Identification and Estimation of Structural VARMA Models Using Higher Order Dynamics

Carlos Velasco

Department of Economics, Universidad Carlos III de Madrid, Getafe (Madrid), Spain

ABSTRACT
We use information from higher order moments to achieve identification of non-Gaussian structural vector autoregressive moving average (SVARMA) models, possibly nonfundamental or noncausal, through a frequency domain criterion based on higher order spectral densities. This allows us to identify the location of the roots of the determinantal lag matrix polynomials and to identify the rotation of the model errors leading to the structural shocks up to sign and permutation. We describe sufficient conditions for global and local parameter identification that rely on simple rank assumptions on the linear dynamics and on finite order serial and component independence conditions for the non-Gaussian structural innovations. We generalize previous univariate analysis to develop asymptotically normal and efficient estimates exploiting second and higher order cumulant dynamics given a particular structural shocks ordering without assumptions on causality or invertibility. Finite sample properties of estimates are explored with real and simulated data.

ARTICLE HISTORY
Received August 2021
Accepted April 2022

KEYWORDS
Causality; Higher-order spectra; Independence of components; Invertibility; Minimum distance; Rank condition

1. Introduction
There is an increasing literature on the application of Structural VARMA (SVARMA) models for the analysis of economic data which tries to solve the identification problem of these models by incorporating information from the distribution of non-Gaussian structural shocks. This information recovered from the data can substitute, at least in part, the restrictions provided by economic theory on the impulse response functions (IRF) of endogenous variables to given shocks. Furthermore, there are many examples where it is not possible to discard nonfundamental solutions to a wide class of dynamic macroeconomic models with features affecting the flow of information used by agents to make decisions, see for example, the surveys in Alessi, Barigozzi, and Capasso (2011) and Gouriéroux, Monfort, and Renne (2020), GMR henceforth. In parallel, the identification of noncausal structural VAR models have been also investigated, see for example, Lanne and Saikkonen (2013).

The identification analysis of SVARMA models has to account for both static (Structural) and dynamic (VARMA) specification. The dynamic problem is related to the location of the determinantal roots of the VAR and VMA polynomials that lead to causal/noncausal or invertible/noninvertible solutions, respectively, guaranteeing that model errors are unpredictable, not just a serially uncorrelated white noise sequence. The static identification relates to the choice of the particular rotation of the reduced form errors that delivers the true vector of structural shocks with proper economic interpretation. Therefore, these true shocks must additionally satisfy some mutual independence condition strengthening the uncorrelation achieved by any transformation using a square root of the covariance matrix.

Under Gaussianity, SVARMA model identification is not possible in absence of further restrictions provided by economic theory because uncorrelation is equivalent to independence and therefore all the infinite sequences obtained by different versions of the lag polynomials obtained by flipping roots and/or rotating the different shocks through Blaschke orthogonal matrices (see e.g., Lippi and Reichlin 1994) would be admissible. However, under non-Gaussianity and the independence component assumption (ICA) of the structural shocks, it is known that static rotations can be identified up to permutation and sign, that is, up to labeling of the shocks, if at most one of the innovations components is Gaussian (see Comon 1994; Hyvärinen et al. 2010), while a condition on higher order cumulants and moments of serially independent errors guarantees the dynamic identification (see Chan and Ho 2004; Chan, Ho, and Tong 2006, GMR). However, these results do not lead to specific methods for designing parameter estimates and inference rules relying on the identifying assumptions which are easy to interpret and check for particular models. Instead, most of available methods are based instead on moment estimates for which local rank conditions are assumed after a basic order condition is imposed or on (Pseudo) ML procedures which have to be further justified. This is precisely the aim of this article, to provide neat inference methods exploiting efficiently global identification conditions based on a complete charac-
terization of the higher order dynamics of observables given linearity and non-Gaussianity of the sequence of structural errors.

Non-Gaussianity has been exploited for identification of dynamic models through conditions on higher order cumulants (Chan, Ho, and Tong 2006; Lanne and Luoto 2021, GMR) or spectral densities (Lii and Rosenblatt 1982), but it has also been imposed through particular probability distribution assumptions on the shocks (Lanne and Lütkepohl 2010) or with heteroscedasticity conditions, see the review in Lütkepohl and Netsuñajev (2017). Then, estimation is performed using ML (e.g., Lii and Rosenblatt 1992) for ARMA; Gouriéroux, Monfort, and Renne (2017) and Lanne, Meitz, and Saikkonen (2017) for SVAR; GMR for SVARMA models or non-Gaussian criteria like ranks (Andrews, Davis, and Breidt 2007). Methods based on higher order moments have been also developed for the univariate case in the frequency and time domains (Lii and Rosenblatt 1982; Gospodinov and Ng 2015, respectively). For multivariate models, as the identification results of Chan, Ho, and Tong (2006) provide no estimation method, GMR proposed a semiparametric 2-step approach. First, the VAR parameters are estimated using 2SLS under causality, and, second, the VMA parameters are estimated using moment conditions on linear combinations of the residuals or by non-Gaussian PMLE. The choice of restrictions on residuals moments of order 2, 3, and 4 derived from ICA aims to improve the efficiency and the chances that they are sufficient for global identification and for the usual local rank conditions. A similar approach is pursued in Lanne and Luoto (2021) for SVAR models by imposing co-kurtosis conditions.

In this article we study the problem of SVARMA identification and estimation extending the frequency domain approach of Velasco and Lobato (2018), henceforth VL, to the multivariate and structural case. VL showed that identification of a possible noncausal or noninvertible ARMA model can be achieved by the sensitivity of higher order spectral densities to the location of the roots of the lags polynomials, unlike the usual second order spectral density, attaining phase identification as noted in Lii and Rosenblatt (1982). VL also investigated parameter estimation using a minimum distance criterion between higher order periodograms and spectral densities of the ARMA model which accounts efficiently for all moment conditions of a given order at all lags.

To extend these ideas to the SVARMA setting, we first show that an identifying frequency domain criterion based on higher order spectral densities arrays can indeed discriminate linear processes which have the same autocovariance function but whose different IRFs are reflected on their higher order dynamics. We are able to reproduce the previous dynamic and static identification results for non-Gaussian vector models assuming only ICA and serial independence up to a given order (third and/or fourth) and to provide some extensions when some nonzero (i.e., non-Gaussian) cumulant condition is violated or when no version of ICA holds but we impose a rank condition on the innovations third order cumulant array. These results rely on a simple non singularity condition on the transfer function of the linear system so that our criterion can evaluate all versions of the model up to a Blaschke factor and the value of higher order cumulants of structural errors.

This global identification provides a constructive method for designing minimum distance parameter estimates, exploiting efficiently all information contained in the dynamics of moments of order 2, 3, and 4 without distributional assumptions or factorizations of the matrix lag polynomials to deal with the simultaneous presence of roots inside and outside the unit circle needed when computing residuals and likelihood functions as in, for example, GMR. Despite SVARMA identification up to a signed permutation is enough for IRF and prediction error variance analyses (given that a particular labeling can be attached to each shock), to obtain standard asymptotic results for our parameter estimates we fix a unique identified version of the model using a particular ordering and sign structure on the innovations. These restrictions could be replaced by alternative statistical conditions or economic information, which then would become overidentification restrictions that could be tested in our framework. We also develop efficient estimates exploiting jointly all moment conditions available and bootstrap approximations for the distribution of estimates.

The rest of the article is organized as follows. Section 2 sets the identification problem and Section 3 provides the basic results. Sections 4 and 5 deal with parameter identification and minimum distance estimation. Section 6 analyzes GMM efficient estimates and bootstrap approximations. Section 7 presents the numerical methods and a simulation experiment. Section 8 reanalyzes Blanchard and Quah (1989) identification of a bivariate system for U.S. GNP growth and unemployment. A series of appendices in the supplemental materials include additional discussion and simulations, together with all proofs.

2. Identification Problem and Assumptions

We consider the SVARMA\((p, q)\) system

\[
\Phi(L) Y_t = \mu + \Theta(L) \varepsilon_t,
\]

where the \(d\)-vector \(\varepsilon_t\) behaves as an iid sequence up to a finite number \(k\) of moments, \(k \geq 3\), with zero mean and covariance matrix \(I_d\), the \(d\)-dimensional identity matrix, but with components not necessarily mutually independent. The vector \(\mu\) is an unknown level parameter and the lag polynomials with matrix coefficients \(\Phi(L) = I_p - \Phi_1 L - \cdots - \Phi_p L^p\) and \(\Theta(L) = \Theta_0 + \Theta_1 L + \cdots + \Theta_q L^q\), \(\Theta_0\) non singular, satisfy \(\det(\Phi(z)) \neq 0\) for \(|z| = 1\). These conditions guarantee the existence of a stationary solution for \(Y_t\). Note that we allow the roots of the determinants of \(\Theta(z)\) or \(\Phi(z)\) to be inside or outside the unit circle so that the expansions of the transfer function \(\Psi(z) := \Phi^{-1}(z) \Theta(z)\) and its inverse \(\Psi^{-1}(z)\) could include powers of \(z\) and \(z^{-1}\), accounting for noncausal or non-invertible systems.

To investigate the identification problems on the location of the roots of the matrix polynomials \(\Phi(z)\) and \(\Theta(z)\) and on the components of \(\varepsilon_t\) determined by \(\Theta_0\), we use the device of the generalized orthogonal Blaschke matrices (BM). Following Lippi and Reichlin (1994), and denoting by \(*\) simultaneous transposition and complex conjugation, a \(d \times d\) matrix \(A(z)\) is a BM if

1. \(A(z)\) has no poles of modulus smaller or equal to unity and
2. \(A(z)^{-1} = A^*(z^{-1})\), that is, \(A(z)A^*(z^{-1}) = I_d\).
Further, for any BM, there exists an integer $r$ and complex $a_j$, $j = 1, \ldots, r$, such that

$$\mathbf{A}(z) = K_0 R(a_1, z) K_1 R(a_2, z) K_2 \cdots K_{r-1} R(a_r, z) K_r,$$

where $|a_j| < 1$, $K_i$ are orthogonal matrices, $K_i K'_i = I_d$, and

$$R(a, z) = \begin{pmatrix} g_a(z) & 0 \\ 0 & I_{d-1} \end{pmatrix}, \quad g_a(z) = \frac{z-a}{1-a^* z},$$

see also the discussion in Hannan (1970, pp. 65–67). For any BM $\mathbf{A}(z)$, we can write

$$Y_t = \Psi(L) \mathbf{A}(L) u_t,$$

(2)

where $u_t = \mathbf{A}(L)^{-1} \varepsilon_t$ is a serially uncorrelated process because its spectral density matrix is constant, $f_{\lambda}(\lambda) = (2\pi)^{-1} A^{-1}(e^{-i\lambda}) A^{-\ast}(e^{i\lambda}) = (2\pi)^{-1} I_d$, though not independent. Then, the spectral density of $Y_t$ implied by the representation (2) for any BM $\mathbf{A}(z)$ and any $u_t$ is always the same,

$$f(\lambda) = \frac{1}{2\pi} \Psi(e^{-i\lambda}) \Psi^*(e^{i\lambda}).$$

The same conclusion arises if $\mathbf{A}$ is the inverse of a BM, and in particular when in representation (1) it holds that $1/a^*$ equals an actual root of $\det(\Psi(z))$, irrespective of being inside or outside the complex unit circle, in a process of flipping the roots of $\det(\Psi(z))$. These facts imply at once that second order information cannot identify the location of these roots with respect to the unit circle, and, even with knowledge of $p$ and $q$, there are infinite VARMA representations with the same second order properties (i.e., $f(\lambda)$), but different IRF $\Psi(L) \mathbf{A}(L)$ and white noise error sequence $u_t = \mathbf{A}(L)^{-1} \varepsilon_t$. These alternative IRFs and errors are associated to invertible/noninvertible and causal/noncausal representations of a model when the roots of $A^{-1}(L)$ match those of $\Theta(L)$ or the roots of $\mathbf{A}(L)$ match those of $\Phi(L)$, respectively, and $\Psi(L) \mathbf{A}(L) = \Phi^{-1}(L) \Theta(L) \mathbf{A}(L)$ is still the IRF of a VARMA($p$, $q$) process with the same $f(\lambda)$. Traditional estimation methods based on Gaussian PML, like Whittle approximation, only consider causal and invertible representations, but still have to deal with the static problem that arises for $\mathbf{A}(L)$ constant.

The static identification problem refers to the well-known lack of identification of standard Structural VAR(MA) models with respect to orthogonal rotations $u_t = K \varepsilon_t$ of the structural errors in absence of further identifying assumptions on the IRF provided by economic theory and/or further model structure (see e.g., Rubio-Ramirez, Waggoner, and Zha (2010) for equality restrictions, and Granziera, Moon, and Schorfheide (2018) for sign restrictions). However, it is possible to consider this static problem within the same framework by allowing constant BM equal to an orthogonal matrix.

To consider all these situations when trying to identify a SVARMA model we extend the BM concept to any matrix $\mathbf{A}(z)$ that satisfies the orthogonality condition 2. and $\mathbf{1}^*$. $\mathbf{A}(z)$ has no poles of modulus equal to unity, but could have some with modulus larger or smaller than unity.

Then, in the representation (1) for a BM $\mathbf{A}(z)$ we allow for $|a_1| > 1$ as well as $|a_j| < 1$, so that there exists an $\eta > 0$ such that $\min_{j=1,\ldots,r} |a_j| - 1 \leq \eta > 0$, where the case $\eta = 0$ is interpreted as $\mathbf{A}(z) = K$ being a constant (in $z$) orthogonal matrix. If $1/a^*_j$ are not restricted to match the roots of $\det\Psi(z)$, both basic and non basic representations of VARMA models can be obtained in the sense of Lippi and Reichlin (1994).

To solve the problem that second order dynamics cannot identify the phase of $\Psi$, we resort to higher order moments and non-Gaussianity as proposed by Lii and Rosenblatt (1982). The $k$th order cumulants of $\varepsilon_t$ can be characterized by the $d^2 \times d^{k-2}$ matrix

$$\mathbf{v}_{k} \Psi^0 := [\mathbf{v}(\kappa_1\cdots\kappa) \mathbf{v}(\kappa_2\cdots\kappa) \cdots \mathbf{v}(\kappa_d\cdots\kappa)],$$

where $\kappa_j\cdots\kappa_j$ is the $d \times d$ matrix with typical $(j, (j, 3, \ldots, k))$-element equal to the $k$th order joint cumulant $\mathrm{cum}(\varepsilon(1)\varepsilon(2), \varepsilon(3)\varepsilon(k))$, $j \in \{1, \ldots, d\}$, see Chan and Ho (2004) and Jammalamadaka, Rao, and Terdik (2006). Then, we find the following compact representation of the spectral density $f_{\lambda,k}$ of $(Y_1, a(1), \ldots, Y_1, a(k))$ for any order $k = 2, 3, \ldots$, and $a = (a(1), \ldots, a(k))$ selecting $k$ elements of $Y_t = \Psi(L) \varepsilon_t$ following a linear model with IRF $\Psi(L)$ and iid innovations $\varepsilon_t$ up to moments of order $k$,

$$f_{\lambda,k}(\lambda) = \frac{1}{(2\pi)^{k-1}} \Psi_{\lambda}^0 (\lambda) \mathbf{v}_{k} \Psi^0 \mathbf{v},$$

see Appendix A in the supplementary materials, where for $\lambda = (\lambda_1, \ldots, \lambda_{k-1})$ we define

$$\Psi_{\lambda} \Psi^0 := [\Psi_{\lambda}(\kappa) (e^{\lambda_1 + \cdots + \kappa_{k-1}}) \otimes \Psi_{\lambda}(\kappa) (e^{-i\kappa_{k-1}}) \otimes \cdots \otimes \Psi_{\lambda}(\kappa) (e^{-i\kappa_{k-1}}) \otimes \Psi_{\lambda}(\kappa) (e^{-i\kappa_{k-1}})],$$

for the usual Kronecker product $\otimes$ on the rows $\Psi_{\lambda}(\kappa)$ of $\Psi$. This representation produces the usual spectral density for $k = 2$, $f_{2}(\lambda) = f(\lambda)$, because for any pair $a(1), a(2)$

$$f_{\lambda,1,2}(\lambda) = \frac{1}{2\pi} \Psi_{\lambda}(\kappa) (e^{\lambda_1 + \kappa}) \otimes \Psi_{\lambda}(\kappa) (e^{-i\kappa_1}) \mathbf{v}_{k} \Psi^0$$

as $\mathbf{v}_{k} \Psi^0 = \mathbf{v}(\varepsilon_t \varepsilon'_t)$ is the usual vector $\mathbf{v}(\varepsilon_t \varepsilon'_t)$ under the imposed normalization.

We now discuss the intuition on why higher order spectral densities with $k \geq 3$ can achieve dynamic identification unlike for $k = 2$. Thus, for given components $a = (a(1), \ldots, a(k))$ of $Y_t$ and any nonconstant BM $\mathbf{A}(z)$,

$$f_{\lambda,k}(\lambda; \mathbf{A}, \mathbf{v}_{k}) = \frac{1}{(2\pi)^{k-1}} \Psi_{\lambda}^0 (\lambda) \mathbf{A}^k (\lambda) \mathbf{v}(\mathbf{v}_{k})$$

is the implied $k$th spectral density for any $k$th order marginal cumulants matrix $\mathbf{v}_{k}$ under the (wrong) assumption that $u_t = \mathbf{A}(L)^{-1} \varepsilon_t$ is an iid $\mathbf{0}(d)$ sequence in (2) (as the true $\varepsilon_t$ recovered for $\mathbf{A}(L) = I_d$ and not just serially uncorrelated).

Under some identification (rank) assumptions, taking $k = 3$ for instance, $f_{a,3}(\lambda; \mathbf{A}, \mathbf{v}_{3})$ does differ from the true density $f_{a,3}(\lambda) = f_{a,3}(\lambda; I_d, \mathbf{v}_{3})$ for all choices of $\mathbf{v}_{3}$ because

---

1 Notice that in Theorem 1 of Lippi and Reichlin (1994) it is fixed that $K_0 = I_d$, but $K_0$ may need to be different from identity.
In this section we discuss general identification results for linear models based on the spectral loss functions $L^0_k$ under the assumption of known dynamics up to a BM $\mathbf{A}$ and the corresponding cumulants of structural shocks. We do not parametrize $\Psi$ or impose a VARMA structure following a nonparametric or coordinate-free approach in the terminology of Hannan and Deistler (1988) and consider, first, dynamic identification using nonconstant BM $\mathbf{A}$, and, second, static components identification using constant $\mathbf{A}$.

**Theorem 1.** Under Assumptions 1(k), 2(3), and 4, for any nonconstant BM $\mathbf{A}(z)$, there exists an $\epsilon > 0$ such that,

$$\inf_{\mathbf{v}\kappa_3} L^0_k(\mathbf{A}, \mathbf{v}\kappa_3) \geq \epsilon > 0.$$
and D. Theorem 1 shows that under a rank condition on \(\text{rank} v_k^0\) it is not possible to choose \(v_k^3\) such that third order dynamics can be replicated for any nonconstant BM \(A(z)\) with an infinite expansion in positive and/or negative powers of \(z\), possibly inverting some roots of \(\Psi\). Further, unlike in GMR or Theorem 4 of Chan and Ho (2004), there is no need to assume bounded moments of order 4.

**Theorem 2.** Under Assumptions 1(k), 3(k), \(k = 3\) or 4, and 4, for any nonconstant BM \(A(z)\), there exists an \(\epsilon > 0\) such that for \(\text{rank} v_k^0(\alpha)\) defined in (3) and (4) for marginal cumulants \(\alpha\),
\[
\inf_{\alpha} L_{\text{k}}(A, v_k^0(\alpha)) \geq \epsilon > 0.
\]

Theorem 2 relies on the ICA particular structure of \(v_k^0(\alpha)\) imposed by Assumption 3(k). In fact, for \(k = 3\), this is just a particular case of Theorem 1, since \(\text{rank} v_k^0(\alpha) = d\) under Assumption 3(k), \(k = 3\), 4. However, the argument of Theorem 1 cannot be directly extended under a generic rank condition on \(v_k^0\) to show identification by \(L_{\text{k}}^0\).

**Theorem 3.** Under Assumption 1(k), 3(k), \(k = 3\) or 4, and Assumption 4, for any constant BM \(K\) different from a SPM, that is, \(K \neq P_d\), there exists an \(\epsilon > 0\) such that
\[
\inf_{\alpha} L_{\text{k}}(K, v_k^0(\alpha)) \geq \epsilon > 0.
\]

Under marginal independence of order \(k\), which implies all the moment conditions of GMR, including the co-kurtosis imposed by Assumption 3, it is not possible to choose \(v_k^3\) and \(D\). Theorem 1 shows that under a rank condition on \(v_k^0\) provided in Theorem 1, since rank \(v_k^0(\alpha) = d\) under Assumption 3(k), \(k = 3\), 4. However, the argument of Theorem 1 cannot be directly extended under a generic rank condition on \(v_k^0\) to show identification by \(L_{\text{k}}^0\).

**Corollary 1.** Under Assumptions 1(k), 3*(k), \(k = 3\) or 4, and 4, Theorem 3 holds.

Thus, static identification is preserved if a single higher order cumulant of a given order \(k\) is zero, because all BM but SPM are discarded, but this is not true if more than one marginal element of \(\alpha_k^0\) is zero as described by Comon (1994). The next result investigates the case when possibly more than one marginal cumulant of the same order \(k = 3\) or \(k = 4\) is zero, but the corresponding marginal cumulants of the other order are nonzero, allowing for at most one component to have simultaneously zero skewness and zero kurtosis for static identification. As before, for dynamic identification all components have to be non-Gaussian up to order four, showing explicitly in a joint loss function with both third and fourth cumulants the moment conditions tradeoff which is only implicit in GMR’s set up.

**Corollary 2.** Under Assumptions 1(k), 3(k), \(k = 3\) or 4, and 4, with index sets \(I_3\) and \(I_4\), subsets of \(\{1, 2, \ldots, d\}\), such that \(\alpha_k^0 = 0\) for \(j \in I_k\), and \#\((I_3 \cap I_4)\) \(\leq 1\), for any constant BM \(A\) different from a SPM, that is, \(A \neq P_d\), there exists an \(\epsilon > 0\) such that
\[
\inf_{\alpha} L_{\text{k}}^0(\alpha_k^0(\alpha_3)) + \inf_{\alpha} L_{\text{k}}^0(\alpha_k^0(\alpha_4)) \geq \epsilon > 0,
\]
while if \#\((I_3 \cap I_4)\) = 0, then (5) also holds for all nonconstant BM \(A\).

### 4. Parameter Identification

In this section we assume that the observed \(d\)-dimensional SVARMA(\(p, q\)) process \(Y_t\) admits the following parameterization
\[
\Phi_{\theta_0}(L) Y_t = \mu + \Theta_{\theta_0}(L) \varepsilon_t, \quad \varepsilon_t \sim iid_{\mathcal{K}} \left(0, A_d, v_k^0(\alpha_k^0), k \in \mathcal{K}\right),
\]
where the index set \(\mathcal{K} \subseteq \{3, 4\}\) is non empty, the lag polynomials \(\Phi_{\theta}(L) := L_1 \Phi_{\theta_1}(\theta) \cdots \Phi_{\theta_p}(\theta) L^p\) and \(\Theta_{\theta}(L) := \Theta_{\theta_0}(\theta) + \Theta_{\theta_1}(\theta) L + \cdots + \Theta_{\theta_q}(\theta) L^q\) depend on a \(m\)-dimensional parameter \(\theta \in S \subseteq \mathbb{R}^m\) and \(iid_{\mathcal{K}}\) means that Assumption 1(k) holds imposing serial iid-ness up to \(k\) moments for all \(k \in \mathcal{K}\). The parameterization \(v_k^0(\alpha_k^0)\) given in (3) and (4) imposes the independence component condition of Assumption 3(k) on the arrays of \(k \in \mathcal{K}\) order cumulants of the standardized error sequence \(\varepsilon_t\) with vectors \(\alpha_k^0 \in \mathbb{R}^d\) of marginal skewness \((k = 3)\) and kurtosis coefficients \((k = 4)\). Then \(\theta_a\) and \(\alpha_0\) denote the true value of the parameters and, since the level \(\mu\) could be estimated by OLS or GLS based on estimates of \(\theta\) and our methods are invariant to \(\mu\), we do not discuss this further.

The \(k\)th order spectral density parametric model for each index \(a = (a(1), \ldots, a(k))\) of components of \(Y_t\) with representation (6) is given for \(k = 3, 4\), by
\[
\Phi_{\theta_0}(L) Y_t = \mu + \Theta_{\theta_0}(L) \varepsilon_t, \quad \varepsilon_t \sim iid_{\mathcal{K}} \left(0, A_d, v_k^0(\alpha_k^0), k \in \mathcal{K}\right),
\]
where for any index set \(\mathcal{K} \subseteq \{3, 4\}\) is non empty, the lag polynomials \(\Phi_{\theta}(L) := L_1 \Phi_{\theta_1}(\theta) \cdots \Phi_{\theta_p}(\theta) L^p\) and \(\Theta_{\theta}(L) := \Theta_{\theta_0}(\theta) + \Theta_{\theta_1}(\theta) L + \cdots + \Theta_{\theta_q}(\theta) L^q\) depend on a \(m\)-dimensional parameter \(\theta \in S \subseteq \mathbb{R}^m\) and \(iid_{\mathcal{K}}\) means that Assumption 1(k) holds imposing serial iid-ness up to \(k\) moments for all \(k \in \mathcal{K}\). The parameterization \(v_k^0(\alpha_k^0)\) given in (3) and (4) imposes the independence component condition of Assumption 3(k) on the arrays of \(k \in \mathcal{K}\) order cumulants of the standardized error sequence \(\varepsilon_t\) with vectors \(\alpha_k^0 \in \mathbb{R}^d\) of marginal skewness \((k = 3)\) and kurtosis coefficients \((k = 4)\). Then \(\theta_a\) and \(\alpha_0\) denote the true value of the parameters and, since the level \(\mu\) could be estimated by OLS or GLS based on estimates of \(\theta\) and our methods are invariant to \(\mu\), we do not discuss this further.

The \(k\)th order spectral density parametric model for each index \(a = (a(1), \ldots, a(k))\) of components of \(Y_t\) with representation (6) is given for \(k = 3, 4\), by
\[
\Phi_{\theta_0}(L) Y_t = \mu + \Theta_{\theta_0}(L) \varepsilon_t, \quad \varepsilon_t \sim iid_{\mathcal{K}} \left(0, A_d, v_k^0(\alpha_k^0), k \in \mathcal{K}\right),
\]
where \( S_k := \left( e_1^{\otimes k}, e_2^{\otimes k}, \ldots, e_d^{\otimes k} \right) \) is a rank \( d \) selection matrix and 
\[ \Psi(\lambda; \theta) := \Phi^{-1}_\theta(e^{-\lambda I}) \Theta_{\theta}(e^{-\lambda I}) \].

For \( k = 2 \) we replace \( v_k^{IC}(\alpha) \) by \( L_k \) to impose normalization and uncorrelation in \( \epsilon_t \).

We assume that the parameterization (6) satisfies the following conditions.

**Assumption 5.**

5.1. For all \( \theta \in \mathcal{S}, \det(\Phi_\theta(z)) \neq 0 \) for \( |z| = 1 \) and \( \Theta(\theta) \) is nonsingular.

5.2. For all \( \theta \neq \theta_0, \Phi^{-1}_\theta(z) \Theta_{\theta}(z) \neq \Phi^{-1}_{\theta_0}(z) \Theta_{\theta_0}(z) \) in a subset of positive measure of \{ \( z \in \mathbb{C} : |z|=1 \) \}.

5.3. \( \theta_0 \in \mathcal{S} \) and \( \mathcal{S} \) is compact.

5.4. \( \Phi_1(\theta), i = 1, \ldots, p, \) and \( \Theta_1(\theta), i = 0, \ldots, q, \) are continuously differentiable for \( \theta \in \mathcal{S} \).

Assumption 5.1 imposes Assumption 4 for each parameterized model and, with the iid condition of order \( k \) on the sequence \( \epsilon_t \) in Assumption 1(k), guarantees that \( Y_t \) with representation (6) is \( k \)-stationary. The identifiability conditions in Assumption 5.2 are satisfied when the parameter space \( \mathcal{S} \) is sufficiently constrained (see Boubacar Mainassara and Francq 2011) as for restricted versions of causal and invertible VARMA models (e.g., echelon or final equations forms) that guarantee that \( \Phi(L) \) and \( \Theta(L) \) are left coprime and that the unique unimodular common left divisor of \( \Phi(L) \) and \( \Theta(L) \) is the identity matrix, see for example, Sec. 12.1 in Lütkepohl (2005), and for parameterizations imposing factorizations of lag polynomials in forward and backward components, for example, Lanne and Saikkonen (2013) and Funovits (2021) for AR and MA polynomials, respectively. Assumption 5.3 is a standard parameter space restriction in asymptotic analysis and, together with Assumption 5.4, allows for uniformity arguments.

Note that parameterizations covering both invertible and noninvertible (or causal and noncausal) solutions are allowed by Assumption 5.2, which identifies uniquely the parametric transfer and impulse response functions. However, it is possible that \( \Phi^{-1}_\theta(z) \Theta_{\theta}(z) \neq \Phi^{-1}_{\theta_0}(z) \Theta_{\theta_0}(z) \) for almost all \( z \) with \( |z| = 1 \), some \( \theta \neq \theta_0 \) and BM \( A(z) \neq I_d \), and therefore \( f_{a,k}(\lambda; \theta, \alpha) \) cannot identify \( \Phi^{-1}_0(z) \Theta_{\theta_0}(z) \). Further, Assumption 5.2 is not sufficient either to identify \( \Phi^{-1}_\theta(z) \Theta_{\theta}(z) \) uniquely from \( f_{a,k}(\lambda; \theta, \alpha) \) for \( k > 2 \) without further restrictions to discard that signed permutations of \( \Phi^{-1}_\theta(z) \Theta_{\theta}(z) \) by a SPM \( A(z) \neq P_d \neq I_d \) could be parameterized by some \( \theta \neq \theta_0 \), see Theorem 4. With this aim, we introduce the following assumptions, which fix the signs of the components of \( \epsilon_t \) and enforce an ordering by either imposing a given structure on \( \Theta_0(\theta_0) \) (6A) or by directly excluding any signed permutations among the columns of the transfer function (6B).

**Assumption 6A.** The diagonal elements of \( \Theta_1(\theta_0) \) are all positive and the elements of \( Y_t \) are ordered so that there is no SPM \( P_d \neq I_d \) such that the absolute value of the product of the diagonal elements of \( \Theta_0(\theta_0) P_d \) is equal or larger than that of \( \Theta_0(\theta_0) \).

The restriction on the diagonal of \( \Theta_0(\theta_0) \), see Pham and Garat (1997) and Lanne and Luoto (2021), fixes the signs of \( \epsilon_t \) so that a positive increment in a component of \( \epsilon_t \) corresponds to a positive increment in the element of \( Y_t \) associated to this shock. This condition can be described as a sign restriction (at lag 0) on the IRF of each endogenous variable with respect to the corresponding error term, giving a unique interpretation of the IRF (as it is automatically imposed when imposing a lower triangular structure in \( \Theta_0(\theta_0) \) using Cholesky identification). Further, Assumption 6A imposes that only the identity permutation \( P_d = I_d \) maximizes the absolute value of the product of the diagonal elements of \( \Theta_0(\theta_0) P_d \), but alternative ordering schemes are possible, see for example, Lanne, Meitz, and Saikkonen (2017).

**Assumption 6B.** For all \( \theta \neq \theta_0, \theta \in \mathcal{S}, \Phi^{-1}_\theta(z) \Theta_{\theta}(z) \neq \Phi^{-1}_{\theta_0}(z) \Theta_{\theta_0}(z) \) for any SPM \( P_d \) in a subset of positive measure of \{ \( z \in \mathbb{C} : |z|=1 \) \}.

Assumption 6B holds when some identifying restrictions have already been imposed on \( \Phi_\theta \) or \( \Theta_\theta \), such as a recursive system assumption with known ordering, or sign and zero restrictions on \( \Theta_0 \) such as \( \Theta_{0,1}(\theta_0) = (++, -0, 0) \) and \( \Theta_{0,2}(\theta_0) = (+, +, -) \) for \( d = 3 \).

Define the \( L^2 \) distance between \( f_{a,k}(\lambda; \theta, \alpha) \) and \( f_{a,k}(\lambda) = f_{a,k}(\lambda; \theta_0, \alpha_0) \) for any \( \alpha \neq \alpha_0 \) and all possible indices \( a \),
\[ L_k(\theta, \alpha) := \sum_a \int_{||\lambda|| < 1} \left| f_{a,k}(\lambda; \theta, \alpha) - f_{a,k}(\lambda) \right|^2 d\lambda, \]
which is continuous in the parameters \( \theta \) and \( \alpha \) and the marginal cumulants \( \alpha \) of order \( k \) by Assumption 5.4 and the definition of \( f_{a,k} \), and define \( L_2(\theta) \) in the same way in terms of \( f_{a,2}(\lambda; \theta) \), which only depends on \( \theta \) because of the normalization \( \nabla(\epsilon_t) = I_d \) under (6). We now show that the conclusions of Theorem 4 extend to parametric loss functions.

**Theorem 5.** Under Assumptions 1(k), 3(k), 5, and 6, \( k \in \mathbb{K} \), for any \( v > 0 \) there exists an \( \epsilon > 0 \) such that
\[ \theta \in \mathcal{S}, \|\theta - \theta_0\| + |\alpha - \alpha_0| \geq v \Rightarrow L_2(\theta) + L_k(\theta, \alpha) \geq \epsilon > 0. \]

This result provides identification of the dynamics and scaling parameters \( \theta \) and the marginal cumulant vector \( \alpha \) in absence of knowledge on the possible noninvertibility (or noncausality) of \( Y_t \). Following the comments to Theorem 4, and despite Assumption 5.2 allows that \( L_2(\theta) = 0 \) for some \( \theta \neq \theta_0 \) because \( \Phi^{-1}_\theta(z) \Theta_{\theta}(z) = \Phi^{-1}_{\theta_0}(z) \Theta_{\theta_0}(z) \) for some BM \( A, L_k \) for \( k > 2 \) are not minimized unless \( \mathbb{Z} = P_d \) and \( \alpha_3 = \frac{P_d^2}{P_d^3} \) or \( \alpha_4 = \frac{P_d^4}{P_d^5} \).

However, Assumption 6 imposes a unique ordering (and sign) to discard any \( \theta \) which leads to permutations of the columns of \( \Phi^{-1}_0(z) \Theta_{\theta_0}(z) \) generating the same (second and k-order) dynamics of \( Y_t \) based on a permuted version of \( \epsilon_t \).

Similarly to the univariate analysis of VBL, \( L_3, L_4 \) could be considered jointly so that both \( \alpha_3, \alpha_4 \) are identified by the ordering of the columns of the transfer function imposed by Assumption 6. But, in contrast to VBL, it is not possible now to identify jointly \( \theta \) and \( \alpha_k \) from only \( L_3 \) without \( L_2 \), because the MA matrix polynomial \( \Theta_{\theta}(z) \) incorporates the scaling as we set \( \nabla(\epsilon_t) = I_d \), so that \( \alpha_k \) are skewness and kurtosis coefficients. See a discussion on alternative parameterizations in Appendix E in the supplementary materials.
5. Parameter Minimum Distance Estimation

Given a time series of $Y_t$, $t = 1, \ldots, T$, we define sample analogs of the distance functions $L_k(\theta, \alpha)$ for $k = 3, 4$ and $L_2(\theta)$ as in Brillinger (1985) and VL,

$$L_{k,T}(\theta, \alpha) := \frac{(2\pi)^{2k-2}}{T^{k-1}} \sum_{\lambda_j} \sum_{a_j} \left| I_{a,k}(\lambda_j; \theta, \alpha) - I_{a,k}(\lambda_j) \right|^2,$$

replacing the true spectral densities by the sample (higher) order periodograms $I_{a,k}$.

$$I_{a,k}(\lambda_1, \ldots, \lambda_{k-1}) := \frac{(2\pi)^{1-k}}{T^{1-k}} \sum_{j_1} \sum_{j_2} \cdots \sum_{j_{k-1}} \left| I_{a,k}(\lambda_{j_1}, \ldots, \lambda_{j_{k-1}}, w_{T,a}(1)(\lambda_1) \cdots w_{T,a}(k-1)(\lambda_{k-1}) \right|,$$

where $w_T(\lambda) = \sum_{t=1}^{T} Y_t e^{-i\lambda t}$ is the discrete Fourier transform (DFT) of $Y_t$. In $L_{k,T}$, the summation in $\lambda_j$ is $\lambda_j = (\lambda_{j_1}, \ldots, \lambda_{j_{k-1}})$ for Fourier frequencies $\lambda_{j_n} = 2\pi j_n/T$ runs for all $j_n = 1, \ldots, T-1$, $n = 1, \ldots, k-1$, excluding $j_a + j_b = \text{mod}(T)$, $\alpha \neq b$, and $j_a + j_b + j_c = \text{mod}(T)$, all $a, b$ and $c$ different, for sample mean correction.

The $d^k$-vector containing all $I_{a,k}$ spectral density functions can be written under Assumptions 1(k) and 3(k) as $(2\pi)^{1-k} \Psi_k(\lambda; \theta)$ $S_k \alpha$, with $\Psi_k(\lambda; \theta) = \Psi_{\infty k}(\lambda; \theta)$, so that for $k = 3, 4$

$$L_{k,T}(\theta, \alpha) = \frac{1}{T^{k-1}} \sum_{\lambda_j} \left( \Psi_k(\lambda_j; \theta) S_k \alpha - \bar{I}_k(\lambda_j) \right)^* \times \left( \Psi_k(\lambda_j; \theta) S_k \alpha - \bar{I}_k(\lambda_j) \right),$$

where $\bar{I}_k(\lambda)$ is the $d^k \times 1$ vector stacking all normalized ($k = 3$) biperiodograms and ($k = 4$) triperiodograms of $Y_t$, $\bar{I}_k(\lambda) := T^{-1} w_T(\lambda) = \cdots \otimes w_T(\lambda_2) \otimes w_T(\lambda_1)$. Correspondingly, $L_{2,T}(\theta) = T^{-1} \sum_{\lambda_j} \left( \Psi_2(\lambda_j; \theta) \text{vec}(I_2) - \bar{I}_2(\lambda_j) \right)^* \left( \Psi_2(\lambda_j; \theta) \text{vec}(I_2) - \bar{I}_2(\lambda_j) \right)$.

We set the following minimum distance parameter estimator for weights $w_k$,

$$\left( \hat{\theta}_{w,T}, \hat{\alpha}_{k,T}, k \in \mathcal{K} \right) := \arg \min_{\theta \in \mathcal{S}, \alpha \in \mathcal{K}} L_{2,T}(\theta) + \sum_{k \in \mathcal{K}} w_k L_{k,T}(\theta, \alpha_k),$$

where $\min_{k \in \mathcal{K}} w_k > 0$. The main purpose of combining loss functions $L_{k,T}$ involving cumulants of different orders $k = 3, 4$, is robustness to lack of identification due to failure of the nonzero cumulant condition of Assumption 3(k) for a single $k > 2$, as efficiency gains are possible but difficult to characterize, even in the univariate case, see Lobato and Velasco (2018). Additionally, we always need to include $L_{2,T}$ in our loss function for scaling identification. Then, consistency of estimates is achieved exploiting that periodogram averages estimate consistently integrals of the true spectral densities.

Theorem 6. Under Assumptions 1(2k), 3(4), 5, 6, $k \in \mathcal{K}$, as $T \to \infty$,

$$\left( \hat{\theta}_{w,T}, \hat{\alpha}_{k,T}, k \in \mathcal{K} \right) \overset{p}{\to} (\theta_0, \alpha_0^0, k \in \mathcal{K}).$$

Consistent estimate of $\hat{\alpha}_k$ can be pursued for $k \notin \mathcal{K}$ using $\hat{\theta}_{w,T}$ when Assumptions 1(2k) and 3(4) hold for both $k = 3, 4$, but only one set of cumulants needs to satisfy the nonzero conditions that guarantee identification. Further, independence and equal distribution of order $2k$ in Assumption 1 are used to facilitate the asymptotic analysis despite are not necessary for identification.

Asymptotic Distribution. Now consider optimal weighting of higher order periodograms replacing $L_{k,T}(\theta, \alpha)$ by a weighted loss function

$$L_{k,T}^{\text{EFF}}(\theta, \alpha) := \frac{1}{T^{k-1}} \sum_{\lambda_j} \left( \Psi_k(\lambda_j; \theta) S_k \alpha - \bar{I}_k(\lambda_j) \right)^* \times \left( \Psi_k(\lambda_j; \theta) S_k \alpha - \bar{I}_k(\lambda_j) \right),$$

where $\lambda_j$ are the frequencies introduced at the beginning of the section and $\Psi_k(\lambda; \theta) = \Psi_{\infty k}(\lambda; \theta)$, evaluated at a preliminary estimate $\hat{\theta}_T \rightarrow \theta_0$. This weighting is not affected by structural identification when $W_k$ does not change when $\Psi(e^{-i\lambda}; \theta_T)$ is replaced by $\Psi(e^{-i\lambda}; \theta_T) A(e^{-i\lambda})$ for a BM $A(z)$.

Further, exploiting that $S_k \Psi_k(\lambda; \theta_0) W_k(\lambda; \theta_0) \Psi_k(\lambda; \theta_0) S_k = S_k S_k = I_d$, we replace $L_{k,T}^{\text{EFF}}$ by the pseudo-profile distance

$$\hat{\alpha}_{k,T}(\theta) := L_{k,T}^{\text{EFF}}(\theta, \hat{\alpha}_{k,T}(\theta)) \text{ using the estimate}$$

$$\hat{\alpha}_{k,T}(\theta) := \frac{1}{T^{k-1}} \sum_{\lambda_j} \text{Re} \left\{ S_k \Psi_k^{-1}(\lambda_j; \theta) \bar{I}_k(\lambda_j) \right\},$$

instead of the solution from the first-order condition for $\alpha$, similar to the proposal in VL. Then we focus on parameter estimates

$$\hat{\theta}_{w,T} := \arg \min_{\theta \in \mathcal{S}} \hat{L}_{w,T}(\theta),$$

$$\hat{L}_{w,T}(\theta) := L_{2,T}(\theta) + \sum_{k \in \mathcal{K}} w_k \hat{L}_{k,T}(\theta),$$

(7)

whose consistency can be deduced by the same arguments as for $\hat{\theta}_{w,T}$, and being asymptotically equivalent to estimates based on minimizing $L_{k,T}^{\text{EFF}}(\theta, \alpha)$, their analysis is simpler.

To investigate the asymptotic distribution of parameter estimates we need further restrictions on the parameterization and a local identification condition. Define

$$H_k(\theta) := (2\pi)^{1-k} \int_{\Phi_k} \text{Re} \left\{ B_k(\lambda; \theta) B_k(\lambda; \theta) \right\} d\lambda$$
where for $k = 3, 4$ we set

$$B_k (\lambda ; \theta) := \Psi_k^{-1} (\lambda ; \theta) \tilde{\Psi}_k (\lambda ; \theta) - S_k S_k^* \tilde{A}_k (\theta) := \sum_{j=1}^{k} B_{kj} (\lambda ; \theta),$$

with $\tilde{A}_k (\theta) := (2\pi)^{1-k} \int_{\Pi_{k-1}}^{\Re} \left\{ \Psi_k^{-1} (\lambda ; \theta) \tilde{\Psi}_k (\lambda ; \theta) \right\} d\lambda,$ and $B_2 (\lambda ; \theta) := \Psi_2^{-1} (\lambda ; \theta) \tilde{\Psi}_2 (\lambda ; \theta) = B_{2,1} (\lambda ; \theta) + B_{2,2} (-\lambda ; \theta),$ with

$$\Psi_k (\lambda ; \theta) := \left( \tilde{\Psi}_k^{(1)} (\lambda ; \theta), \ldots, \tilde{\Psi}_k^{(m)} (\lambda ; \theta) \right),$$

$$\Psi_{k}^{(c)} (\lambda ; \theta) := \frac{\partial}{\partial \theta^c} \Psi_k (\lambda ; \theta).$$

The $B_{kj}, j = 1, \ldots, k,$ are obtained at once by taking derivatives of the $k$-fold multiplicative structure of $\Psi_k.$ Introduce for $\alpha = \{ \alpha_k \}_{k \in \mathbb{K}}$

$$\Sigma (\theta, \alpha) := \sum_{k \in \mathbb{K} \cup \{ 2 \}} w_k (I_m \otimes S_k \alpha_k)^* H_k (\theta) (I_m \otimes S_k \alpha_k),$$

where $w_2 = 1$ and $S_k \alpha_k = \text{vec} (I_d),$ and the following assumption which imposes a rank condition on $\Sigma (\theta_0, \alpha_0)$ and reinforces the smoothness conditions of the parameterization.

**Assumption 7.**

1. Let $\Phi_k (\theta), i = 1, \ldots, p,$ and $\Theta_k (\theta), i = 0, \ldots, q,$ have continuous third order derivatives for all $\theta \in S.$
2. $\theta_0 \in \text{Int} (S).$
3. $\Sigma (\theta_0, \alpha_0) > 0.$

Assumptions 7.1 and 7.2 are standard for the analysis of asymptotic properties of estimators and, while sufficient conditions for the local identification Assumption 7.3 are the full rank of individual Hessian matrices, $H_k (\theta_0) > 0,$ for at least one $k = 2, 3, 4$ with $w_k > 0$ and Assumption 3(k) for $k \in \mathbb{K}.$ $H_2$ is equivalent to the Hessian of PMLE estimates under Gaussianity (and causality and invertibility), like Whittle estimates, which only use second order information, noting that here $H_2$ includes also the scaling parameters in $\Theta_0.$ This indicates that, while usual methods are sufficient for local identification and local asymptotic inference, non-Gaussian information is key to achieve global identification and potential efficiency improvements. Further, the centering terms $\tilde{A}_k (\lambda ; \theta)$ in $B_k (\lambda ; \theta)$ reflect the higher order cumulant estimation effect, with the elements corresponding to non scaling parameters in $\theta$ being zero for causal and invertible processes, as in VL for $d = 1.$

To simplify the asymptotic variance of $\hat{\theta}_{w,T}^\dagger$ we could strengthen Assumption 3(k), $k \in \mathbb{K},$ to ICA of order $k = 8,$ as we do in Appendix D in the supplementary materials to obtain explicit formulas for the variance of (centered) vectors of powers $e_t^{[h]}$ of order $h = 2, 3, 4,$ of $e_t$ and $e_r,$ $t \neq r.$ However, there is no need of nonzero assumptions on marginal cumulants for $k \notin \mathbb{K}.$

Define $C_k (0) := (2\pi)^{1-k} \int_{\Pi_{k-1}} B_k^* (\lambda ; \theta_0) d\lambda$ for $k = 2, 3, 4,$ the row block matrices

$$C_k (j) := \left\{ (2\pi)^{1-k} \int_{\Pi_{k-1}} B_{k,a}^* (\lambda ; \theta_0) e^{-j\lambda a} d\lambda \right\}_{a=1, \ldots, k},$$

$$j = \pm 1, \pm 2, \ldots,$$

and the block matrix

$$\Omega (\theta_0 ; C) := \begin{cases} \Omega_{ab} (\theta_0 ; C) \end{cases}_{a,b=2,3,4},$$

where $\Omega_{ab} (\theta_0 ; C) := \Gamma_{ab}^0 (\theta_0 ; C) + \Gamma_{ab}^0 (\theta_0 ; C) + \Gamma_{ab}^- (\theta_0 ; C),$ with

$$\Gamma_{ab}^\pm (\theta_0 ; C) := \sum_{j=-\infty, \neq \theta_0}^{\infty} C_a (\theta) C_b^* (\theta) \left( e_t^{[a]} - e_r^{[b]} \right) C_b^* (\theta) \left( e_t^{[a]} - e_r^{[b]} \right) C_b^* (\theta).$$

Note that variance-covariance matrices do not depend on $t \neq r$ by stationarity. These definitions provide multivariate generalizations of the score variance expressions developed in VL accounting for possibly nonfundamental processes and allow a compact presentation for all $w$ of the asymptotic distribution of $\hat{\theta}_{w,T}^\dagger$ in next theorem. See Appendix D in the supplementary materials for explicit expressions.

**Theorem 7.** Under Assumptions 1(2k), 3(k), 5, 6, 7, $k \in \mathbb{K},$ as $T \to \infty,$

$$\sqrt{T} \left( \hat{\theta}_{w,T}^\dagger - \theta_0 \right) \to_d N_{m,0} (0, \Sigma^{-1} (\theta_0, \alpha_0) \delta (\alpha_0)),

\quad \times \Omega (\theta_0 ; B) \delta' (\alpha_0) \Sigma^{-1} (\theta_0, \alpha_0),$$

where $\delta (\alpha_0) := \left[ \left( I_m \otimes \text{vec} (I_d) \right)' \right. \left| w_3 (I_m \otimes S_3 \alpha_3^0) \right)' \right| w_4 (I_m \otimes S_4 \alpha_4^0)' \right]'.

As in VL, the terms $\Gamma_{ab}^\pm$ are only different from zero for noncausal or noninvertible models while $\Gamma_{22}^0$ incorporates the scaling estimation effect, which is treated separately in other parameterizations, as in the univariate model of VL and in the one discussed in Appendix E in the supplementary materials. In the same line, $\Gamma_{10}^0, a, b > 2,$ account for the estimation of higher order cumulants under the ICA of Assumption 3(k), see Appendix D in the supplementary materials for details.

We finally show the asymptotic normality of $\hat{\theta}_{w,T}^\dagger$ for a particular $k \in \{ 3, 4 \}$ without assumptions on $\alpha_0^k$ which could contain many zeros, obtaining an easy test of overidentification. The only requisite is that $\hat{\theta}_{w,T}^\dagger$ is consistent and asymptotic normal with identification possibly provided by a different set of cumulants.

**Theorem 8.** Under Assumptions 1(2h), 3(h), 5, 6, 7, $h \in \mathbb{K} \cup \{ k \},$ as $T \to \infty,$ then $T^{1/2} \left( \hat{\theta}_{w,T}^\dagger - \alpha_0^k \right) \to_d N_{d,(0, V_k)}$ where $V_{k, k}, k = 3, 4,$ are given in Appendix B in the supplementary materials.

## 6. Efficient GMM Estimation

In this section we propose estimators exploiting efficiently all the information used by the minimum distance estimates of the previous section by minimizing simultaneously the score functions of $\mathbb{L}_{k,T} (\theta, \alpha)$ for all $k = 2, 3, 4$ without need to specify
w. Denote the gradient vector of the concentrated loss functions for all $k = 2, 3, 4$ by

$$
\mathcal{G}_T(\theta) := \left( \frac{\partial}{\partial \theta^j} \mathcal{L}_2(T, \theta), \frac{\partial}{\partial \theta^j} \mathcal{L}^*_{3,1}(\theta), \frac{\partial}{\partial \theta^j} \mathcal{L}^*_{4,1}(\theta) \right),
$$

and denote the asymptotic variance of $T^{1/2} \mathcal{G}_T(\theta_0)$ by $\mathcal{V}$ depending on $\delta$ and $\Omega$ from Theorem 7, so we can consider the objective function

$$
Q_T(\theta) = \mathcal{G}_T(\theta)^T \mathcal{V}^{-1} \mathcal{G}_T(\theta),
$$

(8)
to optimally weight the information on $\theta_0$ contained in $\mathcal{G}_T(\theta)$. Here $\mathcal{V}$ is a consistent estimator of the matrix $\mathcal{V}$, the Moore-Penrose inverse of $\mathcal{V}$, to account for cases where the asymptotic variance of $\mathcal{G}_T(\theta)$ is default rank because Assumption 3(k) fails for some $k = 3, 4$. Hence, the proposed efficient estimator of $\theta_0$ is a Newton–Raphson step using (8),

$$
\hat{\theta}_{GMM,T} := \tilde{\theta}_T - \left( \mathcal{H}_T(\hat{\theta}_T)^T \mathcal{V}_T \mathcal{H}_T(\hat{\theta}_T) \right)^{-1} \mathcal{H}_T(\hat{\theta}_T)^T \mathcal{V}_T \mathcal{G}_T(\theta),
$$

(9)
where $\mathcal{H}_T(\theta) := (\partial/\partial \theta^j) \mathcal{G}_T(\theta)$, and the initial estimate $\hat{\theta}_T$ satisfying

$$
\mathcal{H}_T(\theta_0) = O_p \left( T^{-1/2} \right)
$$

could be $\hat{\theta}^*_{w,T}$ or any other PML or GMM estimate obtained under appropriate identifying conditions. Given Theorem 7 and (10), the consistency of $\hat{\theta}_{GMM,T}$ is trivial and the next theorem states its asymptotic distribution defining the following column block matrix,

$$
\mathcal{H} := p \lim_{T \to \infty} \mathcal{H}_T(\theta_0)
\quad = \left\{ \left( I_m \otimes S_k \alpha^0_k \right) \mathcal{H}_k(\theta_0) \left( I_m \otimes S_k \alpha^0_k \right) \right\}_{k=2,3,4}.
$$

Theorem 9. Under Assumptions 1(2k), 3(5k), $k \in \{3, 4\}$, 5, 7, (10) and $\mathcal{V}_T \to_p \mathcal{V}^{-1}$ as $T \to \infty$, $\sqrt{T}(\hat{\theta}_{GMM,T} - \theta_0) \to_d \mathcal{N}_m(0, \mathcal{V}^{-1})$.

Note that Theorem 9 holds irrespective of zero values in $\alpha^0_k$ for $k = 3$ and 4, so no need of this part of Assumption 3(k) for both $k = 3, 4$ as far as (10) holds, and, though estimation procedures to obtain candidates for $\hat{\theta}_T$ can rely on some form of non-Gaussianity, they might use only second-order information complemented by economic identification restrictions. For the same reason, Assumption 6 is not needed under (10).

To perform inference we propose a simple parametric bootstrap based on resampling from the empirical distribution of model residuals. To avoid costly re-estimation of large high dimensional models, we can alternatively resample the following linearization of the estimates in terms of the higher order periodograms of $\varepsilon_t$,

$$
\hat{\theta}^*_{w,T} - \theta_0 = \sum_{k \in K \cup \{2\}} \frac{w_k}{T_k^{-1}} \left( I_m \otimes S_k \alpha^0_k \right) \mathcal{B}_k \left( \lambda_j; \theta_0 \right) \mathcal{B}_k \left( \lambda_j; \theta_0 \right) + o_p \left( T^{-1/2} \right),
$$

where for $k = 2$ we replace $\mathcal{B}_k^* \left( \lambda_j; \theta_0 \right)$ by $\mathcal{B}_k^* \left( \lambda_j; \theta_0 \right)$—vec($I_2$). Resampled versions of the estimates are obtained replacing $(\theta_0, \alpha^0_k, k \in K)$ by $(\hat{\theta}^*_{w,T}, \hat{\alpha}^*_k, k \in K)$ and $\mathcal{B}_k^*$ by the periodograms $\mathcal{B}_k^*$ of resampled $\hat{\varepsilon}_t^*$ from the empirical distribution of residuals $\hat{\varepsilon}_t = e_t(\hat{\theta}^*_{w,T})$, $t = 1, \ldots, T$, properly standardized,

$$
\hat{\theta}^*_{w,T} := \hat{\theta}^*_{w,T} + \sum_{k \in K \cup \{2\}} \frac{w_k}{T_k^{-1}} \left( I_m \otimes S_k \alpha^0_k \right) \mathcal{B}_k \left( \lambda_j; \theta_0 \right) \mathcal{B}_k \left( \lambda_j; \theta_0 \right) + o_p \left( T^{-1/2} \right).
$$

This requires only one computation of $\mathcal{B}_k(\lambda_j; \theta_0)$ and $\mathcal{B}_k(\lambda_j; \theta_0)$, but not data simulation or parameter re-estimation. Interestingly, residuals $\hat{\varepsilon}_t$ are obtained directly in the frequency domain after inversion of the residual DFT $\Psi(\lambda; \hat{\theta}^*_{w,T})^{-1}$ without need to care about noninvertible or noncausal roots or rely on filter factorizations.

Similarly, bootstrap versions of $\hat{\theta}_{GMM,T}$ and $\hat{\alpha}^*_k$ can be obtained using the same periodograms of resampled $\hat{\varepsilon}_t^*$ to construct bootstrap significance tests for parameters and over-identification tests based on redundant nonzero cumulants or restrictions on the IRF, which could be justified using an iid property on $\varepsilon_t$ with sufficient bounded moments.

7. Numerical Methods and Simulations
In this section we consider several Monte Carlo experiments to check the finite sample performance of the following identification and estimation procedure based on the results of previous sections. To identify both the correct rotation of the innovations and the location of the AR and MA lag polynomials roots we implement the following algorithm and estimates:

1. $\hat{\theta}_{2,T}$: Causal and invertible reduced form VARMA estimation:
   1.1 Whittle initial estimation, using preliminary causal IV estimation of VAR parameters if $p > 0$.
   1.2 Standardization and rotation of residuals to impose ICA and estimate $\Theta_0$.

2. $\hat{\theta}^*_{w,T}$: Minimum distance structural higher order spectral estimation:
   2.1 Computation of the $2^{(d+i+p)}$ VARMA basic representations of $\hat{\theta}_{2,T}$ obtained by combinations of possibly inverted AR and MA roots.
   2.2 Rotation of the residuals of each representation to match ICA.
   2.3 Minimization of $\mathcal{L}'_{w,T}(\theta)$ in (7) with $w \in \{1, 0 \}, \{1 \}$ using the rotated parameters from 2.2 of each specific root-model configuration as initial estimates.
   2.4 Global minimum: choose the root configuration that minimizes $\mathcal{L}'_{w,T}(\theta)$.
   2.5 Permutation of components: choose the signed permutation of error components to match Assumption 6A among all $2^d d!$ possible ones.
3. $\hat{\Theta}_{GMM,T}$: Local GMM estimation based on preliminary estimates $\hat{\theta}_{w,T}$ from Step 2 and bootstrap estimated variance of the score.

For Step 2.1 we use the explicit construction of the minimal spectral factors of rational spectral densities which invert the roots of lag polynomials in Baglio and Ferrante (2019). We also tried to minimize a penalized version of $L_{4,T}$ forcing each specific root location configuration to obtain initial estimates, with similar results in most situations. For Steps 1.2 and 2.2 we use the reconstruction ICA algorithm implemented in MATLAB RICA function. All $L_{4,T}(\theta)$ functions involved in the different steps use optimal weighting $W_k$ based on the causal and invertible Whittle estimates obtained in Step 1 with ICA rotation of the Choleski estimation of $\Theta_0$ from residuals variance (as $W_k$ is invariant to rotations of innovations and flipping of polynomial roots). We also obtain estimates of $\phi_k$ and compute for invertible models an infeasible version of the estimates $\hat{\theta}_{w,T}$ obtained in Step 1 by finding the signed permuted version that minimizes the $L^2$ distance to the true parameters.

This methodology is subject to the curse of dimensionality, first because it involves higher order dynamic moments whose description in the frequency domain involves summations in $(k - 1)$ dimension frequencies $\lambda_j$ growing with sample size $T$. And second, because our algorithm is based on a sequence of local optimizations (to identify the best representation of the lag polynomials for describing $k$-order dynamics) whose number is growing very fast with the system $(d)$ and model $(p,q)$ dimensions. Still, for moderate sample sizes and model orders, the method seems feasible and open to further refinements, but to speed up computations with many replications, local optimization uses only $k = 2, 3$ contributions, though model choice is performed evaluating also $L_{4,T}^*$ for $w = (1,1)$. Each $L_{4,T}^*$ is normalized by the number of spectral densities of order $k$ in the system, $d_k$, and by $T^{k-2}$, so that their values can be compared across $k$.

In our finite sample experiment we simulate bivariate $(d = 2)$ SVARMA$(p,q)$ systems with $(p,q) = (0,1), (1,0)$ and $(1,1)$, $\Theta(L) = (I_2 + B_1L)\Omega$ and $\Phi_1 = \begin{pmatrix} 0.9 & 0 \\ -0.4 & 0.7 \end{pmatrix}$, $B_1 = \begin{pmatrix} \zeta_1 & 0 \\ 0 & \zeta_2 \end{pmatrix}$, $\Omega = \begin{pmatrix} 10 & 4 \\ -2 & 5 \end{pmatrix}$.

Then, $\Theta_0 = \Omega$ satisfies Assumption 6A and $\Theta_1 = B_1\Omega$ with $|\zeta_j| < 1$ and $\rho_j = \pm 1$, so the two roots of $det(\Theta(z))$ are $-\zeta_j^{-p_j}, j = 1,2$. We set $\theta_0 = (\text{vec}(\Phi_1)', \text{vec}(\Omega)', \text{vec}(B_1)')'$ for $(p,q) = (1,1)$ and drop the first and last elements for $p = 0$ and $q = 0$.

We consider three sets of standardized mutually independent innovations. The first is composed by $\chi^2_{n_j}$ variates, $(n_1, n_2) = (6,1)$, the second by two $t_{n_j}$ variates, $(n_1, n_2) = (6,5)$, and the third by a mixed normal $MN(2.12, 1.41^2; -0.24, 0.58^2; 0.1, 0.9)$ and a $t_6$ variable as in GMR. These last two distributions do not satisfy Assumption 3(k) for $k = 3$ because at least one component is symmetric, so (dynamic) identification only relies on $k = 4$ order cumulants, while Assumption 1(2k) does not hold for these $t$ distributions, as they have five finite moments at most. Sample sizes are $T = 100, 200, 300$ with 1000, 500, 100 replications.

We first report the percentage of correct polynomial roots identification (i.e., location of roots inside and outside the unit circle) across simulations of estimates for two different classes of models. In Table 1 we consider all SVARMA$(p,1)$ model configurations of invertible and noninvertible MA roots for $\zeta_j = 0.5$ when $p = 0$, but only mixed cases for $p = 1$ imposing causality. We also tried the SVARMA$(0,1)$ model in GMR with the same mixed MA roots. The results for chi squared innovations indicate that skewness provides a very precise information for identification in multivariate systems as was found for univariate models in VL, while the additional use of kurtosis information does not help much for the sample sizes considered and, even for the smallest ones, can introduce further noise. For mixed normal and $t$ shocks, our identification procedure is still able to gain information on the location of roots when Assumption 3(3) fails for one component, but not much when it fails for all components as for bivariate $t$ distributions (about 50% of right locations for each of the three MA configurations). When adding information from kurtosis in these $k = 3$ identification-failure situations, the results get better in almost all cases, but with modest improvements in general given the moderate sample sizes considered and the heavy tails of the innovations. Across all setups simulated, unit circle models with MA roots both inside and outside the unit circle are more difficult to identify than purely invertible models.

The results for the SVARMA$(0,1)$ model from GMR also report significant differences for each set of innovations, with kurtosis being only relatively helpful for both symmetric and asymmetric distributions. We observe a similar pattern for SVARMA$(1,1)$ models, but identification of the MA roots is more complicated in the presence of autoregressive dynamics though improving with $T$ in almost all cases except for $t$ innovations.

Table 2 includes SVARMA$(0,1)$ and $(1,1)$ models with all AR roots configurations setting $\Phi_1 = B_1$ for $\zeta_j = 0.5$,
and a mixed MA part with $B_1 = \text{diag}(0.5, 2)$ and do not make any assumption on invertibility or causality for identification. For simple models with $q = 0$ the results are parallel to those of Table 1 with $p = 0$ (though noticeable worse for the mixed case). For VARMA(1, 1) models the rate of success of joint identification of AR and MA is much reduced than when causality is imposed (see Table 1), but now the number of root configurations actually compared is substantially larger, 9 versus 3. As expected, symmetric innovations are much more difficult to handle in this set up, especially for $q = 1$.

In Tables A1 and A2 in Appendix F in the supplementary materials we report the RMSE of estimates for the invertible and noninvertible VARMA(0, 1) models of Table 1 only for the replications which correctly identified the MA root location to make the comparisons meaningful (see Tables A3 and A4 in the supplementary materials for bias results). The behavior of parameter estimates $\hat{\theta}_{w,T}$ for VARMA(0, 1) invertible models in Table A1 in the supplementary materials compares well in terms of bias and RMSE with the unfeasible estimate $\hat{\theta}_{2,T}$ using information on the true value of the parameters. Among the different distributions, the case of the symmetric $t$ distribution is the most complicated given that its heavy tails violate the moment condition in Theorem 6, while the $\chi^2$ distribution is the most informative given its strong skewness. The results for noninvertible models (mixed and noninvertive cases) in Table A2 in the supplementary materials are similar but, in general, estimates are less precise.

Summary results on estimation of higher order cumulants are severely affected by some extreme replications, and confirm the usual intuition that estimation of higher order moments is quite difficult for small and moderate sample sizes and heavy tail distributions, even more complicated for the kurtosis than for the skewness. Cumulant estimates are biased toward zero, that is, underestimate the non-Gaussianity, the bias growing with the magnitude of the cumulant, and work similarly when based on the unfeasible $\theta_{2,T}$.

The results for $T = 200$ and $\chi^2$ innovations confirm that estimates of models with non invertible roots are more imprecise and that the use of higher order moments at least in the invertible case can outperform estimates using only second order moments. The simulations for the causal VARMA(1, 1) model with mixed MA roots for both $T = 100$ and 200 and $\chi^2$ errors and imposing causality, confirm that kurtosis does not contribute much on top of skewness in terms of bias and variability of estimates, in parallel with its reduced identifying information (see Tables A5 and A6 in Appendix F in the supplementary materials). Cumulant estimation also becomes more difficult with model complexity, but improves substantially with sample size.

8. Empirical Analysis

We apply our inference methods to Blanchard and Quah (1989), BQ henceforth, bivariate system for the U.S. real GNP growth and unemployment rate after linear detrending. They fit a SVAR with eight lags (1948Q2-1987Q4, $T = 159$) and use for identification a long-run restriction by which the demand shock has no long-run effect on real GNP in the same way as both supply and demand shocks have no long-run effect on unemployment. Lippi and Reichlin (1994) alternatively propose that these long VAR dynamics could be better approximated by a VARMA model and explore different versions of the IRF obtained by inverting the MA roots of the fundamental VARMA(1, 1) representation deduced from the fitted VAR(8) parameterization. GMR fit VARMA($p$, 1) models to the same dataset for $p \in \{1, \ldots, 6\}$ by PMLE assuming mixed Gaussian distributions, with GMM initial estimates. They choose a mixed invertible/noninvertible MA representation for $p = 4$ by a combination of model selection criteria and correlation diagnostics.

We fit a series of simple causal VARMA($p$, $q$) models to the original dataset to investigate the possible nonfundamentalness of the dynamics of the system using our higher order cumulant identification. We follow the same MA polynomial parameterization and procedure as in the Monte Carlo simulations with equal weighting $w = (1, 1)$ for each higher order loss function $L^T_{k,T}$, $k = 3, 4$, to obtain GMM estimates $\hat{\theta}_{\text{GMM},T}$ with $V$ estimated by bootstrap using the linear representation of the score.

We report in Table 3 the values of $L^T_{2,T}$, $L^T_{3,T}$, $L^T_{4,T}$, and $L^T_{w,T}$ evaluated at the final estimates (and the value $L^T_{2,T}$ obtained by minimizing only $L^T_{2,T}$ as a benchmark for best linear fit), the modulus of the MA roots of the estimated parameterization and the estimates of $(\alpha_3, \alpha_4)$ using Assumption 6A for component identification. We also report the estimates of $\Omega$ for the instantaneous impact of the shocks on the endogenous variables. We use a similar identification strategy as GMR to label error components as transitory or demand and permanent or supply shocks, facilitating an easy comparison to previous analysis.

The VARMA(1, 1) model appears to best fit the data, reporting the smallest value for the overall loss function $L^T_{w,T}$ after
an efficient GMM Newton–Raphson step, just third best for $L_{2,T}$, outperformed by the much larger SVARMA(4, 1) and VAR(8) models, which also did a better job when fitting only second order dynamics attending to $L_{2,T}^0$, with the first model leading to unit MA roots, possibly indicating a model order overspecification.

The SVARMA(1, 1) model provides an invertible solution, but some more complex models present noninvertible dynamics. Focusing on cumulants estimates, invertible solutions report shocks with moderate skewness, possibly of different signs. The SVARMA(1, 1) identifies the first shock as the transitory with negative skewness, while the second shock would be the permanent one with much larger positive asymmetry, so in both cases negative news (decreasing GNP growth and increasing unemployment) tend to be more extreme than positive ones.

Further, typically one shock displays large positive kurtosis (the permanent one for the SVARMA(1, 1)), but there is no conclusive evidence about the kurtosis of the other one, given the large bootstrap standard errors which make most estimates not significant.

Table 3. U.S. real GNP growth and unemployment rate. GMM SVARMA estimates.

| ($p, q$) | (1, 0) | (4, 0) | (8, 0) | (0, 1) | (1, 1) | (2, 1) | (4, 1) | (1, 2) | (2, 2) |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| $L_{2,T}^0$ | 0.9608 | 0.9991 | 0.7694 | 2.6118 | 0.9136 | 0.9342 | 0.8928 | 0.9395 | 0.9371 |
| $L_{2,T}^1$ | 0.9608 | 1.0393 | 0.8325 | 2.6118 | 0.9148 | 0.9329 | 0.8885 | 0.9395 | 0.9364 |
| $L_{2,T}^2$ | 1.0258 | 1.0345 | 1.0343 | 1.3832 | 1.0284 | 1.0349 | 1.0335 | 1.0285 | 1.0297 |
| $L_{2,T}^3$ | 1.1066 | 1.0784 | 1.2122 | 1.8540 | 1.0116 | 1.0693 | 1.1070 | 1.0578 | 1.0622 |
| $L_{w,T}^4$ | 3.0932 | 3.1522 | 3.0790 | 5.8490 | 2.9548 | 3.0370 | 3.0290 | 3.0258 | 3.0284 |

| MA roots | – | – | – | 3.107 | 8.767 | 0.787 | 1.000 | 0.25, 0.51 | 1.13, 1.32 |
|          | – | – | – | 3.107 | 8.767 | 5.277 | 1.000 | 0.51, 2.81 | 1.32, 2.99 |
|          | – | – | – | (Inv.) | (Inv.) | (Mix.) | (Inv.) | (Inv.) | (Inv.) |

| $\hat{\alpha}_{3,T}$ | -0.447 | -0.542 | -1.378 | -0.461 | -0.526 | -0.302 | -0.272 | -0.855 | -0.296 |
| $\hat{\alpha}_{4,T}$ | 1.497 | 1.575 | 0.344 | 0.511 | 1.284 | 1.485 | 1.926 | -1.222 | 1.423 |
| $\hat{\phi}_{1,1}$ | 5.402 | 0.131 | -1.810 | 1.145 | 6.822 | 7.199 | -1.897 | 7.948 | 7.066 |
| $\hat{\phi}_{2,1}$ | 0.865 | 0.747 | 0.419 | 1.470 | 0.808 | 0.918 | 0.712 | 0.234 | 0.818 |
| $\hat{\phi}_{2,2}$ | (1.03) | (1.90) | (0.07) | (0.07) | (0.08) | (0.091) | (0.026) | (0.025) | (0.287) |
| $\hat{\phi}_{2,2}$ | -0.213 | -0.199 | -0.036 | -0.631 | -0.250 | -0.205 | -0.257 | -0.036 | -0.204 |
| $\hat{\phi}_{2,2}$ | (0.088) | (0.088) | (0.070) | (0.040) | (0.136) | (0.122) | (0.027) | (0.194) | (0.173) |

| $\hat{\phi}_{2,2}$ | 0.089 | 0.144 | 0.038 | 0.304 | 0.245 | -0.058 | 0.281 | -0.070 | 0.105 |
| $\hat{\phi}_{2,2}$ | (0.308) | (0.202) | (0.074) | (0.194) | (0.1619) | (0.187) | (0.027) | (0.194) | (0.261) |
| $\hat{\phi}_{2,2}$ | 0.191 | 0.169 | 0.057 | 0.196 | 0.147 | 0.230 | 0.109 | 0.070 | 0.183 |
| $\hat{\phi}_{2,2}$ | (0.111) | (0.097) | (0.073) | (0.208) | (0.140) | (0.156) | (0.057) | (0.170) | (0.192) |

Figure 1. IRF for U.S. GNP growth and unemployment based on our GMM estimates (9) for SVARMA(1, 1), BQ: SVAR(8), and GMR: SVARMA(4, 1).

Table 3. U.S. real GNP growth and unemployment rate. GMM SVARMA estimates.

Note: SVARMA($p, q$) model fitting for U.S. real GNP growth and unemployment rate after linear detrending, 1948Q2–1987Q4, $T = 159$. GMM estimates $\hat{\theta}_{GMM,T}$ in (9) obtained with $\hat{\theta}_{T} = \hat{\theta}_{w,T}$, $w = (1, 1)$ and a bootstrap estimate $\hat{G}_{T}$ with 200 replications of the joint score $G_{T}$ evaluated at $\hat{\theta}_{w,T}$.
components, but the sign of the instantaneous impact of the permanent shock on output is reversed.

Finally, we provide in Figure 1 the plots of the IRFs identified by our estimation methods for the SVARMA(1, 1) model together with those reported by BQ and GMR. We can observe that the effect of the supply (or permanent) shock on both endogenous variables is very close to the results of BQ for the SVAR(8) model, though there is a slight delay in the maximum effect and a quite more persistent effect on output. For the demand (transitory) shock, the shape of both IRFs and the timing of maximal effects are almost the same as for the SVAR model, but now the long run effect on output is not restricted to zero and the IRF displays a very slow rate of decay but close to the horizontal axis. We finally perform an overidentification bootstrap test for the long-run restriction used by BQ based on the resampling method of Section 6. In particular, we cannot reject the plausibility of this identification strategy based on a significance test of the IRF coefficients \( \Phi_{1}^{(1)} \) of GNP growth to the transitory shock at lags \( j = 40, 100 \) computed from the estimated SVARMA(1, 1) model, as could be expected from the large standard errors of estimates. On the other hand, the IRFs obtained by GMR with noninvertible dynamics retain some of the previous properties for the unemployment response, but both output growth IRFs behave quite differently in the long run.

9. Conclusions

In this article we have showed how to achieve identification of non-Gaussian SVARMA models using rank conditions on dynamics and higher order cumulant matrices under serial and component independence conditions of finite order \( k = 3 \) or 4 on the structural shocks sequence. We use an identification criterion in the frequency domain that leads to global and local identification conditions and permits the design of consistent and asymptotically normal parameter estimates which exploit all dynamic and static information in second, third, and fourth order moments. These results provide consistent estimation of IRFs without need to specify the fundamentality of the system and can be combined with different sources of information to proper label the structural shocks of the model or to test relevant hypothesis and overidentification conditions.

Supplementary Materials

The Supplemental Materials contain an Appendix with proofs of results, additional discussion and further numerical results, and Matlab code for the Monte Carlo simulations and empirical section.

Acknowledgments

Thanks to an Associate Editor and two referees, and to J. C. Escanciano, F. J. Hidalgo, I. N. Lobato, G. Sucarrat and seminar participants at Nuffield College, LSE, QMUL, BI Norwegian BS and at Workshop on Time Series Econometrics 2018, Waseda International Symposium 2018, CFE-CM Statistics Conference 2019, EC² Conference on Identification in Macroeconomics 2019 for helpful discussions and comments.

Funding

Financial support from the Ministerio Economía y Competitividad (Spain) grants ECO2017-86009-P and PID2020-114664GB-I00 is gratefully acknowledged.

References

Alessi L., Barigozzi, M., and Capasso, M. (2011), "Non-Fundamentalness in Structural Econometric Models: A Review," International Statistical Review, 79, 16–47. [819]

Andrews, B., Davis, R. A., and Breidt, J. (2007), "Rank-based Estimation for All-pass Time Series Models," Annals of Statistics, 35, 844–869. [820]

Baggio G., and Ferrante, A. (2019), "Estimation of Minimal Spectral Factors of Discrete-Time Rational Spectral Densities," IEEE Transactions on Automatic Control, 64, 396–403. [828]

Blanchard, O., and Quah, D. (1989), "Dynamic Effects of Aggregate Demand and Supply Disturbances," The American Economic Review, 79, 655–673. [820,829]

Boubacar Mainassara, Y. B., and Francq, C. (2011), "Estimating Structural VARMA Models with Uncorrelated but Non-independent Error Terms," Journal of Multivariate Analysis, 102, 496–505. [824]

Brillinger, D. R. (1985), "Fourier Inference: Some Methods for the Analysis of Array and Non-Gaussian Series Data," Water Resources Bulletin, 21, 744–756. [825]

Chan, K.-S., and Ho, L.-H. (2004), "On the Unique Representation of Non-Gaussian Multivariate Linear Processes," Technical Report No 341, Department of Statistics and Actuarial Science, The University of Iowa. [819,821,822,823]

Chan, K.-S., Ho, L.-H., and Tong, H. (2006), "A Note on Time-Reversibility of Multivariate Linear Processes," Biometrika, 93, 221–227. [819,820]

Chen, B., Choi, J. and Escanciano, J.C. (2017), "Testing for Fundamental Vector Moving Average Representations," Quantitative Economics, 8, 149–180. [822]

Comon, P. (1994), "Independent Component Analysis, A New Concept?" Signal Processing, 36, 287–314. [819,822,823]

Funovits, B. (2021), "Identifiability and Estimation of Possibly Non-Invertible SVARMA Models: A New Parametrisation," arXiv:2002.04346. [824]

Gospodinov, N., and Ng, S. (2015), "Minimum Distance Estimation of Possibly Noninvertible Moving Average Models," Journal of Business and Economic Statistics, 33, 403–417. [820]

Gourieroux, C., Monfort, A., and Renne, J.-P. (2017), "Statistical Inference for Independent Component Analysis: Application to Structural VAR Models," Journal of Econometrics, 196, 111–126. [820]

——— (2020), "Identification and Estimation in Non-Fundamental Structural VARMA Models," The Review of Economic Studies, 87, 1915–1953. [819,828]

Granziera, E., Moon, H. R., and Schorfheide, F. (2018), "Inference for VARs Identified with Sign Restrictions," Quantitative Economics, 9, 1087–1121. [821]

Guay, A. (2021), "Identification of Structural Vector Autoregressions Through Higher Unconditional Moments," Journal of Econometrics, 225, 27–46. [823]

Hannan, E. J. (1970), Multiple Time Series, New York: Wiley. [821]

Hannan, E. J., and Deistler, M. (1988), The Statistical Theory of Linear Systems, New York: Wiley. [822]

Hyvärinen, A., Zhang, K., Shimizu, S., and Hoyer, P. O. (2010), "Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity," Journal of Machine Learning Research, 11, 1709–1731. [819]

Jammalamadaka, S. R., Rao, T. S., and Terdik, G. (2006), "Higher Order Cumulants of Random Vectors and Application to Statistical Inference of a Structural Vector Autoregression Model Using Non-Gaussianity," Sankhya: The Indian Journal of Statistics, 68, 326–356. [821]

Keweloeh, S. A. (2021), "A Generalized Method of Moments Estimator for Structural Vector Autoregressions Based on Higher Moments," Journal of Business Economic Statistics, 39, 772–782. [823]

Lanne, M., and Luoto, J. (2021), "GMM Estimation of Non-Gaussian Structural Vector Autoregression," Journal of Business and Economic Statistics, 39, 69–81. [820,823,824]
Lanne, M., and Lütkepohl, H. (2010), "Structural Vector Autoregressions With Nonnormal Residuals," *Journal of Business and Economic Statistics*, 28, 159–168. [820]
Lanne, M., Meitz, M., and Saikkonen, P. (2017), "Identification and Estimation of Non-Gaussian Structural Vector Autoregressions," *Journal of Econometrics*, 196, 288–304. [820,824]
Lanne, M., and Saikkonen, P. (2013), "Noncausal Vector Autoregression," *Econometric Theory*, 29, 447–481. [819,824]
Lii, K.-S., and Rosenblatt, M. (1982), "Deconvolution and Estimation of Transfer Function Phase and Coefficients for Non-Gaussian Linear Processes," *The Annals of Statistics*, 4, 1196–1208. [820,821]
Lobato, I. N., and Velasco, C. (2018), "Efficiency Improvements for Minimum Distance Estimation of Causal and Invertible ARMA Models," *Economics Letters*, 162, 150–152. [825]
Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag. [824]
Lütkepohl, H., and A. Netšunajev (2017), "Structural Vector Autoregressions with Heteroskedasticity: A Review of Different Volatility Models," *Econometrics and Statistics*, 1, 2–18. [820]
Pham, D. T., and Garat, P. (1997), "Blind Separation of Mixture of Independent Source Through a Quasi-Maximum Likelihood Approach," *IEEE Transactions on Signal Processing*, 45, 1712–1725. [824]
Rubio-Ramirez, J. F., Waggoner, D. F., and Zha, T. A. (2010), "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *The Review of Economic Studies*, 77, 665–696. [821]
Velasco, C., and Lobato, I. N. (2018), "Frequency Domain Minimum Distance Inference for Possibly Noninvertible and Noncausal ARMA models," *Annals of Statistics*, 46, 555–579. [820]

Lipari, M., and Reichlin, L. (1994), "VAR Analysis, Nonfundamental Representations, Blaschke Matrices," *Journal of Econometrics*, 63, 307–325. [819,820,821,829]

(1992), "An Approximate Maximum Likelihood Estimation for non-Gaussian non-minimum Phase MA Processes," *Journal of Multivariate Analysis*, 43, 272–299. [820]

---

Lobato, I. N., and Velasco, C. (2018), "Efficiency Improvements for Minimum Distance Estimation of Causal and Invertible ARMA Models," *Economics Letters*, 162, 150–152. [825]
Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag. [824]
Lütkepohl, H., and A. Netšunajev (2017), "Structural Vector Autoregressions with Heteroskedasticity: A Review of Different Volatility Models," *Econometrics and Statistics*, 1, 2–18. [820]
Pham, D. T., and Garat, P. (1997), "Blind Separation of Mixture of Independent Source Through a Quasi-Maximum Likelihood Approach," *IEEE Transactions on Signal Processing*, 45, 1712–1725. [824]
Rubio-Ramirez, J. F., Waggoner, D. F., and Zha, T. A. (2010), "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *The Review of Economic Studies*, 77, 665–696. [821]
Velasco, C., and Lobato, I. N. (2018), "Frequency Domain Minimum Distance Inference for Possibly Noninvertible and Noncausal ARMA models," *Annals of Statistics*, 46, 555–579. [820]