\approx\; & -\left[E_0\left(\frac{\partial h'}{\partial \beta}\right)\Omega E_0\left(\frac{\partial h}{\partial \beta'}\right)\right]^{-1} \\ & E_0\left(\frac{\partial h'}{\partial \beta}\right)\Omega\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t + E_0\left(\frac{\partial h}{\partial \alpha'}\right)\sqrt{T}(\hat{\alpha}_T - \alpha_0)\right\},\end{aligned} \tag{a.11}$$

where

$$h_t = h(\underline{y_t}; \alpha_0, \beta_0), \quad E_0\left(\frac{\partial h'}{\partial \beta}\right) = E_0\left(\frac{\partial h'}{\partial \beta}(\underline{Y_t}; \alpha_0, \beta_0)\right), \quad E_0\left(\frac{\partial h}{\partial \alpha'}\right) = E_0\left(\frac{\partial h}{\partial \alpha}(\underline{Y_t}; \alpha_0, \beta_0)\right).$$

We deduce from (a.5) and (a.11) that:

$$\sqrt{T}\begin{bmatrix}\hat{\alpha}_T-\alpha_0\\ \hat{\beta}_T-\beta_0\end{bmatrix}\approx-\begin{bmatrix}Q_\pi & 0\\ I_{\beta\beta}^{-1}I_{\beta\alpha}Q_\pi & I_{\beta\beta}^{-1}E_0\left(\frac{\partial h'}{\partial\beta}\right)\Omega\end{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix}g_t\\ h_t\end{bmatrix}, \tag{a.12}$$

where $g_t = vec(Z_t W_t')$ (see Subsection C.1 for the notations Z_t and W_t) and where

$$I_{\beta\beta}=E_0\left(\frac{\partial h'}{\partial\beta}\right)\Omega E_0\left(\frac{\partial h}{\partial\beta'}\right),\quad I_{\beta\alpha}=E_0\left(\frac{\partial h'}{\partial\beta}\right)\Omega E_0\left(\frac{\partial h}{\partial\alpha'}\right).$$

The expression of the variance-covariance matrix of the 2-step estimator follows by noting that, under the strict stationarity of Z_t, process $(g_t', h_t')'$ is also stationary, and by applying the central limit theorem:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix}g_t\\ h_t\end{bmatrix}\xrightarrow{d}\mathcal{N}(0,Q),$$

where

$$Q=\sum_{k=-\infty}^{+\infty}Cov_0\left[\begin{pmatrix}g_t\\ h_t\end{pmatrix},\begin{pmatrix}g_{t-k}\\ h_{t-k}\end{pmatrix}\right]=:\begin{bmatrix}Q_{gg} & Q_{gh}\\ Q_{hg} & Q_{hh}\end{bmatrix}.$$

With these notations, we have:

$$\sqrt{T}(\hat{\beta}_T-\beta_0)\xrightarrow{d}\mathcal{N}(0,\Sigma(\Omega)),$$

with

$$\begin{aligned}\Sigma(\Omega) &= I_{\beta\beta}^{-1}E_0\left(\frac{\partial h'}{\partial\beta}\right)\begin{bmatrix}\Omega E_0\left(\frac{\partial h}{\partial\alpha'}\right)Q_\pi & \Omega\end{bmatrix}Q\begin{bmatrix}\Omega E_0\left(\frac{\partial h}{\partial\alpha'}\right)Q_\pi & \Omega\end{bmatrix}'E_0\left(\frac{\partial h}{\partial\beta'}\right)I_{\beta\beta}^{-1}\\ &= \left[E_0\left(\frac{\partial h'}{\partial\beta}\right)\Omega E_0\left(\frac{\partial h}{\partial\beta'}\right)\right]^{-1}E_0\left(\frac{\partial h'}{\partial\beta}\right)\Omega Q_0\Omega' E_0\left(\frac{\partial h}{\partial\beta'}\right)\left[E_0\left(\frac{\partial h'}{\partial\beta}\right)\Omega E_0\left(\frac{\partial h}{\partial\beta'}\right)\right]^{-1},\end{aligned}$$

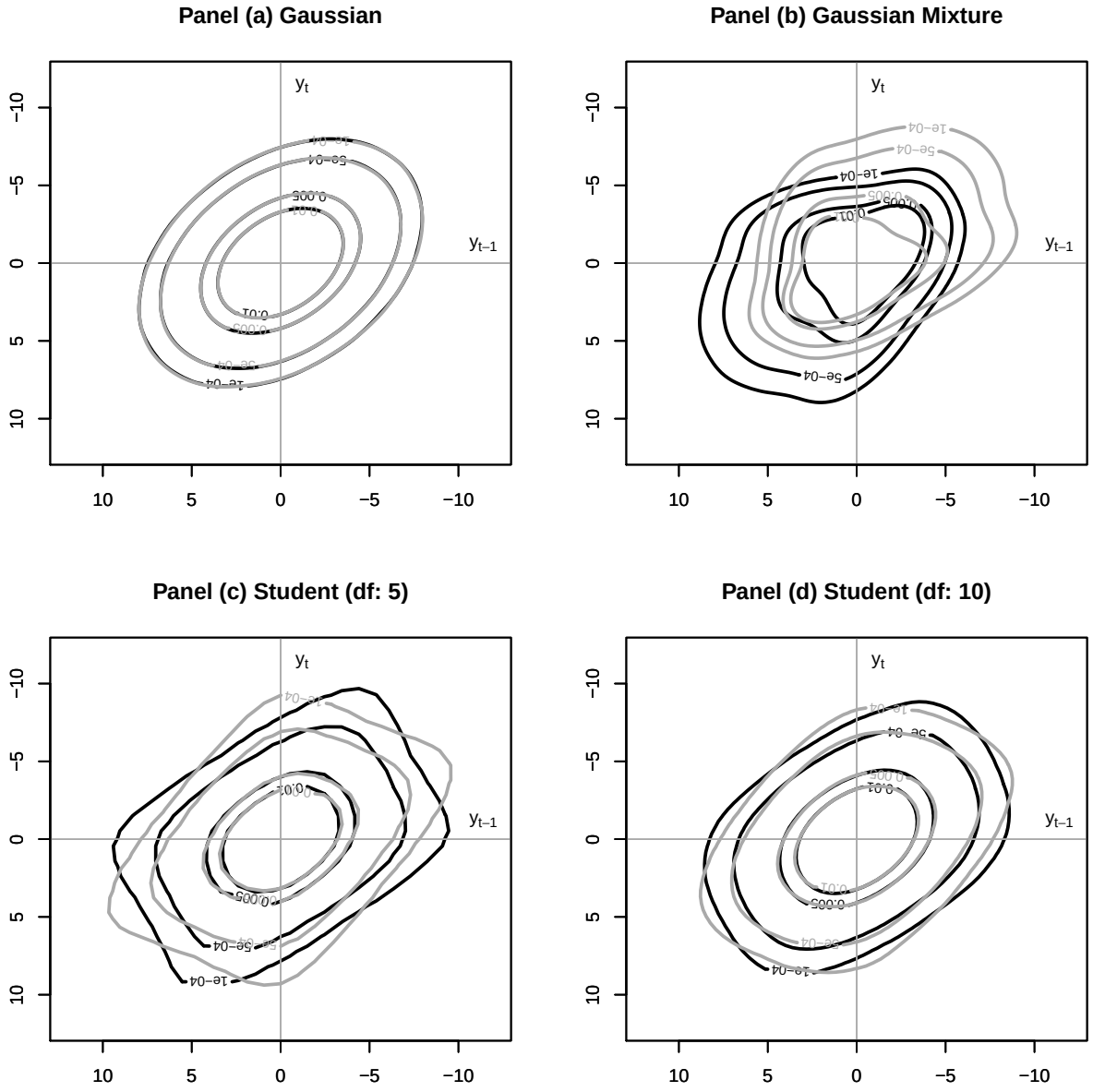
with

$$Q_0=\begin{bmatrix}E_0\left(\frac{\partial h}{\partial\alpha'}\right)Q_\pi & Id\end{bmatrix}Q\begin{bmatrix}E_0\left(\frac{\partial h}{\partial\alpha'}\right)Q_\pi & Id\end{bmatrix}'.$$

The optimal value for Ω is $\Omega^*=Q_0^{-1}$. In this case, we have

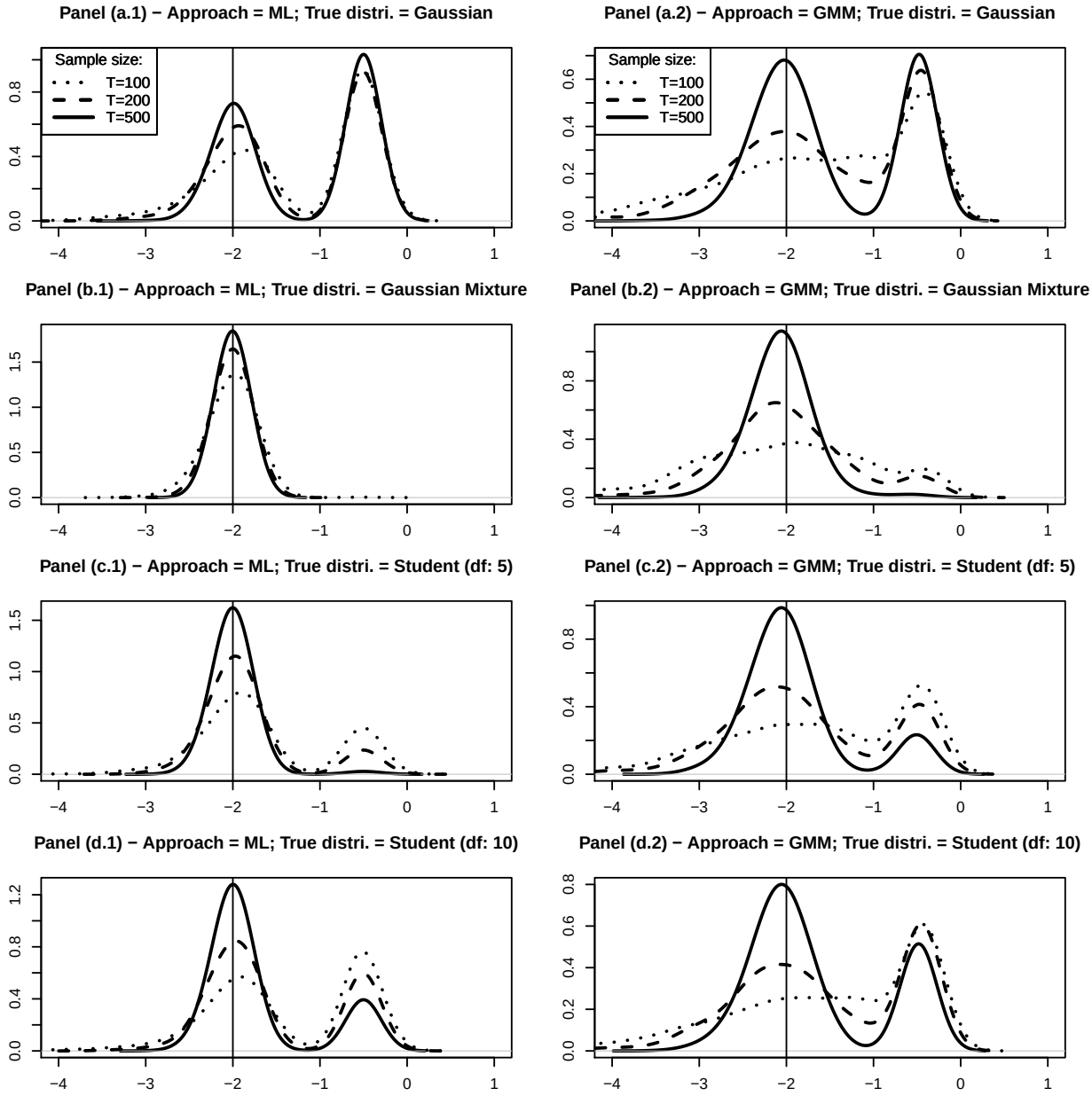
$$\Sigma(\Omega^*)=\left[E_0\left(\frac{\partial h'}{\partial\beta}\right)Q_0^{-1}E_0\left(\frac{\partial h}{\partial\beta'}\right)\right]^{-1}.$$

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



Note: 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 ε_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 ε_ts. For Panel b (mixture of Gaussian distributions), ε_t is drawn from the Gaussian distribution $\mathcal{N}(0,\sigma_1^2)$ with probability p and from $\mathcal{N}(0,\sigma_2^2)$ with probability $1-p$; specifically, we set: $\mu_1 = -0.7$, $\mu_2 = 0.7$, $\sigma_1 = 0.32$, $\sigma_2 = 0.95$, $p = 0.5$, which results in a zero-mean unit-variance distribution with order-3 and order-4 cumulants of 0.85 and 0, respectively.

Figure 2: Monte-Carlo experiments: distribution of θ estimators



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

Table 1: Results of the Monte-Carlo experiment

ε_t's distribution:	Bias	RMSE	MAE	S.D.	$\overline{\sigma_{asy}}$	α =75%	α =90%	α =95%
Panel (a) Maximum Likelihood approach								
Sample size: T=100								
Gaussian	0.74	1.17	0.97	0.91	0.17	0.25	0.31	0.33
Mixture of Gaussian	−0.01	0.24	0.17	0.24	0.19	0.67	0.84	0.88
Student (df: 5)	0.35	0.88	0.61	0.80	0.25	0.51	0.60	0.64
Student (df: 10)	0.59	1.14	0.85	0.98	0.22	0.34	0.42	0.45
Sample size: T=200								
Gaussian	0.72	1.09	0.86	0.81	0.13	0.30	0.37	0.39
Mixture of Gaussian	0.00	0.14	0.11	0.14	0.13	0.74	0.88	0.92
Student (df: 5)	0.17	0.58	0.35	0.55	0.20	0.63	0.77	0.80
Student (df: 10)	0.46	0.88	0.62	0.74	0.17	0.46	0.56	0.58
Sample size: T=500								
Gaussian	0.79	1.09	0.85	0.76	0.08	0.32	0.39	0.41
Mixture of Gaussian	0.00	0.08	0.06	0.08	0.08	0.76	0.89	0.94
Student (df: 5)	0.01	0.22	0.12	0.22	0.14	0.76	0.89	0.93
Student (df: 10)	0.29	0.68	0.39	0.62	0.12	0.62	0.72	0.75
Panel (b) Generalized Method of Moments approach								
Sample size: T=100								
Gaussian	−1.81	50.78	3.24	50.77	189.53	0.34	0.49	0.53
Mixture of Gaussian	−0.54	13.02	1.33	13.01	4.82	0.46	0.65	0.74
Student (df: 5)	0.25	1.36	1.05	1.33	0.52	0.38	0.52	0.56
Student (df: 10)	−1.86	49.58	3.35	49.57	184.67	0.33	0.48	0.52
Sample size: T=200								
Gaussian	0.47	1.10	0.91	1.00	0.26	0.29	0.40	0.46
Mixture of Gaussian	0.00	0.77	0.56	0.77	0.38	0.52	0.70	0.77
Student (df: 5)	0.21	0.97	0.74	0.94	0.32	0.38	0.54	0.61
Student (df: 10)	0.44	1.08	0.87	0.99	0.27	0.33	0.44	0.50
Sample size: T=500								
Gaussian	0.53	0.96	0.73	0.81	0.15	0.34	0.45	0.51
Mixture of Gaussian	−0.05	0.35	0.25	0.35	0.25	0.67	0.82	0.88
Student (df: 5)	0.10	0.61	0.40	0.60	0.21	0.53	0.69	0.75
Student (df: 10)	0.35	0.84	0.60	0.76	0.17	0.43	0.55	0.60

Note: 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 $N = 1000$ samples for each of the four distributions considered for the errors ε_t (see first column) and each of the three considered sample sizes ($T = 100$, 200 or 500). For each simulated sample, we employ the Maximum Likelihood approach (Panel (a)) and the GMM approach (Panel (b)) to estimate $(\theta, V(\varepsilon_t), \gamma)$, where γ characterizes the distribution of the shocks. In the ML approach, γ contains three parameters defining a Gaussian mixture of mean zero and unit variance. When the true distribution is Student, the true and pseudo distributions are different, hence the ML approach is, more precisely, a Pseudo Maximum Likelihood (PML) approach. In the GMM approach, γ contains the order-3 and order-4 cumulants of the shock distribution. Columns 2 to 5 give, respectively: the bias, the root mean-squared error, the mean absolute error and the standard deviation of the estimator of θ. The next column ($\overline{\sigma_{asy}}$) gives the mean (across the N simulations) of the asymptotic standard deviations (based on the Hessian matrix of the log-likelihood function in the ML case and on the formula given in Appendix C.3 in the GMM case). The last three columns indicate the fractions of times (among the N simulations) where the true value of θ lies within the interval $[\hat{\theta} - \phi_\alpha\sigma_{asy}, \hat{\theta} + \phi_\alpha\sigma_{asy}]$ where σ_{asy} denotes the estimate of the asymptotic standard deviation of the estimator $\hat{\theta}$ and where ϕ_α 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*	Normal. test p-values	ϕ		θ		c		μ_1		σ_1		p	
Austria	1870	0.09 / 0.00	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	0.00 / 0.00	0.60	*(0.04)*	2.56	*(0.27)*	1.46	*(0.39)*	0.40	*(1.09)*	4.39	*(1.11)*	0.03	*(0.02)*
Denmark	1820	0.02 / 0.00	−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	0.00 / 0.00	−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	0.06 / 0.00	−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	0.02 / 0.00	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	0.00 / 0.00	0.43	*(0.04)*	11.10	*(6.62)*	0.37	*(0.23)*	0.25	*(0.38)*	2.51	*(0.52)*	0.10	*(0.05)*
Netherlands	1815	0.01 / 0.00	0.54	*(0.14)*	0.37	*(0.14)*	5.95	*(1.43)*	−0.11	*(0.87)*	4.47	*(1.03)*	0.03	*(0.01)*
Norway	1830	0.00 / 0.00	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	0.00 / 0.00	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	0.02 / 0.00	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	0.05 / 0.00	0.80	*(0.17)*	0.83	*(0.15)*	3.30	*(0.23)*	−0.75	*(0.57)*	1.61	*(0.25)*	0.18	*(0.13)*
Ireland	1921	0.22 / 0.02	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	0.06 / 0.00	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	0.04 / 0.00	0.38	*(0.24)*	0.15	*(0.25)*	4.43	*(0.45)*	−2.10	*(1.48)*	2.09	*(0.62)*	0.04	*(0.03)*
Canada	1870	0.01 / 0.00	−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	0.04 / 0.00	0.04	*(0.10)*	−9.11	*(7.74)*	0.50	*(0.42)*	−0.05	*(0.15)*	1.65	*(0.28)*	0.27	*(0.13)*

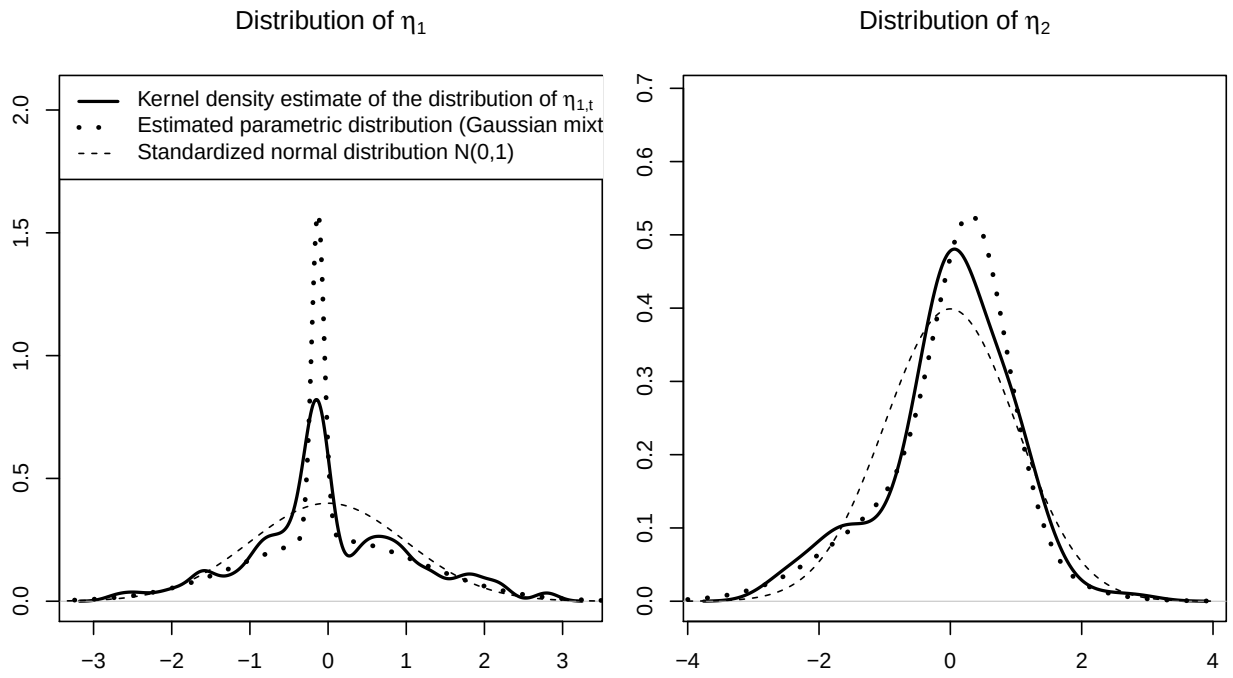
Note: This table reports the results of estimations of ARMA(1,1) processes for long historical series of per capita GDP annual growth rates (data from Bolt and van Zanden, 2014). The model is $y_t = \phi y_{t-1} + c\eta_t - \theta c\eta_{t-1}$, where the distribution of η_t is a Gaussian mixture, i.e. η_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 η_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-values of two time-series normality tests: the first is the Bai and Ng (2005) test, the second is the Lobato and Velasco (2004) test.

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

Approach:	GMM			ML	
	param.	std dev.		param.	std dev.
$\Phi_{1,[1,1]}$	0.883	*(0.540)*		0.665	*(0.202)*
$\Phi_{1,[2,1]}$	−0.425	*(0.355)*		−0.239	*(0.098)*
$\Phi_{1,[1,2]}$	0.064	*(1.323)*		−0.806	*(0.376)*
$\Phi_{1,[2,2]}$	1.153	*(0.722)*		1.727	*(0.197)*
$\Phi_{2,[1,1]}$	0.210	*(0.166)*		0.244	*(0.083)*
$\Phi_{2,[2,1]}$	−0.069	*(0.088)*		−0.036	*(0.039)*
$\Phi_{2,[1,2]}$	1.187	*(1.661)*		2.034	*(0.520)*
$\Phi_{2,[2,2]}$	−0.565	*(0.904)*		−1.264	*(0.249)*
$\Phi_{3,[1,1]}$	−0.165	*(0.172)*		−0.337	*(0.078)*
$\Phi_{3,[2,1]}$	0.019	*(0.091)*		0.037	*(0.038)*
$\Phi_{3,[1,2]}$	−1.763	*(0.995)*		−1.482	*(0.438)*
$\Phi_{3,[2,2]}$	0.600	*(0.688)*		0.609	*(0.209)*
$\Phi_{4,[1,1]}$	0.048	*(0.107)*		0.169	*(0.060)*
$\Phi_{4,[2,1]}$	0.016	*(0.072)*		−0.016	*(0.032)*
$\Phi_{4,[1,2]}$	0.604	*(0.490)*		0.315	*(0.185)*
$\Phi_{4,[2,2]}$	−0.198	*(0.320)*		−0.093	*(0.099)*
$\Theta_{1,1}$	0.814	*(0.308)*		0.406	*(0.199)*
$\Theta_{2,1}$	−0.468	*(0.085)*		0.000	*(0.098)*
$\Theta_{1,2}$	−2.733	*(0.941)*		−2.036	*(0.323)*
$\Theta_{2,2}$	1.496	*(0.455)*		1.770	*(0.153)*
$C_{1,1}$	0.298	*(0.101)*		−0.558	*(0.045)*
$C_{2,1}$	0.172	*(0.037)*		0.197	*(0.017)*
$C_{1,2}$	0.689	*(0.080)*		0.668	*(0.061)*
$C_{2,2}$	0.007	*(0.054)*		−0.012	*(0.018)*
$\kappa_{3,1}$	0.071	*(0.308)*	$\mu_{1,1}$	−0.129	*(0.021)*
$\kappa_{3,2}$	3.136	*(0.085)*	$\mu_{2,1}$	−0.426	*(0.221)*
$\kappa_{4,1}$	1.459	*(0.941)*	$\sigma_{1,1}$	0.080	*(0.012)*
$\kappa_{4,2}$	19.828	*(0.455)*	$\sigma_{2,1}$	1.247	*(0.098)*
			p_1	0.270	*(0.040)*
			p_2	0.444	*(0.117)*

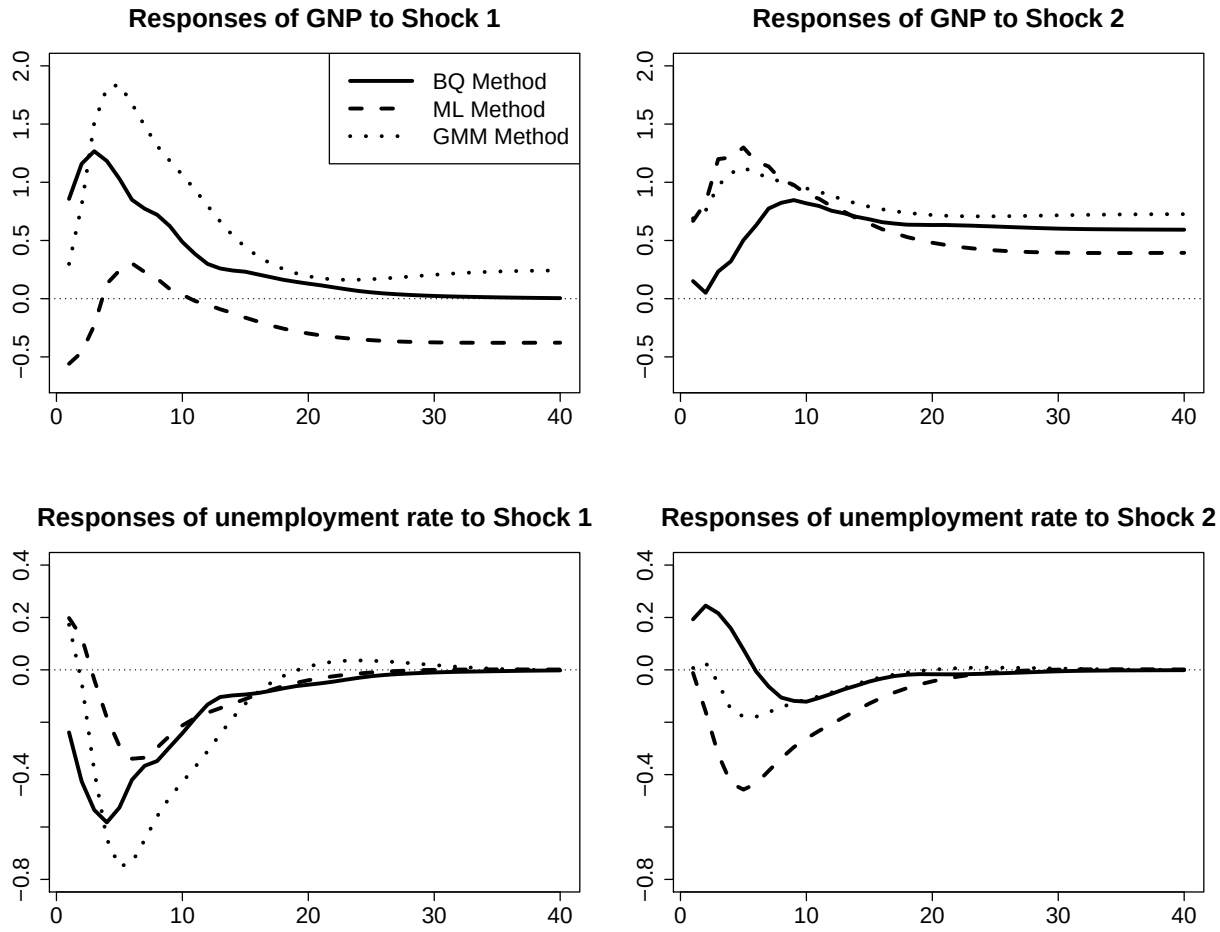
Note: This table reports the results of the estimation of bivariate VARMA(4,1) models. The dataset is the one used in Blanchard and Quah (1989); it covers the period from 1948Q1 to 1987Q4 at the quarterly frequency. The first endogenous variable is the U.S. real GNP growth (i.e. Δgnp, where $gnp = log(GNP)$) and the second endogenous variable is the unemployment rate. The model is: $\Phi(L)Y_t = (I - \Theta L)C\eta_t$, where, for $j \in \{1,2\}$, $\eta_{j,t}$ is drawn from $\mathcal{N}(\mu_{j,1}\sigma^2_{j,1})$ with probability p_j and from $\mathcal{N}(\mu_{j,2}\sigma^2_{j,2})$ with probability $1-p_j$. We impose $E(\eta_{j,t}) = 0$ and $V(\eta_{j,t}) = 1$, which implies that $\mu_{j,2}$ and $\sigma_{j,2}$ can be deduced from $\mu_{j,1}$, $\sigma_{j,1}$ and p_j. Both the Maximum Likelihood (ML) and the 2SLS-GMM approaches are employed to estimate the model (see Subsections 4.1.2 and 4.2, respectively). Asymptotic standard deviations are reported in parentheses. $\Phi_{k,[i,j]}$ is the (i,j) element of Φ_k, with $\Phi(L) = I - \Phi_1 L - \cdots - \Phi_p L^p$.

Figure 3: Estimated shock distributions



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

Figure 4: IRFs – GNP and unemployment rate



Note: This figure compares impulse response functions (IRFs) associated with models resulting from three approaches: the BQ approach (long-run restrictions *à la* Blanchard and Quah, 1989, 8-lag VAR model), the Maximum Likelihood (ML) approach (Subsection 4.1.2, VARMA(4,1) model) and the 2SLS-GMM approach (Subsection 4.2, VARMA(4,1) model). The dataset is the same as that in Blanchard and Quah (1989). In the context of the BQ approach, whereas Shock 1 (first column of plots) is interpreted as a demand shock, Shock 2 (second column of plots) is interpreted as a supply shock. In BQ, by construction, the long-run impact of the demand shock (Shock 1) on real GNP is null. In the 2SLS-GMM and the ML approaches, Shock 1 is defined as the shock having the lower influence on the long-run GNP variance. For the GNP variable, we report the cumulated impacts of the shocks on the (log) GNP growth rate.

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.

Bandi, F. M., B. Perron, A. Tamoni, and C. Tebaldi, 2018, "The Scale of Predictability," *Journal of Econometrics (forthcoming)*.

Beaudry, P., P. Fève, A. Guay, and F. Portier, 2015, "When is Nonfundamentalness in VARs a Real Problem? An Application to News Shocks," NBER Working Papers 21466, National Bureau of Economic Research, Inc.

Bernanke, B. S., 1986, "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25(1), 49–99.

Blanchard, O. J., J.-P. L'Huillier, and G. Lorenzoni, 2013, "News, Noise, and Fluctuations: An Empirical Exploration," *American Economic Review*, 103(7), 3045–3070.

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

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

Bonhomme, S., and J.-M. Robin, 2009, "Consistent Noisy Independent Component Analysis," *Journal of Econometrics*, 149(1), 12–25.

Boudt, K., D. Cornilly, and T. Verdonck, 2018, "Nearest Comoment Estimation with Unobserved Factors," working paper.

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

Box, G. E. P., and D. A. Pierce, 1970, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65(332), 1509–1526.

Breidt, F. J., and R. A. Davis, 1992, "Time-Reversibility, Identifiability and Independence of Innovations for Stationary Time Series," *Journal of Time Series Analysis*, 13(5), 377–390.

Brockwell, P. J., and R. A. Davis, 1991, *Time Series: Theory and Methods*. 2nd ed., Springer-Verlag New York, 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 Non-invertible MA(1) Process," *Journal of Forecasting*, 30(1), 6–30.

Cheng, Q., 1992, "On The Unique Representation of Non-Gaussian Linear Processes," *The Annals of Statistics*, 20, 1143–1145.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson, 2006, "Alternative Procedures for Estimating Vector Autoregressions Identified with Long-Run Restrictions," *Journal of the European Economic Association*, 4(2-3), 475–483.

——— , 2007, "Assessing Structural VARs," in *NBER Macroeconomics Annual*. MIT Press, Cambridge, pp. 1–106.

Cochrane, J. H., 1994, "Shocks," *Carnegie-Rochester Conference Series on Public Policy*, 41, 295–364.

Comon, P., 1994, "Independent Component Analysis, A New Concept?," *Signal Processing*, 36, 287–314.

Darmois, G., 1953, "Analyse Générale des Liaisons Stochastiques: Etude Particulière de l'Analyse Factorielle Linéaire," *Revue de l'Institut International de Statistique*, 21(1/2), 2–8.

Davidson, J., 1994, *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

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.

Engle, R. F., and C. W. J. Granger, 1987, "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55(2), 251–276.

Erceg, C., G. L., and C. Gust, 2005, "Can Long Run Restriction Identify Technology Shocks?," *Journal of the European Economic Association*, 3, 1237–1278.

Eriksson, J., and V. Koivunen, 2004, "Identifiability, Separability and Uniqueness of Linear ICA Models," *IEEE Signal Processing Letters*, 11, 601–604.

Evans, C., and D. Marshall, 2011, "Fundamental Economic Shocks and the Macroeconomy," Discussion Paper 351, Central Bank of Chile.

Faust, J., and E. Leeper, 1997, "When Do Long Run Identifying Restrictions Give Reliable Results?," *Journal of Business and Economic Statistics*, 15, 345–353.

Fernandez-Villaverde, J., J. Rubio-Ramirez, T. Sargent, and M. Watson, 2007, "A,B,C's (and D)'s for Understanding VAR's," *American Economic Review*, 97, 1021–1026.

Fève, P., 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., and L. Gambetti, 2014, "Sufficient Information in Structural VARs," *Journal of Monetary Economics*, 66(C), 124–136.

Francis, N., and V. Ramey, 2005, "Is Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited," *Journal of Monetary Economics*, 52, 1379–1399.

Gagliardini, P., and C. Gouriéroux, 2018, "Identification by Laplace Transforms in Nonlinear Panel or Time Series Models with Unobserved Stochastic Dynamic Effects," *Journal of Econometrics*, forthcoming.

Galbraith, J., A. Ullah, and V. Zinde-Walsh, 2002, "Estimation of the Vector Moving Average Model by Vector Autoregression," *Econometric Reviews*, 21(2), 205–219.

Gali, J., and P. Rabanal, 2005, "Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit Postwar US Data?," in *NBER Macroeconomics Annual*. MIT Press, Cambridge, vol. 19.

Ghurye, S., and I. Olkin, 1961, "A Characterization of the Multivariate Normal Distribution," *Annals of Mathematical Statistics*, 33, 533–541.

Giannone, D., and L. Reichlin, 2006, "Does Information Help Recovering Structural Shocks From Past Observations?," *Journal of European Economic Association*, 4, 455–465.

Gospodinov, N., and S. Ng, 2015, "Minimum Distance Estimation of Possibly Noninvertible MA Process," *Journal of Business and Economic Statistics*, 33, 403–417.

Gouriéroux, C., and J. Jasiak, 2016, "Filtering and Prediction in Noncausal Processes," *Journal of Time Series Analysis*, 37, 405–430.

——— , 2017, "Noncausal Vector Autoregressive Process: Representation, Identification and Semi-Parametric Estimation," *Journal of Econometrics*, 200, 118–134.

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.

Jardet, C., A. Monfort, and F. Pegoraro, 2013, "No-Arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth," *Journal of Banking & Finance*, 37(2), 389–402.

Koop, G., M. H. Pesaran, and S. M. Potter, 1996, "Impulse response analysis in nonlinear multivariate models," *Journal of Econometrics*, 74(1), 119–147.

Kydland, F., and E. Prescott, 1982, "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345–1370.

Lanne, M., H. Lütkepohl, and K. Maciejowska, 2010, "Structural Vector Autoregressions with Markov Switching," *Journal of Economic Dynamics and Control*, 34, 121–131.

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

Lii, K.-S., and M. Rosenblatt, 1992, "An Approximate Maximum Likelihood Estimation for Non-Gaussian Non-minimum Phase Moving Average Processes," *Journal of Multivariate Analysis*, 43, 272–299.

——— , 1996, "Maximum Likelihood Estimation for Non-Gaussian Nonminimum Phase ARMA Sequences," *Statistica Sinica*, 6, 1–22.

Lippi, M., and L. Reichlin, 1993, "The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment," *American Economic Review*, 83, 644–652.

——— , 1994, "VAR Analysis, Nonfundamental Representations, Blaschke Matrices," *Journal of Econometrics*, 63, 307–325.

Ljung, G. M., and G. E. P. Box, 1978, "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 65(2), 297–303.

Lobato, I. N., and C. Velasco, 2004, "A Simple Test of Normality for Time Series," *Econometric Theory*, 20(4), 671–689.

——— , 2018, "Frequency Domain Minimum Distance Inference for Possibly Noninvertible and Noncausal

ARMA Models," *Annals of Statistics*, 46(2), 555–579.

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

Lütkepohl, H., 2013, "Identifying Structural Vector Autoregression via Changes in Volatility," in *VAR Models in Macroeconomics, Essays in Honor of Christopher Sims, Advances in Econometrics*. Thomas B. Fomby , Lutz Kilian , Anthony Murphy, editors, Emerald Group Publishing Limited, vol. 32.

——— , 2014, "Fundamental Problems with Nonfundamental Shocks," in *Essays in Nonlinear Time Series Econometrics*. N., Haldrup, M., Meltz and P., Saikkonen, editors, Oxford University Press.

Lütkepohl, H., and B. Theilen, 1991, "Measures of Multivariate Skewness and Kurtosis for Tests of Nonnormality," *Statistical Papers*, 32(1), 179–193.

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

Pesaran, H. H., and Y. Shin, 1998, "Generalized Impulse Response Analysis in Linear Multivariate Models," *Economics Letters*, 58(1), 17–29.

Rosenblatt, M., 2000, *Gaussian and Non Gaussian Linear Time Series and Random Fields*. Springer Verlag, New York edn.

Rozanov, Y., 1960, "Spectral Properties of Multivariate Stationary Processes and Boundary Properties of Analytic Matrices," *Theory of Probability and its Applications*, 5, 362–376.

——— , 1967, *Stationary Random Processes*. Holden Day edn.

Rubio-Ramirez, J., D. Waggoner, and T. Zha, 2010, "Structural Vector Autoregression: Theory of Identification and Algorithms for Inference," *Review of Economic Studies*, 77, 665–696.

Shapiro, S. S., and M. B. Wilk, 1965, "An analysis of variance test for normality (complete samples)?," *Biometrika*, 52(3-4), 591–611.

Sims, C., 1980, "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

——— , 1986, "Are Forecasting Models Usable for Policy Analysis?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 1–16.

——— , 1989, "Models and their Uses," *American Journal of Agricultural Economics*, 71, 489–494.

Sims, E. R., 2012, "News, Non-Invertibility, and Structural VARs," Working Papers 013, University of Notre Dame, Department of Economics.

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

Söderlind, P., 1999, "Solution and Estimation of RE Macromodels with Optimal Policy," *European Economic Review*, 43(4-6), 813–823.

Uhlig, H., 2005, "What Are the Effects of Monetary Policy on Output? Results From an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381–419.

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

Weiss, G., 1975, "Time-Reversibility of Linear Stochastic Processes," *Journal of Applied Probability*, 12(4), 831–836.

Wu, R., and R. A. Davis, 2010, "Least Absolute Deviation Estimation for General Autoregressive Moving Average Time-Series Models," *Journal of Time Series Analysis*, 31(2), 98–112.