Revisiting Identification and Estimation in Structural VARMA Models

C., GOURIEROUX [1] and A., MONFORT [2]

(September 2014, revised February 2015)

The first author gratefully acknowledges financial support of the chair ACPR/Risk Foundation : Regulation and Systemic Risks, and of the Global Risk Institute. We thank M. Barrigozi, M. Deistler, M. Dungey, A. Guay, A. Hyvarinen, J. Jasiak, M. Lippi, H. Lütkepohl, A. Melino, H. Uhlig for helpful comments. The views expressed in this paper are ours and do not necessarily reflect the views of the Banque de France.

[1]CREST and University of Toronto.
[2]CREST and Banque de France.

Revisiting Identification in Structural VARMA Models

Abstract

The basic assumption of a structural VARMA model (SVARMA) is that it is driven by a white noise whose components are uncorrelated (or independent) and can be interpreted as economic shocks, called "structural" shocks. These models have to face two kinds of identification problems. The first identification problem is "static" and is due to the fact that there is an infinite number of linear transformations of a given random vector making its components uncorrelated. The second identification problem is "dynamic" and is a consequence of the fact that the SVARMA process may have a non invertible AR and/or MA matrix polynomial but, still, has the same second order properties as a VARMA process in which both the AR and MA matrix polynomials are invertible (the fundamental representation). Moreover the standard Box-Jenkins approach 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 underlying the Box-Jenkins type approaches, and that both identification challenges are solved in a non Gaussian framework. We also develop simple new parametric and semi-parametric estimation methods when there is nonfundamentalness in either the moving-average, or the autoregressive dynamics, and discuss the derivation of impulse response functions.

Keywords : Structural VAR, Fundamental Representation, Noncausal Process, Shock, Impulse Response Function, Incomplete Maximum Likelihood, Pseudo-Maximum Likelihood.

1 Introduction

The basic assumption of a structural VARMA model (SVARMA) is that it is driven by a white noise whose components are uncorrelated (or independent) and are interpreted as economic shocks[3], called "structural" shocks. These models have to face two kinds of identification problems.

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

A second identification problem comes from the fact that the stationary SVARMA process may feature a non-invertible autoregressive (AR) or, more often, a non-invertible moving average (MA) matrix polynomial. The latter situation, called nonfundamentalness may occur, for instance, when the SVARMA is deduced from business cycle models [see e.g. Kydland, Prescott (1982), Francis, Ramey (2005), Gali, Rabanal (2005)], or from log-linear approximations of Dynamic Stochastic General Equilibrium (DSGE) models involving rational expectations [see e.g. Hansen, Sargent (1991), Smet, Wouters (2003), Christiano, Eichenbaum, Vigfussen (2007), Leeper, Walker, Yang (2013)]. Typically the matrix MA polynomial is not invertible and the shock vector is not simply linearly linked to the (linear) innovation of the process [see e.g. Lippi, Reichlin (1993), (1994)]. Moreover the nonfundamental SVARMA process has exactly the same second-order properties as another VARMA process with an invertible MA part (the fundamental representation) and, in the Gaussian case, both processes are observationally equivalent. This creates a dynamic identification problem, which is exarcerbated by the fact that the standard Box-Jenkins approach (i.e. the Gaussian Pseudo Maximum Likelihood method based on a VAR approximation

[3]Our paper will not consider the debate about how structural are the parameters and the shocks in SVARMA models [see e.g. Pesaran, Smith (2011) and the reference therein].

of the VARMA process obtained by inverting the MA part) [Box, Jenkins (1970)] provides a consistent estimation of the fundamental representation and, therefore, may lead to misspecified Impulse Response Functions.

The aim of this paper is to explain that these difficulties are due to the Gaussian assumption underlying the Box-Jenkins (BJ) type approaches, and that these identification issues are solved in a non Gaussian framework. We also introduce simple semi-parametric and parametric estimation approaches when there is nonfundamentalness either in the moving-average, or in the autoregressive dynamics and discuss the derivation of impulse response functions.

In Section 2, we consider a vector autoregressive moving average process, with roots of the autoregressive and/or moving average polynomials that are not necessarily outside the unit circle. We focus on its two-sided moving average and autoregressive representations and on the fact that the economic shocks are not necessarily interpretable in terms of causal linear innovations[4]. We review the different types of nonfundamental representations in the moving average dynamics given in the literature. We also explain that ill-located roots can arise in the autogressive dynamics, when rational expectations are included in the model. In particular we show that there exists an infinite number of stationary solutions in any rational expectation model, if we do not impose these solutions to have a finite variance. Some of these solutions have noncausal stationary components. Next we discuss the identification issue in the Gaussian case and explain why the standard Box-Jenkins approach based on Gaussian pseudo-likelihood suffers from these identification issues.

Section 3 is the core of the paper. We consider the case of non Gaussian observable SVARMA processes, based on the recent literature on the so-called noncausal processes [see e.g. Brockwell, Davis (1991), Rosenblatt (2000) for an introduction]. We explain that the standard static and dynamic identification problems encountered in the SVARMA analysis disappear when the shocks are not Gaussian. In Section 4 we suggest new semi-parametric estimation methods to improve the standard SVAR methodology. We first consider a semi-parametric SVARMA with nonfundamental representation in the moving average dynamics only and introduce a two step moment approach to estimate the autoregressive and moving average parameters as well as the distributions of the errors. Then we consider a SVAR model with pos-

[4]See Appendix 1 for precise definitions of the different notions of innovations.

sibly ill-located roots in the AR lag polynomial and explain how the AR parameters can be consistently estimated by covariance estimators.

The construction of Impulse Response Functions (IRF) is discussed in Section 5. We emphasize the significant differences in these constructions when the ill-located roots concern the MA and the AR dynamics, respectively. In particular nonlinear IRF have to be considered for ill-located roots in the AR component. Section 6 concludes. Some complements are gathered in appendices. In particular we introduce and study new parametric estimation approaches of SVARMA models with nonfundamentalness, that are the Incomplete Maximum Likelihood (IML) and the Simulated Pseudo Maximum Likelihood (SPML) approaches, respectively.

2 Dynamic Linear Model and Nonfundamentalness

2.1 The dynamic model

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

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

where Y_t is the n-dimensional vector of observations at date t, ε_t is the n-dimensional vector of errors,

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

and the matrix autoregressive and moving average lag-polynomials are of degree p and q, respectively. [5]

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

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

a.1 : Assumption on errors

i) The errors ε_t are independently and identically distributed. They have some moments : $\exists s > 0, E(\|\varepsilon_t\|^s) < \infty$.

ii) They can be written as $\varepsilon_t = C\eta_t \Leftrightarrow \eta_t = C^{-1}\varepsilon_t$, where the components of η_t are mutually independent.

Assumption a.1 i) on the errors is standard in the literature. For instance for $s = 2$, the existence of second-order moments underlies the Box-Jenkins methodology. We extend the set of possible errors to allow for errors with infinite variance, or even no mean, in line with the financial literature. This will also allow for considering stationary solutions of rational expectation models containing explosive speculative bubbles (see Section 2.3). Assumption a.2 ii) is needed for allowing separate shocks on the system when defining the impulse response functions.

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, Deistler (1988), Chap 2 for more details]. This condition will greatly simplify the discussions in the next sections. It is often forgotten in structural settings and it might be necessary to test for the minimality of the representation. This is clearly out of the scope of this paper.[6]

a.3 : Assumption on the observable process

All the roots of *det* $\Phi(L)$ have a modulus different from 1 and process (Y_t) is strongly stationary.

Under Assumptions a.1 - a.3, the linear dynamic system (2.1)-(2.2) has a unique strongly stationary solution, such that $E\|Y_t\|^s) < \infty$ [see e.g. the discussion in Gourieroux, Zakoian (2014) a,b].

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

[6]See Deistler, Schrader (1979) for a study of identifiability without coprimeness, and Gourieroux, Monfort, Renault (1989) for the test of coprimeness, i.e. common roots, for one-dimensional ARMA processes.

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

$$\begin{aligned} \Phi(L)Y_t &= \Theta(L)\varepsilon_t \\ \iff Y_t &= \Phi(L)^{-1}\Theta(L)\varepsilon_t \equiv \sum_{j=0}^{\infty} A_j L^j \varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}. \end{aligned} \tag{2.3}$$

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

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

From the macroeconomic literature we know that SVARMA models do not always have roots of the moving average or autoregressive operators located outside the unit circle (see Sections 2.2 and 2.3). In that case, it is still possible to invert the autoregressive polynomial operator to get a two-sided moving average representation as the stationary solution :

$$Y_t = \sum_{j=-\infty}^{+\infty} A_j \varepsilon_{t-j}. \tag{2.4}$$

Similarly, if $\det \Theta(z)$ has no roots on the unit circle, we get a two-sided autoregressive representation :

$$\sum_{j=-\infty}^{\infty} B_j Y_{t-j} = \varepsilon_t.$$

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

$$(1 - \varphi L)y_t = (1 - \theta L)\varepsilon_t, \tag{2.5}$$

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

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

$$\Leftrightarrow \quad (1-\frac{1}{\theta}L^{-1})^{-1}(1-\varphi L)y_t = -\theta L\varepsilon_t. \tag{2.6}$$

This formula reveals that :

i) the process has a two-sided autoregressive representation;

ii) the error appropriate for a mathematical analysis of the dynamics is rather the time shifted process $\tilde{\varepsilon}_t = -\theta L\varepsilon_t = -\theta\varepsilon_{t-1}$ [see e.g. Lanne, Saikkonen (2013), Gourieroux, Jasiak (2014)], than ε_t itself. Equation (2.6) becomes :

$y_t - \varphi y_{t-1} = -\frac{1}{\theta}\tilde{\varepsilon}_{t+1} + \tilde{\varepsilon}_t$, with a future value of process $\tilde{\varepsilon}$ on the right hand side.

iii) The error term ε_t is not the causal innovation of process y_t (see Appendix 1), i.e.

$\varepsilon_t \neq y_t - E[y_t|y_{t-1}, y_{t-2}, \ldots]$, if this conditional expectation exists.

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 $\det\Phi(z)$ and $\det\Theta(z)$ are all outside the unit circle. Under this condition, we say that process Y_t has a fundamental (causal) VARMA representation [see e.g. Hansen, Sargent (1980), p18, (1991), p79, and Lippi, Reichlin (1994) for the introduction of this terminology in the macroeconometric literature].[7] Otherwise, ε_t is not equal to the causal innovation and future values are involved in either the pure moving average, or the pure autoregressive representation of the process. It is a nonfundamental VARMA representation. As noted in Lanne, Saikkonen (2011) "nonfundamental solutions have typically been represented by noninvertible moving average models. However, noncausal autoregressive and noninvertible moving average models closely

[7]The term "fundamental" is likely due to Kolmogorov and appears in Rozanov (1960), p367, (1967), p56, to define the "fundamental process", that is the second-order white noise process involved in the Wold decomposition of a weak stationary process. At any date t, the information contained in the current and past values of the fundamental process coincides with the information contained in the current and past values of the observations.

approximate each other".[8] Thus the nonfundamentalness has to be considered in a more symmetric way in both the AR and MA polynomials defining the dynamics.

The macroeconometric specific terminology fundamental vs nonfundamental representation must be related to other standard time series terminologies. In time series a (linear) process is causal (resp. invertible), if it admits a one-sided moving average representation (resp. a one-sided autoregressive representation). For a VARMA model with $n = m$, satisfying the left coprimeness assumption, the process is causal (resp. invertible) if $\det \Phi(z)$ (resp. $\det \Theta(z)$) has all its roots outside the unit circle. Thus the VARMA representation is fundamental if it is both causal and invertible. [9]

2.2 Non-inversibility

There exist different sources of non-inversibility in SVARMA models, that is of ill-located roots of the moving average polynomial [see the discussion in Alessi, Barigozzi, Capasso (2011)].

i) **Lagged impact** Some are due to the dynamics of exogenous variables in the system. A well-known example appears in the comment of the Blanchard, Quah model [Blanchard, Quah (1989)] by Lippi, Reichlin (1993). The productivity, y_t, can be written as :

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

where ε_t denotes the shock on productivity. It may be realistic to assume that the impact of the productivity shock is not instantaneous and is maximal with a lag, i.e. that $\theta > 1$. This is the learning-by-doing hypothesis.

ii) **Non-observability** Non-inversibility can also arise from a lack of observability. Fernandez-Villaverde et al. (2007) give the example of a

[8] Indeed, when $|\theta| > 1$, the truncated version of $[1 - 1/\theta L^{-1}]^{-1}$ is equal to $1 + 1/\theta L^{-1} + \ldots + (1/\theta)^p L^{-p}$, with also all roots in L inside the unit circle, with the same modulus $1/|\theta|$.

[9] 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, Marshall (2005)]. Moreover a fundamental shock may also be defined as the non bubble component of a structural shock, i.e. the component with a permanent effect on the economy in Velinov, Chen (2013).

state space representation of the surplus in a permanent income consumption model [see Lof (2013), Section 3, for another example]. The state space model is of the following type :

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

where c_t (resp. y_t) denotes the consumption (resp. the surplus), $R > 1$ 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, we get the dynamics of y_t as :

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

Thus the root of the moving-average lag-polynomial is equal to $1/aR$. It is smaller than one when $aR > 1$. [10]

iii) **Rational expectation** Other sources of non-inversibility are the rational expectations introduced in the models. In the simple example of Hansen, Sargent (1991) the economic variable y_t, is defined as :

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

and if the information set available at date t is $I_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$, we get :

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

The root of the moving average polynomial is $(1-\beta\theta)/\theta$. This root can be larger or smaller than 1, depending on the values of β and θ.

[10] This reasoning does not hold for $a = 1$, which was precisely the case considered in Fernandez-Villaverde et al. (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.

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

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

$$y_t = \sum_{i=0}^{q} a^{q-i} \varepsilon_{t-i}. \tag{2.7}$$

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

$a \exp(2ik\pi/(q+1)), k = 1, \ldots, q$, with common modulus $|a| < 1$. Therefore $\Theta(L)$ is noninvertible and the MA polynomial is nonfundamental.

v) **Prediction error** When the variable of interest can be interpreted as a prediction error, non-inversibility may also appear [see Hansen, Hodrick (1980)]. For instance if y_t is the price of an asset at t, $E_{t-2}y_t$ can be interpreted as the futures price at $t-2$ (if the agents are risk-neutral) and, also as the forward price (if, moreover, the interest rates are zero). The spread between the spot price and the futures price is $s_t = y_t - E_{t-2}y_t$ and, if y_t is an invertible MA(2) process $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} = \Theta(L)\varepsilon_t$, we get $s_t = \varepsilon_t + \theta_1\varepsilon_{t-1} = \Theta_1(L)\varepsilon_t$, which is not necessarily invertible. 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 Noncausal process

The presence of rational expectations can also imply ill-located roots in the autoregressive dynamics, that is, a process with noncausal components. Let us consider again the univariate linear expectation model for prices à la Taylor (1977), Diba, Grossmann (1988) :

$$y_t = aE_t(y_{t+1}) + z_t, a > 0, \tag{2.8}$$

where z_t is an exogenous variable.

Since the equilibrium equation is obtained by matching the demand and supply, the information set as well as the endogenous and exogenous variables y_t, z_t depend on the structural shocks to the demand and supply. Therefore there are two underlying shocks ε_t and w_t, say. Let us assume that these shocks are independent and square integrable. For expository purpose we assume that $z_t = \varepsilon_t$ is a strong white noise. It is well known that the set of square integrable stationary equilibrium prices depends on coefficient a [see e.g. Blanchard (1978), Gourieroux, Laffont, Monfort (1982)]. If $a < 1$, there is a unique solution $y_t^0 = \varepsilon_t$. If $a > 1$, there is an infinite number of solutions obtained by considering the convex combinations of the forward solution y_t^0 and of the perfect foresight or backward solution :

$$y_t^1 = ay_{t+1}^1 + \varepsilon_t \Longleftrightarrow y_t^1 = \frac{L}{L-a}\varepsilon_t.$$

In fact **this dynamic rational expectation equilibrium model has many more stationary solutions, if we do not restrict the stationary equilibrium price to be square integrable**. In order to reveal other solutions, let us first recall the properties of the noncausal stable AR(1) process [see Gourieroux, Zakoian (2014), a].

A noncausal stable AR(1) process is a strongly stationary process satisfying the noncausal autoregression :

$$y_t^* = \rho y_{t+1}^* + \varepsilon_t^*, |\rho| < 1, \tag{2.9}$$

where the ε_t^* are i.i.d. variables with a stable distribution with stability index $s, 0 < s < 1$. This process is also a Markov process in calendar time and admits a nonlinear autoregressive representation [see e.g. Rosenblatt (2000) and Appendix 1] :

$$y_t^* = g(y_{t-1}^*, \eta_t; s, \rho), \tag{2.10}$$

where (η_t) is a strong Gaussian white noise. The noise η_t is in a one-to-one increasing relationship with y_t^* given y_{t-1}^*.

Since ε_t^* has fat tails the process y_t^* admits neither first, nor second-order unconditional finite moments, nor first and second-order moments conditional on the future. Nevertheless, it is shown in Gourieroux, Zakoian (2014)a that this process has a first-order conditional moment given the past and that this conditional expectation is given by :

$$E(y_{t+1}^*|\underline{y_t^*}) = |\rho|^{(s-1)} y_t^*, \tag{2.11}$$

where the autoregressive coefficient $|\rho|^{(s-1)}$ is larger than 1.

Let us now consider again rational expectation model (2.9), assume that the additional shock w_t is standard Gaussian and consider the noncausal stable AR(1) process defined by :

$$\begin{aligned} y_t^*(s,\rho) &= g[y_{t-1}^*(s,\rho), w_t; s, \rho] && (2.12) \\ \Leftrightarrow y_t^*(s,\rho) &= \rho y_{t+1}^*(s,\rho) + \varepsilon_t^*(s). && (2.13) \end{aligned}$$

We see that the process :

$$y_t(s,\rho) = y_t^*(s,\rho) + \varepsilon_t, \tag{2.14}$$

is a solution to dynamic rational expectation model (2.9), whenever

$$|\rho| = a^{1/(1-s)}, \tag{2.15}$$

since

$$E_t(y_{t+1}) = E_t y_{t+1}^* = |\rho|^{s-1} y_t^* = \frac{1}{a} y_t^* = \frac{1}{a}(y_t - \varepsilon_t).$$

This implies that, **even in the case** $a < 1$, **the RE equilibrium model (2.9) has an infinite number of stationary solutions**. Indeed, the stability index s can be chosen arbitrarily and by taking linear combinations, we find that any process of the type :

$$y_t = \Sigma_s \lambda(s) y_t^*[s, a^{1/(1-s)}] + \varepsilon_t,$$

is a stationary solution [11], since $E_t y_{t+1} = \frac{1}{a}\Sigma_s \lambda(s) y_t^*[s, a^{1|(1-s)}]$.

Equations (2.13)-(2.14) imply :

$$\begin{aligned} & y_t(s,\rho) = \frac{\varepsilon_t^*(s,\rho)}{1-\rho L^{-1}} + \varepsilon_t \\ \Leftrightarrow \quad & (L-\rho) y_t(s,\rho) = L\varepsilon_t^*(s) + (L-\rho)\varepsilon_t. \end{aligned} \tag{2.16}$$

and the VARMA representation of (y_t, z_t) with $z_t = \varepsilon_t$, that is the system :

$$\begin{cases} (L-\rho) y_t(s,\rho) - (L-\rho) z_t &= L\varepsilon_t^*(s), \\ \\ z_t &= \varepsilon_t, \end{cases}$$

is such that one root of the autoregressive polynomial is inside the unit circle. Such solutions are interesting, since $y_t^*(s,\rho)$ is a stationary speculative bubble component [see Gourieroux, Zakoian (2014) a]. Note that the two noises (ε_t) and $(\varepsilon_t^*(s))$ are independent. Indeed by (2.13) $y_t^*(s,\rho)$ is a nonlinear function of the current and lagged values of w_t. In particular the process $[y_t^*(s,\rho)]$ is independent of (ε_t), as well as the process $[\varepsilon_t^*(s)]$ by (2.13).

Also note that the standard Box-Jenkins approach applied to the bivariate series $[y_t(s,\rho), z_t]$ is unable to find the underlying shocks $\varepsilon_t^*(s)$. First the Box-Jenkins method assumes the square integrability of $y_t(s,\rho)$ whereas this process has no mean. Second the Box-Jenkins approach is a linear approach, which cannot accommodate the nonlinear innovation w_t in (2.12). However, it will be seen in Section 3, if both process y_t and z_t are observable, it is possible to identify the errors $(\varepsilon_t, \varepsilon_t^*(s))$ of the nonfundamental VARMA representation of y_t, and by using the link between the causal and noncausal representations of y_t to recover the economic shocks (ε_t, w_t) themselves (see Section 5). Thus, contrary to a common belief [see e.g. Lanne, Saikkonen (2011), p1], "the presence of noncausality "(does not)" necessarily indicate that the agents are able to forecast a part of the future values of the economic variable in question by information unknown to the econometrician". Indeed we have to focus on economic shocks (ε_t, w_t), and the

[11] It is usual to select the forward solution as if it were the unique solution of the rational expectation model in the case $a < 1$. This practice neglects all stationary solutions with speculative bubbles [see e.g. Pesaran, Smith (2011), p7, Leeper et al. (2013), eq. (4) and eq. (17) for recent examples of this practice].

shocks $[\varepsilon_t^*(s), \varepsilon_t]$, are just convenient tools for estimation purpose. In particular the impulse response functions [12] have to be derived by applying shocks to (ε_t, w_t), and not to nonfundamental forecast errors $[\varepsilon_t^*(s), \varepsilon_t]$. The multipliers are not derived from the nonfundamental VARMA representation, but from the associated nonlinear causal autoregressive representation [see Koop, Pesaran, Potter (1996), Gourieroux, Jasiak (2005), for the definition of impulse response functions in a general framework and Section 5].

2.4 The limits of the Gaussian approach

The Box-Jenkins methodology is based on the assumption of Gaussian errors. Let us first discuss the identification issue for a Gaussian VAR. The results are similar for a Gaussian VARMA. The observable process satisfies the autogressive dynamics :

$$\Phi(L)Y_t = C\eta_t, \eta_t \sim IIN(0, Id), \tag{2.17}$$

where $\Phi(L) = Id - \Phi_1 L \ldots - \Phi_p L^p$. The joint distribution of the process depends on parameters $C, \Phi_1, \ldots, \Phi_p$ through the first and second-order moments of the process, or equivalently through the matrix spectral density :

$$f(w) = \frac{1}{2\pi}\Phi(\exp iw)CC'\Phi(\exp(-iw))'. \tag{2.18}$$

Several pairs $\Phi(L), C$ yield the same spectral density, and, therefore, observationally equivalent Gaussian processes, leading to static and dynamic identification issues.

i) **static identification issue :** CC' can be identified, but not C itself.

ii) **dynamic identification issue :** The different polynomials in $\Phi(L)$ are deduced from the fundamental solution (i.e. such that the roots of $\det \Phi(z)$ are outside the unit circle) by an appropriate use of Blaschke matrices [see e.g., Hansen, Sargent (1981), Lippi, Reichlin (1994)].

The Box-Jenkins approach consists in estimating parameters $\Omega = CC', \Phi_1, \ldots, \Phi_p$ by maximizing the Gaussian log-likelihood (if η_t is Gaussian), or the pseudo log-likelihood (if η_t is non-Gaussian) :

[12] The impulse response analysis describes how structural shocks propagate through the macroeconomy.

$$(\hat{\Phi}, \hat{\Omega}) = \arg\max_{\Phi,\Omega} \sum_{t=1}^{T} \{-\frac{n}{2}\log 2\pi - \frac{1}{2}\log\det\Omega - \frac{[\Phi(L)Y_t]'\Omega^{-1}[\Phi(L)Y_t)]}{2}\}. \tag{2.19}$$

In other words $\hat{\Phi}$ is obtained from the O.L.S. estimators equation by equation, and $\hat{\Omega}$ is the empirical variance-covariance matrix of the associated residuals.

These (pseudo) maximum likelihood estimators converge when T tends to infinity, and the estimated autoregressive polynomial $\hat{\Phi}(L)$ converges to the fundamental solution [13] associated with $\Phi(L)$ (i.e. by inverting in an appropriate way the ill-located roots through Blaschke matrices and orthonormal transformations [14]). Therefore, it is not consistent whenever the true VAR defined in (2.18) features nonfundamentalness.

To adjust for this lack of consistency, it is proposed in the literature to deduce all the remaining solutions by applying the Blaschke matrices [15] [see Lippi, Reichlin (1991)]. This approach is computationally demanding and does not solve the dynamic identification issue.

3 The identification issues in the non Gaussian SVARMA

Let us consider the VARMA model,

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

$$\text{where : } \Phi(L) = Id - \Phi_1 L - \ldots - \Phi_p L^p, \Theta(L) = Id - \Theta_1 L - \ldots - \Theta_q L^q, \tag{3.2}$$

and the components $\eta_{1t}, \ldots, \eta_{n,t}$ of the error term are independent (and also serially independent by Assumption a.1).

[13] This result is well-known, if the errors have finite variance. It is also valid for errors with fat tails [see Davis, Resnick (1986)].

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

[15] See Leeper et al. (2013), p1123-1124 for a practical example of the use of Blaschke matrices.

This additional cross-sectional independence assumption is needed for impulse response analysis. Let us for instance assume that η_{1t} has an economic interpretation, such as a technological shock. To ensure that a change in η_{1t} captures only the effect of this shock, it is necessary to eliminate any link between η_{1t} and the other components, i.e. to assume their independence.

In this respect we follow the structural VARMA literature by assuming that all structural shocks are serially and mutually orthogonal. But we emphasize that the appropriate notion of "orthogonality" is independence. Indeed the absence of correlation usually considered in the SVARMA literature is not sufficient for deriving the impulse response function and its confidence intervals. It is also insufficient for comparing the impulse response functions deduced from a nonlinear DSGE and from its SVARMA approximation, or for applying Bayesian techniques à la Sims-Litterman . Distributional assumptions are required. [16] We point out below that both static and dynamic identification challenges are solved when the errors are independent, with at most one Gaussian error. We discuss first the static case, then the dynamic case.

3.1 How to orthogonalize the shocks

Let us first consider a static framework. When the noise is Gaussian $\eta_t \sim N(0, Id)$, we cannot identify C and η_t given the knowledge of $C\eta_t$. Indeed we have :

$$C\eta_t = C^*\eta_t^*, \text{with } \eta_t^* \sim N(0, Id),$$

as soon as :

$$C^* = CQ \text{ and } \eta_t^* = Q'\eta_t,$$

where Q is an orthogonal matrix : $QQ' = Id$.

Thus, in a Gaussian framework, there are several ways of selecting the $\eta's$ and thus of "orthogonalizing" the shocks. These different possibilities underlie the recursive identification scheme proposed by Sims (1980), (1989).

[16] For all these problems, the shocks are implicitly assumed Gaussian, but the Gaussian hypothesis is never tested in practice. This assumption is even explicit in some papers [see e.g. Forni et alii (2013)].

[17] Additional structural short run restrictions [see e.g. Bernanke (1980), Sims (1986), Rubio-Ramirez, Waggoner, Zha (2010)], and long run restrictions [18] are sometimes introduced in the applied literature to reduce the set of possibilities [see e.g. Blanchard, Quah (1989), Faust, Leeper (1997), Erceg, Guerrieri, Gust (2005), Christiano, Eichenbaum, Vigfusson (2007)], as well as sign restrictions [Uhlig (2005), Chari, Kehoe, McGrattan (2008), Mountford, Uhlig (2009)].[19] These "restrictions imposed in the usual style of identification are neither essential to constructing a model, nor innocuous", (since they can introduce misspecifications) [Sims (1980)].

Indeed there is no identification problem in a non Gaussian framework as shown in the following Proposition [see Comon (1994), Theorem 11 for square integrable variable, Eriksson, Koivunen (2004), Theorem 3 ii) and the references therein, for variables with fat tails] :

Proposition 1 : Let us consider two vectors η_t, η_t^* of the same size n, with independent components, with continuous distributions and satisfying a linear relationship : $\eta_t^* = M\eta_t$, where M is invertible. Then, these components are such that $\eta*_{i,t} = \sigma_i \eta_{\pi(i),t}, i = 1, \ldots, n$, where π is a permutation of $\{1, \ldots, n\}$, and σ_i a scale parameter, possibly negative, whenever at most one component of η_t is Gaussian.

The proof is a direct consequence of the Darmois, Skitovich characterization of the multivariate normal distribution [see Darmois (1953), Skitovich (1953), Ghurye, Olkin (1961), Theorem p 533, Kagan, Linnik, Rao (1973), Th 10.3.1]. This identification result underlies independent component analysis (ICA), which is the analogue of the principal component analysis (PCA) when components are required to be independent rather than being simply uncorrelated [see e.g. Hyvarinen, Karhunen, Oja (2001)]. In practice consistent estimators of matrix C are obtained by considering cross-moment conditions, or tail properties or Pseudo-Maximum Likelihood approach [Gourieroux, Monfort (2015)].

[17] See also Klein (2000).

[18] Typically permanent shocks to output are associated with technology shocks.

[19] An alternative consists in leaving the linear dynamic framework by considering Markov Switching SVAR [see Lanne, Lutkepohl, Maciejowska (2010), Lutkepohl (2013), Herwatz, Lutkepohl (2014), Velinov, Chen (2013)]. This extended framework allows to test the identification restrictions. In this note we will stay in a pure SVARMA framework.

An important consequence of Proposition 1 is the following : Let us consider the error terms $\varepsilon_t = C\eta_t$ in the VARMA model (3.1), with independent components for the η_t, at most one Gaussian component and a lower triangular matrix T (not diagonal). Then the components of $T\varepsilon_t$ cannot be mutually independent except in a very special case. Indeed the random vectors $\eta_t^* = T\varepsilon_t$ and η_t satisfy a one-to-one linear relationship $\eta_t^* = TC\eta_t$. Thus by Proposition 1, we deduce that : $TC = \wedge P$, where $\wedge$ is a diagonal matrix and P a permutation matrix. Thus matrix C must be the product of a triangular matrix by a permutation matrix, which is a very specific situation. **This implies that, in a non-Gaussian framework, the recursive scheme proposed by Sims can be used to find uncorrelated components, but not, in general, independent components.**

3.2 Identification of the relevant nonfundamental representation

We have seen in Section 3.1 above that, in a non-Gaussian framework, there is no real identification problem for the "static" part of the SVARMA model. What can be said about the identification of its dynamic part? The question concerns the uniqueness of the two-sided moving average representation :

$$Y_t = \sum_{j=-\infty}^{+\infty} A_j \eta_{t-j}, \text{ with } \sum_{j=.\infty}^{+\infty} ||A_j|| < \infty, \tag{3.3}$$

where the components of η_t are mutually independent. The following result (Chan, Ho, Tong (2006), Theorem 1) extends the Darmois, Skitovich result to Gaussian processes [20] and is valid under the assumption of non-degenerate transfer function :

Assumption a.4 : The determinant of the transfer function

$\tilde{A}(w) = \sum_{j=-\infty}^{+\infty} A(j) \exp(-ijw)$ is not zero almost everywhere on the interval $(-\pi, \pi)$.

This assumption is introduced to avoid linear dependence between the columns of matrix polynomial $A(L)$. If, for instance $Y_t = A_1(L)\eta_{1t} + A_2(L)\eta_{2t}$,

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

say, with $A_2(L) = \lambda A_1(L)$, we can write $y_t = A_1(L)(\eta_{1t} + \lambda\eta_{2t})$, and replace the bidimensional noise (η_{1t}, η_{2t}) by the one dimensional noise $\eta_{1t} + \lambda\eta_{2t}$.

Proposition 2 : Under Assumption a.4 of non-degenerate transfer function, let us consider two moving average representations of a non Gaussian process (Y_t) :

$$Y_t = \sum_{j=-\infty}^{+\infty} A_j \eta_{t-j} = \sum_{j=-\infty}^{+\infty} A_j^* \eta_{t-j}^*, \forall t.$$

Then

$$\eta_{i,t-m(i)}^* = \sigma_i \eta_{\pi(i),t}, A_{i,j}^* = \frac{1}{\sigma_i} A_{i,j+m(i)}$$

where $\eta_{i,t}$ is the i^{th}-component of η_t and $A_{i,j}$ the i^{th}-column of A_j,
for some scalars σ_i, possibly negative, integers $m(i)$ and permutation π of the set $\{1, 2, \ldots, m\}$, if one of the two following conditions is satisfied :
Condition C_1 : the components of η_t are identically distributed.
Condition C_2 : each component of η_t has a nonzero r^{th} cumulant, with $r \geq 3$, and a finite moment of order s where s is an even integer greater than[21] r.

Thus, the two-sided moving average representation is unique up to a permutation, a change of scale and a time shift, possibly depending on the component.

To understand how Proposition 2 solves the dynamic identification issue, let us consider a bivariate MA(1) model :

$$\begin{cases} y_{1t} &= \eta_{1t} - \theta_{1,1}\eta_{1,t-1} - \theta_{1,2}\eta_{2,t-1}, \\ \\ y_{2t} &= \eta_{2t} - \theta_{2,1}\eta_{1,t-1} - \theta_{2,2}\eta_{2,t-1}, \end{cases}$$

where $\eta_{1,t}, \eta_{2,t}, t$ varying, are identically distributed and non-Gaussian. Then the other moving average representations with i.i.d. components of the error terms are either of the type[22] :

[21] Condition C2 implies that all the components of η_t are not Gaussian.
[22] The equalities below are in distribution.

$$\begin{cases} y_{1t} &= \eta_{1,t-m(1)} - \theta_{1,1}\eta_{1,t-1-m(1)} - \theta_{1,2}\eta_{2,t-1-m(2)}, \\ y_{2t} &= \eta_{2,t-m(2)} - \theta_{2,1}\eta_{1,t-1-m(1)} - \theta_{2,2}\eta_{2,t-1-m(2)}, \end{cases}$$

or of the type :

$$\begin{cases} y_{1t} &= \eta_{2,t-m(2)} - \theta_{1,1}\eta_{2,t-1-m(2)} - \theta_{1,2}\eta_{1,t-1-m(1)}, \\ y_{2t} &= \eta_{1,t-m(1)} - \theta_{2,1}\eta_{2,t-1-m(2)} - \theta_{1,2}\eta_{2,t-1-m(2)}. \end{cases}$$

Thus they differ only by a redenomination of the errors.

A similar identification result has been recently derived when the components of η_t have fat tails [see Gourieroux, Zakoian (2014)b], and is applicable for rational expectation models with non square integrable solutions [see Section 2.3 iii)].[23]

Proposition 2 has far reaching consequences. In particular **if a non-Gaussian stationary process has a VARMA representation with serially independent errors, this representation is unique and all the second order equivalent representations have serially uncorrelated, but not independent, errors.**

4 Semi-parametric estimation of models with nonfundamentalness or noncausal component.

Proposition 2 suggests better semi-parametric estimation methods[24] than the Gaussian pseudo-maximum likelihood used in the Box-Jenkins methodology. These alternative methods will provide consistent estimators of the true two-sided moving average polynomial $A(L)$ [or equivalently the true $\Phi(L), \Theta(L)$

[23]Note that the identification result in Chen, Choi, Escanciano (2012), Theorem 1, is much less powerful. This result provides conditions to check if the fundamental representation is the right one, but cannot be used to find the correct nonfundamental representation, otherwise.

[24]Parametric estimation methods are introduced and discussed in Appendices 2 and 3.

in the VARMA representation]. They also provide consistent nonparametric estimators of the distribution of the components of the error term.

We consider successively a SVARMA causal model with nonfundamentalness in the moving average dynamics and a SVAR model with noncausal components. In both cases we introduce appropriate moment methods, based on moment restrictions deduced from the independence assumption on the components of error η. These estimation methods are consistent. The derivation of their asymptotic properties is out of the scope of this paper, but their finite sample properties are easily derived by bootstrap [see Gourieroux, Jasiak (2015) for simulating trajectories of a mixed causal/noncausal VAR(1) process].

4.1 Semi-parametric estimation of a SVARMA model with nonfundamentalness

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

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

where the components of η_t are both serially and cross-sectionally independent with $E(\eta_t) = 0, V(\eta_t) = Id$. We assume that the roots of the determinant of the autoregressive polynomial are well- located, but make no assumption on the roots of the moving average. We denote by f_j the common probability density function of the $\eta_{j,t}, t = 1, \ldots, T$. We have to consistently estimate the parameters Φ, C_0, C_1 as well as the functional parameters $f_j, j = 1, \ldots, n$.

i) Pure moving average process

Let us first consider the case $\Phi = 0$, that is, a pure moving average process and focus on the estimation of the moving average matrix coefficients C_0, C_1.

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

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

by using the independence assumptions.

The expression of this joint Laplace transform can be used to compute the expressions of the first, second, third (fourth) cross-moments of Y_t as functions of C_0, C_1 and of the first, second, third (fourth) marginal moments of the errors $\eta_{j,t}$. [25]

By focusing on the first and second moments only, we know that the corresponding GMM estimators will not provide consistent results (see the discussion in Section 2.4). But the identification of parameters and then the consistency of associated moments methods are generally achieved if we also consider higher order moments.

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

$$E(y_{j,t}y_{k,t}), j, k = 1, \ldots, n,$$

$$E(y_{j,t}^3), j = 1, \ldots, n,$$

$$E(y_{j,t}^2 y_{k,t-1}), j, k = 1, \ldots, n$$

$$E(y_{j,t}y_{k,t-1}^2), j, k = 1, \ldots, n.$$

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

The number of unknown parameters to estimate are the elements of C_0, C_1 and the third-order moments of the $\eta_{j,t}, j = 1, \ldots, n$ (since their first and second-order moments are already known and their cross third-order moments are equal to zero because of the cross-sectional independence). Thus this number of parameters is $n + 2n^2$. Therefore the order condition for identification, that is,

$n(n+1)/2 + n + 2n^2 > n + 2n^2$, is satisfied.

ii) The general case

Let us now consider the general specification (4.1). Since the process is causal, η_t and η_{t-1} are independent from Y_{t-2}. Thus we can estimate the autoregressive matrix coefficient by projecting Y_t on Y_{t-1}, with instruments Y_{t-2}. The corresponding instrumental variable (IV) estimator of Φ is :

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

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

Once Φ has been estimated, we deduce the associated IV residuals :

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

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

Then in a second-step we can apply to observations $\hat{Z}_t$ the estimation method for pure MA process introduced in the subsection above and deduce consistent estimates of C_0, C_1.[26]

Example : In the one-dimensional case :

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

the estimators of the parameters are :

$$\begin{aligned}
\hat{\varphi} &= (\Sigma_t y_t y_{t-2}/(\Sigma_t y_{t-1} y_{t-2}), \\
\hat{\theta} &= -(\Sigma y_t^2 y_{t-1})/(\Sigma y_t y_{t-1}^2), \text{[see Appendix 4 iii)]} \\
\hat{c} &= \frac{1}{T}\Sigma_t \hat{\eta}_t^{*2}, \\
\text{where } \hat{\eta}_t^* &= \frac{1-\hat{\varphi}L}{1-\hat{\theta}L} y_t,
\end{aligned}$$

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

iii) Estimation of the error distribution

Once Φ, C_0, C_1 have been estimated, we deduce consistent approximations of the errors :

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

[26] If $C_1 = 0, C_0$ can be directly estimated by ICA [see e.g. Chen, Choi, Escanciano (2012)].

where the inverse $(\hat{C}_0 + \hat{C}_1 L)^{-1}$ has to be computed carefully, with backward expansions for roots larger than one, forward expansions, otherwise.

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

4.2 Semi-parametric estimation of a SVAR model with noncausal component.

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

$$Y_t = \Phi Y_{t-1} + \varepsilon_t, \tag{4.5}$$

with i.i.d. non Gaussian error terms, with zero mean. This process can always be written as :

$$Y_t = A_1 Y_{1t}^* + A_2 Y_{2t}^*,$$

where Y_{1t}^* is a pure autoregressive causal process of order 1 and dimension K and Y_{2t}^* a pure noncausal autoregressive process of order 1 and dimension $n - K$ [see Gourieroux, Jasiak (2015)]. Thus this process generally contains causal as well as noncausal components.

The Box-Jenkins approach, based on a Gaussian pseudo-likelihood, estimates the Seemingly Unrelated Regression model (4.14) by ordinary least squares, or equivalently by an instrumental variable approach with instrument Y_{t-1}. This method is valid if Y_{t-1} and ε_t are uncorrelated, that is, if all the roots of $\det \Phi(z)$ are outside the unit circle. It is not valid if some roots are inside the unit circle, since Y_t has a two-sided moving average representation that creates correlation between Y_{t-1} and ε_t.

However the serial independence between the errors implies moment restrictions of the type :

$$Cov[a(Y_t - \Phi Y_{t-1}), b(Y_{t-h} - \Phi Y_{t-h-1})] = 0, \tag{4.6}$$

for any lag h, and any square integrable functions a and b.[27] These restrictions can be the basis for covariance estimators. In particular, for a square integrable process Y_{t-1}, we have :

[27] When the process has fat tails as in the case of explosive bubble solutions of rational expectation models in Section 2.3, we have to select a and b functions, such as $a(y) = \exp(ia'y), b(y) = \exp(ib'y)$, say, to ensure the existence of the covariance in (4.6).

$$E[(Y_t - \Phi Y_{t-1})b(Y_{t-1} - \Phi Y_{t-2})] = 0, \text{ for any square integrable function } b. \tag{4.7}$$

In other words, the variables $b(Y_{t-1} - \Phi_0 Y_{t-2})$, where Φ_0 is the true value of the autoregressive matrix, are valid instruments for the SVAR model (4.5), and the moment conditions (4.6) are used to estimate jointly Φ and these instruments.

In practice a set of functions b has to be selected to apply the Generalized Method of Moments (GMM). When choosing only the identity function, we consider second-order moments only and the dynamic identification issue is not solved. Other functions, such as polynomial functions of degree 2, 3, have to be added to identify which representation is the right one[28],

The GMM can be applied in several steps.

i) Step 1 : Consistent estimation.

First select sets of functions a and b, and calibrate parameter Φ from the associated moment conditions without trying to introduce optimal weights. The corresponding estimator $\tilde{\Phi}$ is consistent and can be used to deduce approximated instrumental variables $\tilde{b}(Y_{t-1} - \tilde{\Phi} Y_{t-2})$, where the set of $\tilde{b}$ functions can be larger than the initial set of functions b.

ii) Step 2 : Efficiency improvement.
Apply to the model :

$$Y_t = \Phi Y_{t-1} + \varepsilon_t,$$

a Two Stage Instrumental Variable (2SIV) approach based on instruments $\tilde{b}(Y_{t-1} - \tilde{\Phi} Y_{t-2})$. This second step provides closed form estimators once $\tilde{\Phi}$ has been derived and avoids an additional nonlinear optimization.

iii) Step 3 : Diagnostic tools for serial independence.

[28] In the univariate case, it is easily seen, taking the identity function for a and b, that φ satisfies the second-order equation $\varphi_0 - \varphi(1 + \varphi_0^2) + \varphi^2 \varphi_0 = 0$, where φ_0 is the true value of φ and this equation has two solutions φ_0 and $1/\varphi_0$. Moreover, adding the condition $E[(y_t - \varphi y_{t-1})\ (y_{t-1} - \varphi y_{t-2})^2] = 0$, we get a third order equation in φ, which admits φ_0 as a solution, but not $1/\varphi_0$, in general.

Let us denote $\hat{\varepsilon}_t = Y_t - \hat{\Phi} Y_{t-1}$, the associated residuals, where $\hat{\Phi}$ denotes the 2SIV estimator. We have to check if the vectors ε_t are serially independent. We can perform tests of serial independence, based for instance on power cross-moments at any lag, for all these estimated shocks. The standard analysis of ACF and cross ACF of the Box-Jenkins approach has to be completed by considering the ACF and cross ACF on nonlinear transforms of the $\hat{\varepsilon}_t$. Typically square ACF and square cross ACF have to be considered to detect possible ARCH effects creating nonlinear dependence, or cross ACF between ε_t^2 and ε_{t-h} to detect leverage effects, and so on.[29]

If the serial independence hypothesis is rejected, the SVAR model cannot be used for computing impulse response functions. Otherwise, we have still to check if the vector of errors linearly depends on independent "structural" shocks.

iv) Step 3 : Analysis of mutual independence.
The hypothesis of mutual independence can be written as :

$H_0 = \{\exists\, C\colon\, \varepsilon_t = C\eta_t$, the components of η_t being independent $\}$.

The test of this hypothesis and the estimation of matrix C if the hypothesis is not rejected can be done by applying Independent Component Analysis [see e.g. Hyvarinen, Karhunen, Oja (2001), Hyvarinen et al. (2008), Moneta et al. (2013), Gourieroux, Monfort (2015)].

5 Shocks and Impulse Response Functions

There exist two ways to construct Impulse Response Functions, either by introducing shocks on errors, or shocks on parameters. We will focus on the first approach, which is the most frequently used by macroeconomists [see e.g. Borovicka, Hansen, Scheinkman (2014) for the second approach].

There are minimal requirements on the errors to be shocked.

i) They have to be serially and cross-sectionally independent (like the $\eta_t's$ in the models of the previous sections).[30]

[29] The generalized spectrum approach proposed in Hong (1999), Chen et al. (2012) considers complex exponential transforms $\exp(ia'\hat{\varepsilon}_t)$ and $\exp(ib'\hat{\varepsilon}_{t-h})$, say.

[30] Cross-sectional no correlation is not sufficient as shown by the following example. Let

ii) The observable macrovariables have to be functions of the current and lagged values of these errors.

iii) These errors need clear economic interpretations. In this respect the overidentifying restrictions of causality, long run behaviour, nonfundamentalness, play a crucial role.

For a causal SVARMA model with possibly non-invertible moving average the derivation of the IRF are standard. We just recall these derivations with special emphasis on dynamic and stochastic shocks.

The case of SVAR model with noncausal components is significantly different, since linear noncausal components have first to be transformed into nonlinear causal components before applying the shock. As a consequence nonlinear IRF have to be considered.

5.1 IRF for a SVARMA model with non-invertible moving average

Let us consider the model :

$$Y_t = \Phi Y_{t-1} + C_0\eta_t + C_1\eta_{t-1}, \tag{5.1}$$

where the roots of the determinant of the autoregressive polynomial are well-located and the components of the errors are independent, with distributions $f_j, j = 1, \ldots, n$, respectively.

i) The traditional IRF

The standard practice consists in considering deterministic shocks on a component of η_t, η_{1t} say, and in deriving the impact on the future expected values of the macrovariables. This error-shock methodology has been put forth in Sims (1980). For such a transitory shock at date T, denoted by $\delta\eta_{1T}$, the impacts on the Y_{T+h}, denoted by $\delta Y_{T+h},$, are computed recursively by :

$$\delta Y_{T+h} = \hat{\Phi}\delta Y_{T+h-1} + \hat{C}_0\delta\eta_{T+h} + \hat{C}_1\delta\eta_{T+h-1}, h \geq 0, \tag{5.2}$$

us consider the two errors $\varepsilon_{1,t}, \varepsilon_{2,t} = \varepsilon_{1,t}^2 - 1$, where $\varepsilon_{1,t} \sim N(0,1)$. These errors are uncorrelated : Cov $(\varepsilon_{1,t}, \varepsilon_{2,t}) = E(\varepsilon_{1,t}^3) - E(\varepsilon_{1,t}) = 0$, but in a deterministic relationship. Thus a shock on $\varepsilon_{1,t}$ has a deterministic impact on $\varepsilon_{2,t}$.

where :

$$\delta Y_{T-1} = 0, \delta\eta_{T-1} = 0, \delta\eta_T = \begin{pmatrix} \delta\eta_{1,T} \\ 0 \end{pmatrix}, \delta\eta_{T+h} = 0, h \geq 1. \tag{5.3}$$

and the parameters are replaced by their estimates.

This practice has at least two limits.

• The shock on error η_{1t} is very specific, since it corresponds to a drift on the distribution f_1 on η_{1t}. In other words the other characteristics of the distribution, like its variance, skewness, or tails have not been changed.

• By focusing on the effect on the expected future values, that is on the conditional mean profile, we have no information on the uncertainty of these impacts.[31]

ii) Stochastic shocks

The standard practice can be improved as follows. The transitory shock of interest is defined by means of some change on distribution f_1, passing from f_1 to g_1, say, or after substitution of the estimate from $\hat{f}_1$ to $\hat{g}_1$.

Then we can simulate future values of the errors without and with shocks. Without shocks, simulated errors denoted by $\eta_T^s, \ldots, \eta_{T+h}^s$ are drawn independently in the distribution $\otimes_j \hat{f}_j$. With transitory shocks, the simulated errors denoted by $\tilde{\eta}_T^s, \ldots, \tilde{\eta}_{T+h}^s$ are drawn independently in $\hat{g}_1 \otimes (\otimes_{j\neq 1}\hat{f}_j)$ for $\tilde{\eta}_T^s$, in $\otimes_j \hat{f}_j$, for $\tilde{\eta}_{T+h}^s, h \geq 1$.

Then we deduce the simulated paths of macrovariables with and without shocks $\tilde{Y}_{T+h}^s$ and Y_{T+h}^s, respectively, by applying recursively equation (5.1) with the shocked and unshocked simulated errors, respectively, after having replaced the parameters by their estimated counterparts and η_{T-1} by its predicted value $\hat{\eta}_{T-1}$.

This approach can be replicated S times and used to derive the shocked and unshocked predictive distributions of the Y_{T+h}. In particular, this approach is appropriate for deriving the impact of the stochastic shocks on the

[31] We discuss below the uncertainty due to the dynamic shocks, not the uncertainty due to the replacement of the parameters by their estimates [see e.g. Runkle (1987), Lutkepohl (1990), Sims, Zha (1999)] for this second type of uncertainty].

quantiles of the predictive distribution of any component $Y_{j,T+h}, j = 1, \ldots, n$, for instance on the term structure of the Values-at-Risk.

5.2 IRF for a SVAR model with noncausal component

The approach to compute the IRF described in Section 5.1 cannot be used for a process with noncausal components. Indeed some errors of the VAR specification are interpretable as innovations in reverse time and depend on the future of the process of interest. They cannot be directly shocked [see the minimal requirement ii) at the beginning of Section 5]. To show how to proceed, let us consider the RE equilibrium model discussed in Section 2.3.

Under the assumptions of Section 2.3, we get a bivariate VAR (1) model :

$$\begin{cases} (L-\rho)y_t - (L-\rho)z_t &= L\varepsilon_t^*, \\ z_t &= \varepsilon_t, \end{cases} \tag{5.4}$$

where the two strong white noises (ε_t) and (ε_t^*) are independent. However the two underlying structural errors are the errors on demand and supply ε_t and ω_t, not directly ε_t and ε_t^*. Thus the IRF has to be deduced from shocks on ε_t and w_t. The solution of this RE equilibrium model can be rewritten as [see eq (2.13)-(2.15)] :

$$\begin{cases} y_t &= y_t^* + \varepsilon_t, \\ z_t &= \varepsilon_t, \end{cases} \tag{5.5}$$

$$\text{where } y_t^* = g(y_{t-1}^*, w_t; s, \rho). \tag{5.6}$$

System (5.5)-(5.6) is causal, since (ε_t, w_t) is independent of the lagged values of y_t and z_t, but is nonlinear due to the autoregressive equation (5.6).

IRF can be deduced from stochastic shocks on the structural errors ε_t, w_t, by applying the nonlinear recursive scheme (5.5)-(5.6) [see Koop et al. (1996), Gourieroux, Jasiak (2005) for nonlinear IRF].

6 Concluding remarks

We have shown in this paper that the static and dynamic identification difficulties encountered in the analysis of SVARMA models are due to the poor

performance of the estimation method used in the Box-Jenkins methodology, namely the Gaussian pseudo maximum likelihood approach. This approach suffers from the lack of identification existing in the Gaussian SVARMA. Whenever the shocks are not Gaussian, the SVARMA becomes identified up to change of scale, drift on time, and permutation of equations.

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

Moreover it is important to keep the structural VARMA dynamics and not to replace it by a VAR model with higher lags. Indeed simple consistent estimation methods can be introduced for nonfundamental representation in the moving average dynamics, and are rather different from the IV approaches used when the nonfundamental representation concerns the autoregressive dynamics. Similarly the methods to simulate a path of the process and to derive the impulse response function are very different if the autoregressive dynamics is causal or not. In particular nonlinear impulse response functions may have to be used when there is a nonfundamentalness in the autoregressive dynamics.

Because it focuses on the second-order properties the SVARMA literature often introduces "incredible" identification assumptions that entail misspecification and naive interpretations of VARMA residuals. To paraphrase Sims (1980) : "Nonlinear analysis is getting easier, both because of improved techniques and because of better computational hardware. This weakens the excuse that second-order analysis has to be followed just since it is simple". Misspecified analysis could not serve as a useful tool for economic policy, nevertheless, as shown in the parametric and semi-parametric analysis developed in Section 4, SVARMA and SVAR can still be useful for economic policy, provided that the independence assumptions required for impulse response analysis are tested and the appropriate estimation methods are used.

The methods developed in this paper can be extended in several directions. First the asymptotic Gaussian distributions of the various estimators proposed can be derived and testing procedures, in particular tests of fundamentalness, can be obtained. Second, the identification and estimation results might be extended to the case of more errors than observables. Indeed identification results exist when the errors are not Gaussian [see e.g. Th

3.1. in Eriksson, Koïvunen (2004) in the static case, Gourieroux, Zakoian (2014)b for stable multivariate processes, or Gagliardini-Gourieroux (2015) for a non Gaussian factor model]. This possibility to identify the dynamics when $m > n$ and the errors are not Gaussian would be important in the discussion of the effect of omitted variables [see Giannone, Reichlin (2006), Lütkepohl (2014)].

R E F E R E N C E S

Alessi, L., Barigozzi, M., and M., Capasso (2011) : " Nonfundamentalness in Structural Econometric Models : A Review", International Statistical Review, 79, 16-47.

Andrews, B., Breidt, J., and R., Davis (2006) : "Maximum Likelihood Estimation for All-Pass Time Series Models", Journal of Multivariate Analysis, 97, 1638-1659.

Bernanke, B. (1986) : "Alternative Explanations of the Money-Income Correlation", vol 25 of Carnegie-Rochester Conference Series on Public Policy, 49-99, North-Holland.

Blanchard, O. (1978) : "Backward and Forward Solutions for Economics with Rational Expectations", DP 627, Harvard University.

Blanchard, O., and D., Quah (1989) : "The Dynamic Effects of Aggregate Demand and Supply Disturbances", American Economic Review, 79, 655-673.

Borovicka, J., Hansen, L., and J., Scheinkman (2014) : "Shock Elasticities and Impulse Response", DP University of Chicago.

Box, C., and G., Jenkins (1970) : "Time Series Analysis, Forecasting and Control", Holden-Day, San Francisco.

Breidt, F., and R., Davis (1992) : "Time Reversibility, Identifiability and Independence of Innovation for Stationary Time Series", Journal of Time Series Analysis, 13, 377-390.

Breidt, F., Davis, R., Lii, K., and M., Rosenblatt (1991) : "Maximum Likelihood Estimation for Noncausal Autoregressive Processes", Journal of Multivariate Analysis, 36, 175-198.

Brockwell, P., and R., Davis (1991) : "Time Series : Theory and Methods", 2nd ed., New-York, Springer Verlag.

Chan, K., Ho, L., and H., Tong (2006) : "A Note on Time-Reversibility of Multivariate Linear Processes"., Biometrika, 93, 221-227.

Chari, V., Kehoe, P., and E., McGrattan (2008) : "Are Structural VAR's with Long Run Restrictions Useful in Developing Business Cycle Theory ?", Journal of Monetary Economics, 55, 1337-1352.

Chen, B., Choi, J., and J.C., Escanciano (2012) : "Testing for Fundamental Vector Moving Average Representation", DP Indiana University.

Chen, M., Davis, R., and L., Song (2011) : "Inference for Regression Models with Errors from a Non-Invertible MA(1) Process", Journal of Forecasting, 30, 6-60.

Cheng, Q. (1992) : "On The Unique Representation of Non-Gaussian Linear Processes, Ann. Statist., 20, 1143-1145.

Christiano, L., Eichenbaum, M., and C., Evans (1999) : "Monetary Policy Shocks : What Have We Learned and to What End?", Handbook of Macroeconomics, Vol 1 A, 65-148, Elsevier Science, North-Holland.

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

Christiano, L., Eichenbaum, M., and R., Vigfusson (2007) : "Assessing Structural VAR's", in Acemoglu, D., Rogoff, K., Woodford, M. (eds), NBER Macroeconomics Annual 2006, MIT Press, Cambridge, 1-106.

Comon, P. (1994) : "Independent Component Analysis, A New Concept ?", Signal Processing, 36, 287-314.

Darmois, G. (1953) : "Analyse Generale des Liaisons Stochastiques", Rev. Inst. Internat. Stat., 21, 2-8.

Davis, R., and S., Resnick (1986) : "Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages", The Annals of Statistics, 14, 533-558.

Davis, R., and L., Song (2012) : "Noncausal Vector AR Processes with Application to Economic Time Series, DP Columbia University.

Deistler, M., and J., Schrader (1979) : "Linear Models with Autocorrelated Errors : Identifiability without Minimality Assumptions", Econometrica, 47, 495-504.

Diba, B., and H., Grossman (1988) : "Explosive Rational Bubbles in Stock Prices ?", The American Economic Review, 78, 520-530.

Economic Sciences Prize Committee (2011) : "Empirical Macroeconomics", The Royal Swedish Academy of Sciences, Thomas Sargent and Christopher Sims.

Erceg, C., Guerrieri, 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 (2005) : "Fundamental Economic Shocks and the Macroeconomy", DP 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., Rubio-Ramirez, J., Sargent, T., 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 SVAR", TSE working paper 12-286.

Fève, P., Matheron, J., and J., 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", CEPR Discussion paper n° DP7840.

Forni, M., Gambetti, L., Lippi, M., and L., Sala (2013) : "Noisy News in Business Cycle", mimeo.

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., Gourieroux (2015) : "Identification by Laplace Transforms in Nonlinear Panel or Time Series Models with Unobserved Stochastic Dynamic Effects", CREST DP.

Gali, J., and P., Rabanal (2005) : "Technology Shocks and Aggregate Fluctuations : How Well Does the RBC Model Fit Postwar US Data?", in Gertler, M., Rogoff, K., (eds), NBER Macroeconomics Annual 2004, Vol 19, MIT Press, Cambridge, 231-283.

Gallant, R., Rossi, P., and G., Tauchen (1993) : "Nonlinear Dynamic Structures", Econometrica, 61, 871-907.

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.

Gourieroux, C., and J., Jasiak (2005) : "Nonlinear Impulse Response Functions", Annales d'Economie et de Statistique, 78, 1-33.

Gourieroux, C., and J., Jasiak (2014) : "Filtering and Prediction in Noncausal Processes", CREST DP.

Gourieroux, C., and J., Jasiak (2015) : "Semi-Parametric Estimation of Noncausal Vector Autoregression", CREST DP.

Gourieroux, C., Laffont, J.J., and A., Monfort (1982) : "Rational Expectation in Dynamic Linear Models : Analysis of the Solutions", Econometrica, 50, 409-425.

Gourieroux, C., and A., Monfort (1996) : "Simulation Based Econometric Methods", Oxford University Press.

Gourieroux, C., and A., Monfort (2015) : "Stastistical Inference for Independent Component Analysis", CREST DP.

Gourieroux, C., Monfort, A., and E., Renault (1989) : "Testing for Common Roots", Econometrica, 57, 171-185.

Gourieroux, C., Monfort, A., and A., Trognon (1985) : "A General Approach to Serial Correlation", Econometric Theory, 1, 315-340.

Gourieroux, C., and J.M., Zakoian (2014)a : "Explosive Bubble Modelling by Noncausal Process", CREST DP.

Gourieroux, C., and J.M., Zakoian (2014)b : "On Uniqueness of Moving Average Representation of Heavy Tailed Stationary Processes", CREST DP.

Hannan, E., and M., Deistler (1988) : "The Statistical Theory of Linear Systems", Wiley, New-York, Reprint in SIAM Classics in Applied Mathematics, 2012.

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.

Hansen, L., and T., Sargent (1981) : "Exact Linear Rational Expectations Models : Specification and Estimation", Federal Reserve Bank of Minneapolis, Report 71, published in 1991 in Hansen, L., Sargent, T. (eds) "Rational Expectation Econometrics", Boulder Colorado Westview Press, 45-71.

Hansen, L., and T., Sargent (1991) : "Two Difficulties in Interpreting

Vector Autoregressions", in L., Hansen and T., Sargent, editors, Rational Expectations Econometrics, Westview Press, Boulder, 77-120.

Herwartz, H., and H., Lutkepohl (2014) : "Structural Vector Autoregressions with Markov Switching : Combining Conventional with Statistical Identification of Shocks", forthcoming Journal of Econometrics.

Hong, Y. (1999) : "Hypothesis Testing in Time Series via the Empirical Characteristic Function : A Generalized Spectral Density Approach", Journal of the American Statistical Association, 94, 1201-1220.

Hyvarinen, A., Karhunen, J., and E., Oja (2001) : "Independent Component Analysis", New-York, Wiley.

Hyvarinen, A., Shimizu, S., and P., Hoyer (2008) : "Causal Modelling Combining Instantaneous and Lagged Effects : An Identifiable Model Based on Non-Gaussianity", in Proceedings of the 25th International Conference on Machine Learning, Helsinki, 424-433.

Hyvarinen, A., Zhang, K., Shimizu, S., and P., Hoyer (2010) : "Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity", Journal of Machine Learning Research, 11, 1709-1731.

Kagan, A., Linnik, Y., and C., Rao (1973) : "Characterization Problems in Mathematical Statistics", New-York, Wiley.

Klein, P. (2000) : "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model", Journal of Economic Dynamics and Control, 24, 1405-1423.

Koop, G., Pesaran, M., and S., Potter (1996) : "Impulse Response Analysis in Nonlinear Multivariate Models", Journal of Econometrics, 74, 119-147.

Kydland, F., and E., Prescott (1982) :"Time to Build and Aggregate Fluctuations", Econometrica, 50, 1345-1370.

Lanne, M., Luoma, A., and J., Luoto (2012) : "Bayesian Model Selection and Forecasting in Noncausal Autoregressive Models", Journal of Applied Econometrics, 7, 812-830.

Lanne, M., and H., Lutkepohl (2010) : "Structural Vector Autoregressions with Nonnormal Residuals", Journal of Business and Economic Statistics, 28, 159-168.

Lanne, M., Lutkepohl, H., and K., Maciejowska (2010) : "Structural Vector Autoregressions with Markov Switching", Journal of Economic Dynamics and Control, 34, 121-131.

Lanne, M., and P., Saikkonen (2011) : "NonCausal Autoregression for Economic Time Series", Journal of Time Series Econometrics, 3, Article 2.

Lanne, M., and P., Saikkonen (2013) : "Noncausal Vector Autoregressions", Econometric Theory, 29, 447-481.

Leeper, E., Walker, T., and S., Yang (2013) : "Fiscal Foresight and Information Flows", Econometrica, 81, 1115-1145.

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

Lippi, M., and L., Reichlin (1994) : "VAR Analysis, Nonfundamental Representations, Blaschke Matrices", Journal of Econometrics, 63, 307-325.

Lof, M. (2013) : "Noncausality and Asset Pricing", Stud. Nonlinear Dyn., 17, 211-220.

Lutkepohl, H. (1990) : "Asymptotic Distribution of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models", The Review of Economics and Statistics, 72, 116-125.

Lutkepohl, H. (2013) : "Identifying Structural Vector Autoregression via Changes in Volatility", in VAR Models in Macroeconomics, Essays in Honor of Christopher Sims, Advances in Econometrics, Vol 32, 169-203.

Lutkepohl, H. (2014) : "Fundamental Problems with Nonfundamental Shocks", in Essays in Nonlinear Time Series Econometrics, ed. N., Haldrup, M., Meltz and P., Saikkonen, Oxford University Press, 198-214.

Moneta, A., Entner, D., Hoyer, P., and A., Load (2013) : "Causal In-

ference by Independent Component Analysis : Theory and Applications", Oxford Bulletin of Economics and Statistics, 75, 705-730.

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

Muller, U. (2013) : "Risk of Bayesian Inference with Misspecified Models and the Sandwich Covariance Matrix", Econometrica, 81, 1805-1849.

Pesaran, H., and R., Smith (2011) : "Beyond the DSGE Straitjacket", DP 1138, Cambridge University.

Rosenblatt, M. (2000) : "Gaussian and Non Gaussian Linear Time Series and Random Fields", Springer Verlag, New-York.

Rozanov, Y. (1960) : "Spectral Properties of Multivariate Stationary Processes and Boundary Properties of Analytic Matrices", Theory of Probability and its Applications, 5, 362-376.

Rozanov, Y. (1967) : "Stationary Random Processes", Holden-Day.

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

Runkle, D. (1987) : "Vector Autoregressions and Reality", Journal of Business and Economic Statistics, 5, 437-442.

Sims, C. (1980) : "Macroeconomics and Reality", Econometrica, 48, 1-48.

Sims, C. (1986) : "Are Forecasting Models Usable for Policy Analysis ?", Federal Reserve Bank of Minneapolis Quarterly Review, 10, 1-16.

Sims, C. (1989) : "Models and their Uses", American Journal of Agricultural Economics, 71, 489-494.

Sims, C., and T., Zha (1999) : "Error Bands for Impulse Responses", Econometrica, 67, 1113-1155.

Skitovich, V. (1953) : "On a Property of a Normal Distribution", Doklady

Academii nauk SSSR, 89, 217-219 (in Russian).

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.

Swanson, N., and C., Granger (1997) : "Impulse Response Function Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions", JASA, 92, 357-367.

Taylor, J. (1977) : "Condition for Unique Equilibrium in Stochastic Macroeconomic Models with Rational Expectation", Econometrica, 45, 1372-1385.

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 (2013) : "Are There Bubbles in Stock Prices ? Testing for Fundamental Shocks", DP 1375, DIW, Berlin.

Appendix 1

What is an innovation ?

The term "innovation" is largely used in the applied macroeconomic literature, but often with different meanings and underlying assumptions. The aim of this appendix is to discuss this (these) notion (s). For expository purpose we consider a Markov process Y_t.

i) Definitions for a square integrable process.

Two notions of innovations are considered in practice :

The weak linear innovation is : $u_t = Y_t - EL(Y_t|Y_{t-1})$, where $EL(Y_t|Y_{t-1})$ is the best mean square approximation of Y_t by an affine function of Y_{t-1}.

The strong linear innovation is : $u_t = Y_t - E(Y_t|Y_{t-1})$, where $E(Y_t|Y_{t-1})$ denotes the conditional expectation of Y_t given Y_{t-1}.

ii) Definition for an infinite variance process

In this case the notion of weak innovation has no meaning. But the notion of strong innovation still exists if the conditional distribution of Y_t given Y_{t-1} admits a first-order moment. This can occur when the process Y_t admits a first-order unconditional moment, but also sometimes for a process without a first-order unconditional moment, as shown with the example of the noncausal stable AR(1) process (see Section 2.3).

iii) Nonlinear innovation

The previous notions of innovation provide information on the accuracy of a prediction of Y_t, but no information when our interest is to predict a nonlinear transform of Y_t. A more appropriate notion is the nonlinear innovation ε_t, say, such that :

$$Y_t = g(Y_{t-1}, \varepsilon_t),$$

where the ε_t variables are i.i.d. standard normal and in an increasing relationship with Y_t for given Y_{t-1}.

This nonlinear autoregressive model is obtained as follows. Let us consider for expository purpose a one-dimensional process Y_t and denote $F(y|Y_{t-1})$ the conditional cdf of Y_t given Y_{t-1}. It is well-known that the variable :

$$U_t = F(Y_t|Y_{t-1})$$

follows a uniform distribution on $[0,1]$ for any Y_{t-1} and, in particular, U_t is independent of Y_{t-1}. By inverting the conditional cdf, we get :

$$Y_t = F^{-1}(U_t|Y_{t-1}),$$

and by defining $\varepsilon_t = \Phi^{-1}(U_t)$, where Φ is the cdf of the standard normal, we get :

$$\begin{aligned} Y_t &= F^{-1}[\Phi(\varepsilon_t)|Y_{t-1}], \\ Y_t &= g(Y_{t-1}, \varepsilon_t). \text{ (say)}, \end{aligned}$$

where g is strictly increasing in ε_t.

This notion of nonlinear innovation has three advantages :

- It can be used to evaluate the accuracy of the strong prediction of any nonlinear transform of Y_t.
- The $\varepsilon_t's$ are serially independent, whereas the weak and strong linear innovations are serially uncorrelated, but in general dependent. Thus, the weak and strong linear innovations cannot be used in general to model the errors at the basis of impulse response analysis.
- By recursive substitutions, we deduce a nonlinear moving average representation of Y_t as function of independent innovations $\varepsilon_t, \varepsilon_{t-1}...$ and by series expansion the Volterra representation of the Markov process (which generally differs from its Wold representation).

Appendix 2

Parametric estimation of noncausal SVAR: the main steps of the ML and Bayes procedures

When the distribution of the error term is parametrically specified, maximum likelihood approaches [32] and appropriate Bayesian approaches can be used, even with ill-located roots in either the autoregressive and/or the moving average dynamics [see e.g. Andrews, Breidt, Davis (2006), Lanne, Saikkonen (2013), Davis, Song (2012), Gourieroux, Jasiak (2014) for ML approach, Lanne, Luoma, Luoto (2012) for Bayesian approach] .However, these approaches require to compute the likelihood functions in each regime of ill-located roots.[33] We discuss below these approaches for SVAR models.

i) The ML procedure

Let us detail the ML procedure for a SVAR model :

$$Y_t - \Phi_1 Y_{t-1} \ldots - \Phi_p Y_{t-p} = C\eta_t, \tag{A.1}$$

say, where the $\eta_t's$ are i.i.d, with i.i.d. components. The common non-Gaussian distribution of the $\eta_{j,t}$ can be parametrized. For instance, this might be a Student distribution parametrized by the degree of freedom ν.

For disentangling the causal and noncausal components of the stationary solution of (A.1), we have to consider the roots of $\det \Phi(z)$. For stationarity we assume that there are no roots on the unit circle (Assumption a.3).

r roots are well-located and $s = np - r$ ones are ill-located.

The expression of the likelihood function of model (A.1) depends on the number of ill-located roots and this number s depends on the parameters of the model : $s = s(\Phi)$, say. Thus the log-likelihood function can be written as :

[32] Called approximate maximum likelihood (AML) in the literature on noncausal time series.

[33] The need for distinguishing the different regimes is not specific to linear dynamic models, but also arises in nonlinear autoregressive models, say, whenever the process of interest is assumed strictly stationary. As an illustration let us consider a one dimensional autoregressive model : $y_t = a(y_{t-1}, \varepsilon_t)$, say, where a is invertible with respect to y_{t-1}. So we can also write $y_{t-1} = b(y_t, \varepsilon_t)$, with $b = a^{-1}$. By recursive substitution we can write y_t function of $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_t, y_{t-k-1}$, or y_t function of $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}, y_{t+h}$. By considering a large value of h, only one of the two nonlinear moving-average representations of y_t will exist. Thus the backward or forward expressions of the stationary solution (Y_t) depends on the stability properties of functions a and $b = a^{-1}$.

$$L(\Phi, C, \nu) = \sum_{s=0}^{np} \{L_s[\Phi, C, \nu] \mathbb{1}_{s(\Phi)=s}\}, \tag{A.2}$$

where $L_s(\Phi, C, \nu)$ is the expression of the log-likelihood function in regime s.

Since there are no roots on the unit circle, the subsets of parameters Φ such that $s(\Phi) = s$ are disjoint open sets. Moreover, due to the identifiability of the MA representation, there exists no distribution of process y belonging to two different regimes, and the standard ML theory applies.

The ML approach can be implemented along the following steps,

step 1 : Estimate the autoregressive order p by the standard BJ methodology, which provides a consistent estimation of the causal SVAR representation of the process (which is a weak representation if the roots of $\det \Phi(L)$ are not all outside the unit circle).

step 2 : Then maximize the log-likelihood function (A.2). Since the likelihood function is not differentiable due to the regimes, this has to be done by first maximizing the regime specific log-likelihoods L_s under the constraints $s(\Phi) = s$, and then by selecting the regime s with the largest value of the regime specific log-likelihood function.

From a practical point of view, the difficulties come from :

i) the number of admissible regimes (but this number can be significantly diminished under structural restrictions).

ii) the derivation of closed form expressions of the regime specific log-likelihoods [see e.g. Davis, Song (2012), or Lanne, Saikkonen (2013), in special cases].

ii) The Bayes procedure

The standard Bayesian analysis of a SVAR(1) model, say, considers the likelihood function corresponding to the fundamental solution, that is,

$$l(\Phi, c, \nu) = \Pi_{t=1}^{T} \frac{1}{|detC|} g[C^{-1}(y_t - \Phi y_{t-1}), \nu],$$

where $g(.,\nu)$ denotes the joint pdf of independent variables with identical Student distributions, say, and in completing by a prior on parameters Φ, C, ν. Usually this prior distribution attributes positive weights on values of Φ such that some roots of $\det(Id - \Phi z)$ are ill-located. This standard Bayesian approach will lead to inconsistent estimators for large T. Indeed, when some roots are ill-located, the Bayesian model assumes that (Y_t) may be nonstationary whereas the process is assumed strictly stationary in each regime in our framework.

Thus an appropriate Bayesian analysis has also to disentangle the different regimes as it is done in the ML approach. More precisely, the joint p.d.f. has to be written as :

$$l(\Phi, C, \nu) = \exp L(\Phi, C, \nu),$$

where $L(\Phi, C, \nu)$ is given in (A.2), before specifying a prior, which can weight all the regimes.

Appendix 3

Incomplete Maximum Likelihood and Simulated Pseudo-Maximum Likelihood for nonfundamental SVARMA models.

We have noted in Appendix 2 the difficulty in implementing maximum likelihood or Bayesian approaches, for noncausal SVAR models. Indeed the (approximated) log-likelihood has different expressions according to the location of the roots of the determinant of the autoregressive polynomials (see however the discussion in Appendix A.4.1).

In this section we introduce new parametric consistent estimation methods which circumvent the regimes of ill-located roots in the moving average dynamics of a SVARMA process. We assume well-located roots for the autoregressive component. We try to get simple estimation methods by avoiding nonlinear optimizations with respect to a large number of parameters. This explains why we focus on multistep approaches and use when possible instrumental variable approaches. The cost of simplicity is a reasonable loss of efficiency.

To introduce the estimation approach, we first discuss the case of a one-dimensional MA(1) process before considering the general framework of a SVARMA process.

A.3.1 The one-dimensional MA(1) process

Let us consider a one-dimensional MA(1) process :

$$y_t = \varepsilon_t - \theta\varepsilon_{t-1}, \tag{A.3}$$

where the $\varepsilon_t's$ are independent.

Suppose that we observe $y_1, \ldots, y_T$. If the common distribution of the $\varepsilon_t's$ is $N(0, \sigma^2)$, the model is not identifiable. Indeed the distribution of $(y_1, \ldots, y_T)$ is multivariate normal, with zero mean and a variance-covariance matrix defined by $V(y_t) = \sigma^2(1+\theta^2)$, $cov(y_t, y_{t-1}) = -\sigma^2\theta$, and $cov(y_t, y_{t-h}) = 0, \forall h \geq 2$. Obviously the two sets of parameters (θ, σ^2) and $(\frac{1}{\theta}, \sigma^2\theta^2)$ give the same distribution.

If ε_t is not Gaussian, Proposition 2 shows that the model is identifiable[34]. Let us denote by $g(\varepsilon; \gamma)$ the common p.d.f. of the $\varepsilon_t's$, where γ is an unknown parameter.

i) When $|\theta[< 1$, we can invert equation (A.3) in the standard way in order to get ε_t as a function of current and lagged values of process Y as :

$$\varepsilon_t = \sum_{h=0}^{\infty} \theta^h y_{t-h}. \tag{A.4}$$

Then the log-likelihood function is approximated by :

$$L_1^a(\theta, \gamma) = \sum_{t=1}^{T} \log g(\sum_{h=0}^{t-1} \theta^h y_{t-h}; \gamma), \tag{A.5}$$

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

ii) When $|\theta| > 1$, equation (A.3) can still be inverted, but in reversed time. We get :

[34] See Appendix 4 for a more detailed discussion of non-identifiability of a MA(1) process and the links with reversibility.

$$\begin{aligned} y_t &= \varepsilon_t - \theta\varepsilon_{t-1} \\ \Leftrightarrow \quad -\frac{y_{t+1}}{\theta} &= \varepsilon_t - \frac{1}{\theta}\varepsilon_{t+1} \\ \Leftrightarrow \quad \varepsilon_t &= -\sum_{h=0}^{\infty} \frac{1}{\theta^{h+1}} y_{t+h+1}. \end{aligned} \tag{A.6}$$

The log-likelihood function is approximated by :

$$L_2^a(\theta,\gamma) = \sum_{t=1}^{T} \log\left\{ \frac{1}{|\theta|} g(-\sum_{h=0}^{T-t-1} \frac{1}{\theta^{h+1}} y_{t+h+1};\gamma) \right\}, \tag{A.7}$$

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

iii) Simple formulas do not exist when $|\theta| = 1$.

Then the practice of ARMA modelling consists in assuming $|\theta| \neq 1$, and in considering the approximated log-likelihood function :

$$L^a(\theta,\gamma) = L_1^a(\theta,\gamma)\mathbb{1}_{|\theta|<1} + L_2^a(\theta,\gamma)\mathbb{1}_{|\theta|>1}. \tag{A.8}$$

The optimization is performed in two steps :

First step : we optimize separately the log-likelihood on each regime to get :

$$\begin{aligned} (\hat{\theta}_1, \hat{\gamma}_1) &= \arg\max_{\theta,\gamma,|\theta|<1} L_1^a(\theta,\gamma), \\ (\hat{\theta}_2, \hat{\gamma}_2) &= \arg\max_{\theta,\gamma,|\theta|>1} L_2^a(\theta,\gamma). \end{aligned}$$

Second step : the approximated ML estimator is :

$$\begin{aligned} (\hat{\theta}, \hat{\gamma}) &= (\hat{\theta}_1, \hat{\gamma}_1), \text{if } \hat{L}_1^a = L_1^a(\hat{\theta}_1, \hat{\gamma}_1) > \hat{L}_2^a = L_2^a(\hat{\theta}_2, \hat{\gamma}_2), \\ &= (\hat{\theta}_2, \hat{\gamma}_2), \text{ otherwise.} \end{aligned}$$

The approach above has at least three drawbacks.

i) First, we do not know how to simply approximate the log-likelihood function in the case $|\theta| = 1$.

ii) Second, this approach is difficult to implement in the multidimensional case, since the number of regimes for ill-located roots increases and the expressions of ε_t as two sided autoregressive functions of y_t have to be derived.

iii) Third, the focus on the regimes when approximating the log-likelihood function gives the misleading impression of a lack of continuity of the exact log-likelihood function w.r.t. θ at $|\theta| = 1$, whereas this exact log-likelihood is continuous.[35] Let us derive it for the discussion. We get :

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

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

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

and the exact log-likelihood is :

$$L(\theta, \gamma) = \log\{\int \Pi_{t=1}^{T} g(\sum_{h=0}^{t-1} \theta^h y_{t-h} + \theta^t \varepsilon; \gamma) g(\varepsilon; \gamma) d\varepsilon.\} \tag{A.9}$$

The exact log-likelihood is generally a differentiable function of θ. However, its expression is not appropriate for deriving the asymptotic properties of the ML estimators. For such a derivation, it is usually shown that it can be approximated by appropriate sums, [$L_1^a(\theta, \gamma)$, or $L_2^a(\theta, \gamma)$ in our example] for which the standard asymptotic theories apply [see e.g. Lanne, Saikkonen (2013) for this approach applied to VAR models with noncausal components].

The approaches proposed in the next subsection are based on optimization criteria decomposed into sums which are appropriate for applying standard asymptotic theory and these criteria are sufficiently regular to avoid the introduction of the different regimes of ill-located roots.

A.3.2 Incomplete Maximum Likelihood (IML), PML and SPML approaches

i) Principle of the IML approach

[35] Such an exact log-likelihood is for instance used in the Gaussian case, with $|\theta| < 1$ by Chen, Davis, Song (2011) to analyze the properties of the ML estimator of a moving-average parameter close to non-inversibility.

The principle of Incomplete Maximum Likelihood (IML) is easily explained for the MA(1) process discussed above.

Let us separate the observations by throwing away an observation every third observation. The set of observations becomes :

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

We have a loss of information since observations $y_3, y_6, \ldots, y_{3j}, \ldots$ are not taken into account with the advantage that the pairs of observations $(y_{3j-1}, y_{3j-2}), j$ varying, are i.i.d.. Thus the exact log-likelihood function corresponding to these incomplete observations is easily computed, naturally expressed as a sum and the standard asymptotic theory will apply.

ii) IML and PML approaches

Let us now describe how the IML approach can be adapted to the multidimensional framework. For expository purpose, we consider a VARMA (1,1) model :

$$Y_t - \Phi Y_{t-1} = \varepsilon_t - \Theta \varepsilon_{t-1}, \tag{A.10}$$

where the errors ε_t are serially i.i.d., not Gaussian, with p.d.f. $g(\varepsilon, \Gamma)$. The joint distribution of the errors is chosen such that :

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

where the η_t are serially independent, with a same distribution with parameter γ, say; thus $\Gamma = (C, \gamma)$. Such a choice of the joint distribution of the components of ε_t is required in order to get mutually independent shocks when defining the impulse response functions. [36]

[36]It is often assumed that the distribution of the error term ε_t of the VAR belongs to the standard multivariate Student family, but this assumption is not appropriate for the analysis of impulse responses since this family does not include the case of independent components. We can assume, for instance, that the i^{th} component of η_t, follows a univariate Student distribution with $\nu(i)$ degrees of freedom. Additional identification restrictions can be introduced to fix the denomination of the errors, i.e. to solve the problem of multiplicity by change of scale and permutations. For instance, a permutation is excluded, if we impose either that the degrees of freedom $\nu(i)$ are in an increasing order, or that the elements of the first row of matrix C are in an increasing order. The second condition is

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

Model (A.10) is a Seemingly Unrelated Regression (SUR) model :

$$Y_t = \Phi Y_{t-1} + v_t, \tag{A.11}$$

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

Next, let us consider the joint p.d.f. of $v_t = \varepsilon_t - \Theta\varepsilon_{t-1}, v_{t-1} = \varepsilon_{t-1} - \Theta\varepsilon_{t-2}$. This p.d.f. is given by :

$$h(v_t, v_{t-1}; \Theta, \Gamma) = \int g(v_t + \Theta v_{t-1} + \Theta^2\varepsilon; \Gamma) g(v_{t-1} + \Theta\varepsilon; \Gamma) g(\varepsilon; \Gamma) d\varepsilon. \tag{A.12}$$

The two step IML estimator of (Θ, Γ) is the solution of :

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

The two step IML estimator has standard asymptotic properties, irrespective of the location of the roots of $\det \Theta(z)$. It is in particular consistent, asymptotically normal, and its asymptotic variance-covariance matrix can be derived.

All the observations of (y_t) are used if $\Phi \neq 0$, but in a non optimal way. Other consistent estimators as simple to implement and using observations in a more efficient way can be based on the same idea.

We can consider the estimator solution of :

$$(\tilde{\Theta}, \tilde{\Gamma}) = \arg\max_{\Theta,\Gamma} \sum_{t=2}^{T} \log h(Y_t - \hat{\Phi} Y_{t-1}, Y_{t-1} - \hat{\Phi} Y_{t-2}; \Theta, \Gamma). \tag{A.13}$$

preferable, since it is compatible with the limiting case of equal degrees of freedom.

This two step Pseudo Maximum Likelihood (PML) estimator is using the information on all the $v_t's$ [see e.g. Gourieroux, Monfort, Trognon (1985)]. It is also consistent, asymptotically normal, but the asymptotic variance-covariance matrix is now computed by a sandwich formula involving a general central limit theorem.

iii) SPML approach

The IML and PML likelihood functions depend on integrals of the same dimension as the VARMA system. We can approximate the integral in function h by simulation to get a two-step Simulated Pseudo Maximum Likelihood (SPML) estimator. The numerical optimization of the approximated PML criterion is :

$$(\Theta^*, \Gamma^*) = \arg\max_{\Theta,\Gamma} \sum_{t=1}^{T} \log \hat{h}(Y_t - \hat{\Phi} Y_{t-1}, Y_{t-1} - \hat{\Phi} Y_{t-2}; \Theta, \Gamma), \qquad (A.14)$$

where :

$$\hat{h}(v_t, v_{t-1}; \Theta, \Gamma) = \frac{1}{S} \sum_{s=1}^{S} \{g(v_t + \Theta v_{t-1} + \Theta^2 \varepsilon^s(\Gamma), g(v_{t-1} + \Theta \varepsilon^s(\Gamma); \Gamma)\}, \quad (A.15)$$

and the $\varepsilon^s(\Gamma)$ are drawn independently in distribution $g(\varepsilon; \Gamma)$.[37]

When the number S of simulations tends to infinity sufficiently fast w.r.t. the number T of observations, the SPML estimator has the same asymptotic properties as the PML estimator itself [see Gourieroux, Monfort (1996) for a general presentation of Simulation Based Estimation Methods].

iv) Remarks

Remark 1 : The objective functions used in the IML or PML approaches are well-specified and misspecified log-likelihood functions, respectively. They can be used to develop Bayesian approaches [see Muller (2013) for Bayesian approaches with misspecified models].

[37] As usual the same basic drawings must be kept when Γ is modified in the optimization algorithm.

Remark 2 : The IML , PML, SPML approaches provide subefficient consistent estimators. Even if the lack of efficiency is expected reasonable, we can look for more efficient estimators. This can be done as follows:

First step : Apply a IML (PML, SPML) approach. This provides a consistent estimator $\hat{\Theta}$ of Θ and of the true regime of nonfundamentalness. Then we can derive the expression of the (approximated) log-likelihood function $L_s^a(\Phi, \Theta, \Gamma)$ corresponding to this regime s. In a second step, we maximize this approximated L_s^a to derive the efficient ML estimator. The advantage of this two-step approach compared to the first step ML approach is to focus on a single regime instead of a large number. This two step approach is a kind of pre-test estimation approach, in which the one step IML (PML, SPML) is used to detect the right regime of ill-located roots.

Remark 3 : Why not consider a simulated maximum likelihood (SML) approach ? The SML estimators would be defined as :

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

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

Appendix 4

Identifiability, reversibility and estimation in 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.

i) **Limit optimization problem in the approximate ML method**

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

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

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

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

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

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

with $E_\theta(y_t|y_{-\infty}^{t-1}) = -\sum_{h=1}^{\infty}\theta^h y_{t-h}$ and $E_\theta(y_{t+1}|y_{t+2}^{\infty}) = -\sum_{h=1}^{\infty}\frac{1}{\theta^{h+1}}y_{t+h+1}$

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

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

The limit optimization problem is :

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

$$\text{with} \quad L_1^a(\theta,\sigma^2) = \log\sigma^2 + \frac{1}{\sigma^2}E_0(y_t - E_\theta(y_t|y_{-\infty}^{t-1}))^2,$$

$$L_2^a(\theta,\sigma^2) = \log(\theta^2\sigma^2) + \frac{1}{\theta^2\sigma^2}E_0(y_{t+1} - E_\theta(y_{t+1}|y_{t+2}^{\infty})^2).$$

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

Let us first assume that $|\theta_0| < 1$ and let us consider the solutions of the limit optimization problem.

In order to minimize $L_1(\theta,\sigma^2)$ on $|\theta| < 1$, we can concentrate with respect to σ^2 and get :

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

The minimum is reached for $\theta = \theta_0$ and the value at the minimum is : $\log\sigma_0^2 + 1$.

In order to minimize $L_2(\theta,\sigma^2)$ on $|\theta| > 1$, we can put $\theta^2\sigma^2 = \tilde{\sigma}^2$ we get :

$$\min_\theta \log E_0[y_{t+1} - E_{1/\theta}(y_{t+1}|y_{-\infty}^{t})] + 1.$$

The minimum is reached for $\theta = \dfrac{1}{\theta_0}$ and the minimum is again $\log\sigma_0^2 + 1$.

When $|\theta_0|$ is larger than 1, we can see that $L_1(\theta,\sigma^2)$ is optimal for $\dfrac{1}{\theta_0}$ and $L_2(\theta,\sigma^2)$ for θ_0 and we still have two inverse values of θ giving the same optimum namely $\log(\theta_0^2\sigma_0^2) + 1$. The model is not asymptotically identifiable.

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

ii) **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, 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 $(c\theta^2, \frac{1}{\theta})$ 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.

iii) **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 4.1).

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

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

and therefore :

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

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