Generalized Covariance-Based Inference for Models Partially Identified from Independence Restrictions

Christian Gourieroux *, Joann Jasiak ‡

October 3, 2023
preliminary version

Abstract

This paper develops statistical inference methods for a class of partially identified models, where the errors are known functions of observations and the parameters satisfy either serial or/and cross-sectional independence conditions. This class of models includes the independent component analysis (ICA) models, Structural Vector Autoregressive (SVAR) models and multivariate mixed causal-noncausal models. We use the Generalized Covariance (GCov) estimator to compute the residual-based portmanteau statistic for testing the error independence hypothesis. Next, we build the confidence sets for the identified sets of parameters by inverting the test statistic. We also discuss the choice (design) of these statistics. The approach is illustrated by simulations examining the under-identification condition in an ICA model and by an application to financial return series.

Keywords: Generalized Covariance (GCov) Estimator, Portmanteau Statistic, Partial Identification, Implied Identified Sets, Independent Component Analysis, Structural VAR.

MSC Codes: 62F03, 62F25, 62P20, 62M10

*University of Toronto, Toulouse School of Economics and CREST, *e-mail*: gouriero@ensae.fr

‡York University, Canada, *e-mail*: jasiakj@yorku.ca

The authors gratefully acknowledge financial support of the Natural Sciences and Engineering Council of Canada (NSERC)

1 Introduction

This paper develops statistical inference methods for a class of partially identified models, where the errors are known functions of the observations and the parameters satisfy either serial or/and cross-sectional independence conditions. This class of models contains the independent component analysis (ICA) models, Structural Vector Autoregressive (SVAR) models, multivariate mixed causal-noncausal models and nonlinear dynamic models with predictable drift and volatility [1].

We introduce the notions of a true identified set and an implied identified set. Our approach consists in considering serial and cross-sectional dependence measures based on the multivariate objective function of the Generalized Covariance (GCov) estimator and a portmanteau test statistic for testing the independence hypothesis. Next, we build the confidence sets for the identified sets of parameters by inverting the tests of the independence hypothesis based on the residual-based portmanteau statistic. This statistic involves nonlinear covariance restrictions that are satisfied under the independence condition. We discuss the choice of covariance restrictions and their impact on the dimension and accuracy of the confidence set.

The identification problems are particularly important in macroeconomic applications, where the theoretical structural vector autoregressive (SVAR) processes with cross-sectionally independent errors have more parameters than the estimable statistical VAR models because of simultaneity effects. Various restrictions, including zero coefficient restrictions and sign restriction have been proposed to eliminate this identification problem[see e.g. Granziera, Moon, Schorfheide (2018)]. The recent macroeconomic literature considers an alternative approach for the SVAR model under the assumption of non-Gaussian errors [Guay (2021), Gourieroux, Monfort, Renne (2020)]. Then, as long as at most one of the errors is Gaussian, the independent component analysis (ICA) can be used to identify the parameters and the cross-sectionally independent latent sources. However, if the non-Gaussianity condition does not hold and if more than one error has a Gaussian or close to a Gaussian distribution, the parameters of the orthogonal rotation matrix become non-identifiable.

We show that despite of the lack of identification, asymptotically valid CI can be provided. When they are empty, they are informative in the sense of signaling that the estimation results may be spurious. The approach is based on a dependence measure with a known limiting distribution, and involves the nonlinear autocovariances, i.e. autocovariances of nonlinear

[1] The class of models considered differs from the class of models in which the identified set is characterized by a finite number of moment inequalities, used commonly for micro-econometric applications [see e.g. Chernozhukov, Hong, Tamer (2007), Canay, Shaikh (2017)]

functions of errors. It can be compared with other statistics based on the covariance distance, considered for example by Davis, Wan (2022) and Chu (2023). The advantage of the proposed dependence measure is that its limiting distribution is known, while the covariance distance based measures require respectively the use of bootstrap and approximations in practice.

The paper is organized as follows. Section 2 describes the semi-parametric models with independence restrictions on the error terms and introduces dependence measures based on the (auto) covariances of transformed errors. Section 3 considers the identification, estimation and testing. It shows how to estimate the implied identified sets pointwise and from the asymptotic confidence sets. Section 4 defines the overidentification problems that arise in the partial identification framework and develops statistical methods for overidentification analysis. Section 5 provides a simulation study that illustrates the performance of the method in application to independent component analysis. It also contains an application of the proposed method to financial return series. Section 6 concludes. Proofs are given in the Appendices.

2 Semi-Parametric Models and Error Dependence Measures

This Section describes the semi-parametric models with errors satisfying either serial and/or cross-sectional independence assumptions. We consider the static Independent Component Analysis models [Comon (1994), Eriksson, Koivunen (2004)] and multivariate dynamic models including the nonlinear Markov models and mixed causal-noncausal models [Gourieroux, Zakoian (2017), Gourieroux, Jasiak (2017)]. Other dynamic models in which the identification problem may arise are also discussed, especially the Structural VAR (SVAR) models. For each of these models, we design dependence measures based on the covariances between nonlinear error transformations.

2.1 Independent Component Analysis

Consider a static model defined by the following system of equations:

$$CY_i = u_i, \quad i = 1, ..., n, \tag{2.1}$$

where $Y_i, i = 1, ..., n$ are of dimension K, the errors $u_i, i = 1, ..., n$ are assumed to be serially independent and identically distributed, and their components $u_{ki}, k = 1, ..., K$, are cross-sectionally independent. We denote $f = (f_k, k = 1, \ldots, K)$ the set of densities f_k of the

$u_{ki}, k = 1, \dots, K$. The matrix C is an invertible matrix $K \times K$ of parameters with the diagonal elements, which are set equal to 1, without loss of generality.

The system (2.1) can be rewritten as:

$$Y_i = C^{-1} u_i, \quad i = 1, ..., n, \tag{2.2}$$

The identification and estimation of matrix C is a multivariate deconvolution problem, as the distribution of any component Y_{ji} is a convoluate of distributions of the error components (called sources) $u_{ki}, k = 1, ..., K$. The mixing matrix C^{-1} and the sources u_i are not always identifiable. For example, if the components $u_{ki}, k = 1, ..., K$ are standard normal, i.e. N(0,1) variables, we can identify $C^{-1}(C^{-1})'$, but not the mixing matrix itself. However, if at most one component u_{ki} follows a Gaussian distribution, then the matrix C^{-1}, the sources u_i and their distributions are identifiable [Comon (1994), Eriksson, Koivunen (2004)]. There can also arise intermediate situations with partial identification, when several components are Gaussian and the others are non-Gaussian [Guay (2021)].

Let us now introduce an error dependence measure. First, the system is rewritten equation by equation as:

$$c_k' Y_i = u_{ki}, \quad k = 1, ..., K, i = 1, ..., n, \tag{2.3}$$

where c_k' is the k^{th} row of matrix C. Next, we apply to (2.3), a set of L nonlinear transforms $a(.)$ satisfying Assumptions A.1 given in Section 3.1, to get:

$$a[c_k' Y_i] = a[u_{ki}], k = 1, ..., K, i = 1, ..., n. \tag{2.4}$$

Let $\theta = (c_1', ..., c_K')'$ denote the vector of unknown parameters. The cross-sectional dependence measure is defined as:[2]

$$\xi(\theta) = \sum_{k,l} \sum_{k>l} Tr[\Gamma_{k,l}(\theta)\Gamma_{l,l}(\theta)^{-1}\Gamma_{l,k}(\theta)\Gamma_{k,k}(\theta)^{-1}], \tag{2.5}$$

with

$$\Gamma_{k,l}(\theta) = Cov[a(c_k' Y), a(c_l' Y)] = \Gamma_{l,k}(\theta)', \ \forall k, l. \tag{2.6}$$

The dependence measure is a multivariate portmanteau measure [Hosking (1980)] that involves pairwise canonical correlations [Anderson (1999)]. If the model (2.1) is well-specified with the true value θ_0 of parameter θ and true densities $f_{k0}, k = 1, \dots, K$ of the sources,

[2]This measure depends on both θ and f and should be denoted $\xi(\theta, f)$. The dependence on f is omitted to simplify the notation.

this dependence measure is equal to 0. This measure depends on the number and type of covariances appearing in (2.5), called the design later on. In this case, the design is

$$\mathcal{D} = [a, K(K-1)/2], \tag{2.7}$$

where a is the set of transformations, $K(K-1)/2$ is the number of terms in the double sum and the total number of covariances is:

$$d = \dim \mathcal{D} = L^2 K(K-1)/2. \tag{2.8}$$

The dimension of the parameter is : $\dim \theta = K(K-1)$, because the diagonal elements of C are equal to 1. It is also useful to evaluate how many covariances are informative about θ. This can be locally measured by the rank of $\partial\gamma(\theta)/\partial\theta'$,

$$r = Rk[\frac{\partial\gamma(\theta)}{\partial\theta'}], \tag{2.9}$$

where γ is the vector representation of all covariances obtained by stacking the subvectors $\gamma_{kl}(\theta) = vec\Gamma_{kl}(\theta),\ k > l$, into a column vector.

The nonlinear transformations need to possibly accommodate fat tails of the sources distributions. We know that the non-Gaussian Pseudo-Maximum Likelihood (PML) can provide consistent estimates of the mixing matrix if the tails satisfy some regularity conditions, because the first-order conditions of PML optimization are zero covariance conditions on the score [Gourieroux, Monfort, Renne (2017)]. For example, the transformation associated with a Cauchy PML is :

$$a(y) = -\frac{\partial}{\partial m}\log[1 + (y-m)^2] = 2\frac{y-m}{1+(y-m)^2}.$$

This transformation ensures the existence of power moments of $a(Y)$ for the Cauchy, whereas they do not exist for the variable Y itself.

Other transformations, such as the power transforms can be used. For example, let us consider the ICA for a process Y of dimension 2. The components of the process are denoted by Y_1 and Y_2. The source components, denoted by u_1 and u_2, have zero means and unitary variances. The portmanteau statistic can be based on the power covariances: $Cov(Y_1^j, Y_2^k),\ j,k = 1,2,3$. These covariances are not always informative about the mixing matrix C. In particular, some covariances can be equal to zero for any matrix C and a specific source distribution. These cases are described in Table 1 (see On-line Appendix 1 for the proof).

Table 1: Conditions of Non-Informative Power Covariance

Covariance:	Non-Informative when:
$Cov(Y_1, Y_2)$	always non-informative
$Cov(Y_1^2, Y_2)$, $Cov(Y_1, Y_2^2)$	$Eu_1^3 = Eu_2^3 = 0$
$Cov(Y_1^2, Y_2^2)$	$E(u_1^4) + E(u_2^4) - 6 = 0$
$Cov(Y_1^3, Y_2)$, $Cov(Y_1, Y_2^3)$	$E(u_1^4) = E(u_2^4) = 3$
$Cov(Y_1^2, Y_2^3)$, $Cov(Y_1^3, Y_2^2)$	$E(u_1^3) = E(u_1^5) = E(u_2^3) = E(u_2^5) = 0$
$Cov(Y_1^3, Y_2^3)$	$E(u_1^3) = E(u_2^3),\ E(u_1^4) = E(u_2^4),\ E(u_1^6) = E(u_2^6)$

All power covariances are non-informative when u_1 and u_2 are Gaussian, which corresponds to the lack of identification of the mixing matrix in Gaussian ICA. Nevertheless, if u_1 is Gaussian and u_2 follows a skewed distribution, or a distribution with fat tails, such that the kurtosis is greater than 3, $E(u_2^4) > 3$, then some power covariances become informative and the mixing matrix can be identified.

2.2 Dynamic Model with Serial Independence

Let us consider the model:

$$g(\tilde{Y}_t, \theta) = u_t, \tag{2.10}$$

where the $u_t's$ are i.i.d, of dimension K and have finite moments up to order 4, $\tilde{Y}_t = (Y_t, Y_{t-1})$, is observable, θ is an unknown parameter (that can be partially identifiable) and g a known function.

Example 1: **Model with predictable drift and volatility**

Let us consider a univariate process (y_t) with predictable drift, volatility and risk premium. The nonlinear dynamic model is:

$$y_t = m(y_{t-1}; \theta_1) + \theta_3 \sigma(y_{t-1}; \theta_2) + \sigma(y_{t-1}; \theta_2) u_t, \tag{2.11}$$

where the errors $u_t,\ t = 1, ..., T$ are i.i.d. and function $\sigma()$ is assumed positive. The model contains the drift parameter θ_1, volatility parameter θ_2 and risk premium parameter θ_3. This model is a Markov equivalent of the ARCH-M model [Engle, Lilien, Robbins (1987)] and extends the Double Autoregressive (DAR) model [Borkovec, Kluppelberg (2001), Ling (2007)] by introducing a risk premium. It can be rewritten under the specification (2.10) for the conditionally demeaned and standardized y_t as:

$$g(\tilde{y}_t; \theta) = [y_t - m(y_{t-1}; \theta_1) - \theta_3 \sigma(y_{t-1}; \theta_2)] / \sigma(y_{t-1}; \theta_2) = u_t. \tag{2.12}$$

The parameters $\theta_1, \theta_2, \theta_3$ are not always identifiable. The identification depends on the functional form of the drift and volatility functions and on the distribution of the errors. In particular, it can be difficult to disentangle the regular part of the drift from the risk premium.

Example 2: **Stochastic Volatility Model**

Let us consider the observations on an asset return y_t and on implied volatility σ_t computed from an at-the-money option written on this asset. A bivariate dynamic model for observed variable (y_t, σ_t) can be written as:

$$\begin{aligned} y_t &= \theta_1 y_{t-1} + \sigma_t u_{1t}, \\ \log \sigma_t &= \theta_2 + \theta_3 \log \sigma_{t-1} + u_{2t}. \end{aligned}$$

The condition of mutual independence of u_{1t} and u_{2t} is a condition of no leverage effect [Black (1976)]. When the independence is satisfied, the error u_{2t} can be interpreted as a primitive volatility shock and used to derive the associated impulse response functions. However, identification of the impulse response functions may become an issue if the errors are cross-sectionally dependent, for example a linear mixture of independent sources.

Example 3: **Causal SVAR(1) Model**

The multivariate SVAR(1) model is:

$$\Phi_0 Y_t + \Phi_1 Y_{t-1} = u_t, \tag{2.13}$$

where the $u_t's$ are i.i.d, the diagonal elements of Φ_0 are set equal to 1 and the eigenvalues of $\Phi_0^{-1}\Phi_1$ are of modulus strictly less than 1. Then, the parameter vector is $\theta = [(vec\Phi_0)', (vec\Phi_1)']'$ of dimension $\dim \theta = 2K^2 - K$. The partial identification issue comes from the simultaneity that can be captured by the matrix Φ_0 as well as by the cross sectional dependence of the $u_t's$.

Example 4: **Multivariate Mixed Causal-Noncausal Model**

These models have been shown in the literature to reproduce the bubbles and local trends observed in financial data, such as the commodity prices [Gourieroux, Zakoian (2017), Gourieroux, Jasiak (2017)]. The mixed VAR(1) model has the form (2.13), but the eigenvalues of $\Phi_0^{-1}\Phi_1$ are of modulus different from 1, not necessarily strictly less than 1. The noncausal directions of eigenvectors associated with the eigenvalues outside the unit circle generate local explosive patterns perceived as bubbles or local trends.

In Section 2.1 we showed that a set of L transformations can be used to replace system (2.9) by:

$$a[g(\tilde{Y}_t;\theta)] = a(u_t). \tag{2.14}$$

Let us next consider the measure of serial dependence:

$$\xi(\theta) = \sum_{h=1}^{H} Tr[\Gamma(h;\theta)\Gamma(0;\theta)^{-1}\Gamma(h;\theta)'\Gamma(0;\theta)^{-1}], \tag{2.15}$$

where $\Gamma(h,\theta) = Cov(a[g(\tilde{Y}_t;\theta)], a[g(\tilde{Y}_{t-h};\theta)])$ and $\tilde{Y}_{t-h} = (Y'_{t-h}, Y'_{t-h-1})'$.
The above equation defines a portmanteau serial dependence measure applied to the transformations a of the error term [see, Gourieroux, Jasiak (2022)]. The design is now:

$$\mathcal{D} = [a, H],$$

where H is the number of terms in the portmanteau measure (2.14) and the total number of covariances is:

$$d = dim\mathcal{D} = L^2 H.$$

2.3 Dynamic Models with Serial and Cross-Sectional Independence Restrictions

The serial and cross-sectional independence restrictions can be imposed on the model (2.10):

$$g(\tilde{Y}_t;\theta) = u_t,\ t = 1, ..., T, \tag{2.16}$$

where u_t are i.i.d. variables and the components $u_{1t}, ..., u_{Kt}$ are independent. In particular, these independence restrictions can be applied to a SVAR(1) model:

$$\Phi_0 Y_t + \Phi_1 Y_{t-1} = u_t \iff Y_t = \Psi Y_{t-1} + B u_t,$$

with $\Psi = \Phi_0^{-1}\Phi_1$ and $B = \Phi_0^{-1}$.
This approach allows for distinguishing the errors $\epsilon_t = Bu_t$ in the above system of equations from the primitive shocks $u_{1t}, ..., u_{Kt}$ assumed mutually independent [see e.g. Gourieroux, Monfort, Renne (2017), Lanne, Meitz, Saikkonen (2017), Davis, Ng (2022) for a discussion on the differences between the errors and the primitive shocks in structural models].

Then, the model can be written as a system of equations for each component:

$$g_k(\tilde{Y}_t;\theta) = u_{k,t},\ k = 1, ..., K,\ t = 1, ..., T. \tag{2.17}$$

The associated dependence measure is:

$$\xi(\theta) = \sum_{h=1}^{H} \sum_{k,l:} \sum_{k \neq l} Tr[\Gamma_{kl}(h;\theta)\Gamma_{ll}(0;\theta)^{-1}\Gamma_{lk}(h;\theta)'\Gamma_{kk}(0;\theta)^{-1}], \tag{2.18}$$

with

$$\Gamma_{kl}(h,\theta) = Cov(a[g_k(\tilde{Y}_t;\theta)], a[g_l(\tilde{Y}_{t-h};\theta)]). \tag{2.19}$$

Then, the design is $\mathcal{D} = [a,\ HK(K-1)/2]$ of dimension $d = dim\mathcal{D} = HL^2K(K-1)/2$.

3 Identification, Estimation and Tests

The models of Section 2 and the associated dependence measures have a common structure that can be used as a general representation for statistical inference. We start with discussing the notion of an identified set. Next, we describe the pointwise estimation and confidence set estimation methods.

3.1 Identification

Suppose that model (2.10) satisfies the following assumptions:

Assumption A.1:

i) The model is well-specified with θ_0 the true value of parameter θ and f_0 the true probability density function (pdf) of error u_t (that may satisfy the independence restrictions on the components $u_{k,t},\ k = 1, ..., K$ in the models of Sections 2.1, 2.3.). The components of u_t have finite fourth order moments.

ii) The process (Y_t) is strictly stationary, geometrically mixing.

iii) The function g is invertible with respect to Y_t and differentiable.

Under Assumption A.1 iii), the model can be rewritten as:

$$Y_t = g^{-1}(u_t, Y_{t-1};\theta_0),\ \ u_t \sim f_0, \tag{3.1}$$

as a nonlinear autoregression that can be univariate or multivariate and includes the static (i.i.d.) linear model of Section 2.1. There exist several solutions to the autoregressive model

(3.1) and this multiplicity can be reduced by using the assumption A.ii) on the existence and uniqueness of a strictly stationary solution. This implies a restriction on the true θ_0, f_0.

The differentiability condition in A.1 iii) allows for applying the Jacobian formula and insures the existence of the true transition density of Y_t given Y_{t-1} denoted by $l(y_t|y_{t-1};\theta_0,f_0)$. The transition density depends on both types of true parameters. Thus we have a semi-parametric model in which we can define the true identified set for parameter θ.

Definition 1:
For a given error distribution f_0, the identified set for θ_0 is: $\Theta_0(f_0) = \{\theta, l(y_t|y_{t-1};\theta,f_0) = l(y_t|y_{t-1};\theta_0,f_0)$ almost everywhere in $y_t, y_{t-1}\}$.

In general, this identified set depends on f_0 and is not reduced to a singleton. Under the rank regularity condition, the identified set is locally a manifold of dimension $dim\,\Theta_0(f_0)$ that is strictly positive in the presence of partial identification . However there can exist models and distributions f_0 such that $\Theta_0(f_0) = \{\theta_0\}$ is a singleton corresponding to the identifiable true value θ_0. In this case the dimension of $\Theta_0(f_0)$ is equal to zero. We illustrate this case in the example below.

Example 2 (follows) : Causal SVAR(1) model
Let us consider a SVAR(1) model:

$$\Phi_0 Y_t + \Phi_1 Y_{t-1} = u_t,$$

where the diagonal elements of Φ_0 are equal to 1. This model can be rewritten as:

$$Y_t = \Phi_0^{-1}\Phi_1 Y_{t-1} + \Phi_0^{-1}u_t.$$

i) If u_t is Gaussian $u_t \sim N(0,\Sigma)$ with Σ unknown, we can identify $\Phi_0^{-1}\Phi_1$ and $\Phi_0^{-1}\Sigma(\Phi_0^{-1})'$, i.e. only the transformation $\Phi_0^{-1}\Phi_1$ of the parameters of interest Φ_0 and Φ_1. The dimension of the identified set is $K(K-1)$ equal to the number of parameters in matrix Φ_0.

ii) If the components u_{kt}'s of u_t's are independent, and, for example, t-distributed, possibly with different degrees of freedom, then both $\Phi_0^{-1}\Phi_1$ and Φ_0^{-1} are identifiable and the identified set is reduced to a singleton.

iii) If these components are independent, with p Gaussian components, the other ones t-distributed, Φ_0, Φ_1 are partly identified with a degree of underidentification depending on p.

The true identified set depends on the entire conditional density, whereas the portmanteau dependence measure is based on a finite number of (true) covariances that are not necessarily informative (see the discussion in Section 2.1). For example, let us consider the dependence measure defined in Section 2.2:

$$\xi_0(\theta) = \sum_{h=1}^{H} Tr[\Gamma_0(h;\theta)\Gamma_0(0;\theta)^{-1}\Gamma_0(h;\theta)'\Gamma_0(0;\theta)^{-1}],$$

where the subscript 0 is used to point out that the above measure is evaluated at the true values θ_0, f_0.

Definition 2: The implied identified set associated with $\xi_0(\theta)$ is :

$$\Theta_0^*(f_0) = \{\theta : \xi_0(\theta) = 0\}.$$

By construction and Assumption A.1, we have:

$$\theta_0 \in \Theta_0(f_0) \subset \Theta_0^*(f_0). \tag{3.2}$$

$\Theta_0^*(f_0)$ is another (sub)manifold of dim $\Theta_0^*(f_0) \geq$ dim $\Theta_0(f_0)$.

If the set of covariances is sufficiently informative, we expect that:

$$\theta_0 \in \Theta_0(f_0) = \Theta_0^*(f_0).$$

If additionally θ_0 is point identified, we have $\{\theta_0\} = \Theta_0(f_0) = \Theta_0^*(f_0)$, although in general this is not the case.

The portmanteau measure and the implied identified set depend on the design $\mathcal{D}$. To keep the exposition simple, this dependence has not been specified yet. It will be clarified later in the discussion of "overidentification" of Section 4.

3.2 Pointwise Estimation of the Implied Identified Set

The true identified set as well as the implied identified set depend on the true DGP and are unknown. They can be estimated as follows:

i) In the first step, we replace the portmanteau dependence measure by a consistent estimator. For example, $\xi_0(\theta)$ defined in Section 2.2 will be replaced by:

$$\hat{\xi}_T(\theta) = \sum_{h=1}^{H} Tr[\hat{\Gamma}_T(h;\theta)\hat{\Gamma}_T(0;\theta)^{-1}\hat{\Gamma}_T(h;\theta)'\hat{\Gamma}_T(0;\theta)^{-1}], \tag{3.3}$$

where $\hat{\Gamma}_T(h;\theta)$ is the sample analogue of $\Gamma_0(h;\theta)$ computed from the observations $Y_t,\ t = 1,..,T$.

ii) The estimated implied identified set is:

$$\hat{\Theta}_T^* = \{\theta : \theta = Argmin_{\theta \in \Theta}\, \hat{\xi}_T(\theta)\}. \tag{3.4}$$

In practice, the estimated implied identified set is equivalently defined from the First-Order Conditions (FOC) as:

$$\hat{\Theta}_T^* = \{\partial \hat{\xi}_T(\theta)/\partial\theta = 0, \partial^2 \hat{\xi}_T(\theta)/\partial\theta\partial\theta' >> 0\},$$

where $>>$ is the Loewner ordering on symmetric matrices.

Proposition 1: Under Assumption A.1 and additional assumptions AA given in Online Appendix 2, $\hat{\Theta}_T^*(f_0)$ is a consistent estimator of $\Theta_0^*(f_0)$ in the sense that the Hausdorff distance between these two sets tends to 0, when T tends to infinity.

The consistency of the estimated implied identified set in the sense of the convergence to its theoretical counterpart follows from the results in Shi, Shum (2015) [see also Chernozhukov et al. (2007) for similar arguments]. This convergence result does no imply the convergence of the dimension $dim\, \hat{\Theta}_T^*(f_0)$ to $dim\, \Theta_0^*(f_0)$ because the function $dim(.)$ is not continuous.

Proposition 1 provides a result that is valid for partial identification. In particular, it is valid if $\{\theta_0\} = \Theta_0(f_0) = \Theta_0^*(f_0)$. In this special case, it can be written as follows [see Gourieroux, Jasiak (2022)]: "There exists a solution $\hat{\theta}_T = Argmin_{\theta \in \Theta}\, \hat{\xi}_T(\theta)$ that converges to the true value θ_0 ". This solution is called the Generalized Covariance (GCov) estimator of θ_0. In the general case, $\hat{\Theta}_T^*$ is the GCov estimator of the implied identified set.

The partial identification of parameter θ implies also the partial identification of the error terms $U_T = (u_1, ..., u_T)'$. A consistent estimator of the implied identified set of errors is derived from the estimated residuals as:

$$\hat{U}_T^* = \{\hat{u}(T) : \hat{u}_t(T) = g(Y_t, Y_{t-1}; \theta), t = 1, ..., T \text{ with } \theta \in \hat{\Theta}_T^*\}.$$

Thus, the error terms are only partially recovered. This estimated residual set is doubly stochastic with the direct effect of Y_t, Y_{t-1} and the asymptotic uncertainty on Θ_T^*. In particular, under the maintained hypothesis of i.i.d. error terms u_t, we cannot expect to get the independence with identical distribution of the sets $\{\hat{u}_{tT} \text{ with } \hat{u}_T \in \hat{U}_T^*\},\ t = 1, ..., T$.

The same remark holds for other (stochastic) functions of parameter θ, such as the impulse response functions [Rubio-Ramirez, Waggoner, Zha (2010), Moon, Schorfheide (2012), Granziera, Moon, Schorfheide (2018), Gourieroux, Jasiak (2023)].

3.3 Asymptotic Confidence Set of the Implied Identified Set

In the case of identifiable θ_0 with $\Theta_0^*(f_0) = \Theta_0(f_0) = \{\theta_0\}$, we know that the minimized $\hat{\xi}_T(\theta)$ is chi-square distributed and the portmanteau test statistic computed from the residuals is such that:

$$T\hat{\xi}_T(\hat{\theta}_T) \xrightarrow{d} \chi^2(dim\,\mathcal{D} - dim\,\theta), \tag{3.5}$$

in which the degree of freedom is equal to the difference between the dimension of the design and the dimension of the parameter. This requires a sufficiently large number of auto-covariances, i.e. the order condition $dim\,\mathcal{D} \geq dim\,\theta$.

This asymptotically valid result can be extended to partial identification as follows:

Proposition 2:

i) Consider $\hat{\xi}_T^* = Min_\theta\, \hat{\xi}_T(\theta)$. Then,

$T\hat{\xi}_T^* \xrightarrow{d} \chi^2[dim\,\mathcal{D} + dim\,\Theta_0^* - dim\theta]$, when $T \to \infty$, if $dim\,\mathcal{D} + dim\,\Theta_0^* - dim\theta \geq 0$,

ii) An asymptotically valid confidence set of the implied identified set at level $1-\alpha$ is

$$CS\hat{\Theta}_T^* = \{\theta : T\hat{\xi}_T(\theta) \leq \chi^2_{1-\alpha}(dim\mathcal{D} + dim\Theta_0^* - dim\theta)\},$$

where $\chi^2_{1-\alpha}(.)$ denotes the $(1-\alpha)$-quantile of a chi-square distribution. This confidence set is such that:

$$\lim_{T\to\infty} P_0[CS\hat{\Theta}_T^* \supset \Theta_0^*(f_o)] = 1-\alpha,$$

for any true distribution P_0.

These distributional properties hold when the GCov estimator is used, as it is semi-parametrically efficient. Similar results have been derived for the Generalized Method of Moments (GMM) estimator [3] in Shi, Shum (2015)], but require a two-step estimation method with an inverse of an optimal weighting matrix estimator of the non-centered moments of interest in the second step. Since the number of non-centered moments is in general large, this inversion is computationally cumbersome. This difficulty is circumvented when the GCov estimator is used, as the GCov estimator is based on centered moments and the inversion concerns a matrix of size (L, L) only, instead of size $(dim\mathcal{D}, dim\mathcal{D})$. Similarly, the covariance distances used in Francq, Roy, Zakoian (2005), Davis, Wan (2022) and Davis, Ng (2022) for the identified cases are not optimally weighted. As a consequence the asymptotic distribution of the test statistic can be a non-centered chi-square, not distribution-free and

[3] See, e.g. Lanne, Luoto (2021) for GMM estimation of this type of models in the identified case.

need to be approximated by bootstrap [see for example Davis, Wan (2022)]. Alternative residual dependence measures written in the frequency domain and considering independence beyond pairwise independence have also been introduced in Velasco, Lobato (2018), Velasco (2022), Chu (2023).

3.4 Test of the Pairwise Independence Hypothesis

Proposition 2 i) is the analogue of the generalized Wald statistic introduced in Szroeter (1983) for the Generalized Method of Moments in the identified case. It is still valid for the GCov-based inference and in a framework of partial identification.

This statistic can be used to test the null hypothesis $H_0 = \{\Gamma_0(h) = 0, h = 1, ..., H\}$ versus the alternative $H_1 = \{\text{There exists } h, h = 1, ..., H \text{ with } \Gamma_0(h) \neq 0\}$, where $\Gamma_0(h) = Cov_0(u_t, u_{t-h})$. As H_0 is satisfied under the hypothesis of independence of the u_t's, it can also be used to test the pairwise independence hypothesis. It extends the residual based portmanteau tests introduced for ARMA models to the nonlinear dynamic framework [see, Box, Pierce (1970), Chitturi (1974), Hosking (1980) for linear dynamic models, De Gooijer (2023) for multivariate nonlinear models with martingale difference errors]. The null hypothesis H_0 is asymptotically rejected at 5% if the confidence set of the implied identified set is empty. In such a case, it is interesting to examine closely the components of the portmanteau test statistic $\hat{\xi}_T^*$. For example, in a model with serial dependence, the components are the lagged values: $\hat{\xi}_T^* = \sum_{h=1}^{H} \hat{\xi}_{T,h}^*$. The components $\hat{\xi}_{T,h}^*$, $h = 1, ..., H$ can be reported in a graph as a function of h to help detect which lag(s) and which nonlinear autocovariances cause the rejection of the null hypothesis, by analogy to the graphical analysis of autocorrelation function (ACF).

In the SVAR model with primitive shocks that are mutually independent (see Section 2.3) such covariance restrictions can help identify the model parameters. Then, the associated test can also be considered as a test of some identifying restrictions, which is examined in the next Section.

4 Overidentification

Let us now describe the consequences of increasing the number of autocovariances in the portmanteau measure for a given model and discuss overidentification in partially identified models.

4.1 The effects of overidentification

Let us consider a given model and two nested designs for defining the portmanteau measure, denoted by $\mathcal{D}_1 \subset \mathcal{D}_2$. The associated measures are denoted by $\xi_1(\theta), \xi_2(\theta)$, respectively. The designs can be nested either by increasing the number of transformations a, and/or by increasing the lag H. We get the following result:

Proposition 3:

If $\mathcal{D}_1 \subset \mathcal{D}_2$, then $\xi_1(\theta) \leq \xi_2(\theta)$, $\hat{\xi}_{1T}(\theta) \leq \hat{\xi}_{2T}(\theta)$ and $\hat{\xi}^*_{1T} \leq \hat{\xi}^*_{2T}$.

From Proposition 3 it follows that the implied identified sets are such that $\Theta^*_{01}(f_0) \supset \Theta^*_{02}(f_0)$. Therefore, the increase of the design can have two effects: it either strictly reduces the implied identified set, i.e. the dimension of the implied identified set, or improves the efficiency of the confidence set of the implied identified set, when these latter sets are equal.

This corresponds to the problem of overidentification of the implied identified set.

4.2 The Asymptotic Distributions

The statistical inference with two nested designs is based on the asymptotic distributional properties of the estimated identified sets and of the optimal values of the portmanteau statistics. The following propositions are shown in Appendices 2 and 3.

Proposition 4: Let us assume that parameter θ is identifiable from the two portmanteau optimizations. Then, if $\mathcal{D}_1 \subset \mathcal{D}_2$, we have:

i) $V\hat{\theta}_{1T} >> V\hat{\theta}_{2T}$;

ii) $T(\hat{\theta}_{1T} - \hat{\theta}_{2T})'(V\hat{\theta}_{1T} - V\hat{\theta}_{2T})^+(\hat{\theta}_{1T} - \hat{\theta}_{2T}) \xrightarrow{d} \chi^2[Rk(V\hat{\theta}_{1T} - V\hat{\theta}_{2T})]$, when $T \to \infty$,

where A^+ denotes the generalized inverse of matrix A.

A similar result under partial identification can be written by considering cuts of confidence sets. More precisely, let us consider a scalar function of parameter θ, denoted by $\zeta(\theta)$. Then, for each design $\mathcal{D}_j,\ j = 1, 2$, we can search for an interval $[m\hat{\zeta}_{j,T}, M\hat{\zeta}_{j,T}],\ j = 1, 2$, where:

$m\hat{\zeta}_{j,T} = \min[\zeta(\theta)]$ for $\theta \in CS\hat{\Theta}_{j,T}$,

$M\hat{\zeta}_{j,T} = \max[\zeta(\theta)]$ for $\theta \in CS\hat{\Theta}_{j,T}$.

Then, by Proposition 1, we have:

$$[m\hat{\zeta}_{2,T}, M\hat{\zeta}_{2,T}] \subset [m\hat{\zeta}_{1,T}, M\hat{\zeta}_{1,T}],$$

if $d_1 = d_2$, i.e. the degrees of implied under-identification are equal.

An alternative approach can be based on a comparison of the values of the objective functions. For example, we have the following result:

Proposition 5:

Let us assume $\mathcal{D}_1 \subset \mathcal{D}_2$. Then the difference:

$$T\hat{\xi}^*_{2T} - T\hat{\xi}^*_{1T} = T\hat{\xi}_{2T}(\hat{\theta}_{1T}) - T\hat{\xi}_{1T}(\hat{\theta}_{1T}),$$

follows asymptotically a mixture of chi-square distributions.

In particular, even if $T\hat{\xi}^*_{1T}$ and $T\hat{\xi}^*_{2T}$ are asymptotically chi-square distributed, the statistics $T\hat{\xi}^*_{1T}$ and $T\hat{\xi}^*_{2T} - T\hat{\xi}^*_{1T}$ are not independent and $T\hat{\xi}^*_{2T} - T\hat{\xi}^*_{1T}$ is not chi-square distributed, in general.

However this latter condition is satisfied if $d_1 + dim\,\Theta^*_0 - dim\theta = 0$, since $\xi^*_{1T} = 0$. This condition is satisfied if parameter θ is just identified under design $\mathcal{D}_1$. This explains the interpretation of ξ^*_{2T} as an over-identification test [Szroeter (1983)]. In particular the just identification condition holds for the AR(p) model estimated from the first p autocorrelations as design $\mathcal{D}_1$.

4.3 Estimating the dimension of the implied identified set

Under partial identification, the asymptotic confidence set in Proposition 3 as well as the asymptotic results in Propositions 4 and 5 are not directly applicable, as they depend on the dimension of the implied identified set, which may be unknown. Moreover, we mentioned that it cannot be consistently estimated from the dimension of Θ^*_T that is itself difficult to evaluate in practice. However, under weak regularity conditions, we have:

$$\dim\theta - \dim\Theta^*_0(f_0) = Rk[\frac{\partial\gamma_0(\theta_0)}{\partial\theta'}]. \tag{4.1}$$

Hence, the degree of partial (under) identification is:

$$\dim\Theta^*_0(f_0) = \dim\theta - Rk[\frac{\partial\gamma_0(\theta_0)}{\partial\theta'}]. \tag{4.2}$$

Therefore, it is equivalent to estimate $\dim\Theta^*_0(f_0)$, or $Rk[\frac{\partial\gamma_0(\theta_0)}{\partial\theta'}]$. We know that:

$$Rk[\frac{\partial\gamma_0(\theta_0)}{\partial\theta'}] = Rk[\frac{\partial\gamma_0(\theta_0)'}{\partial\theta}\frac{\partial\gamma_0(\theta_0)}{\partial\theta'}],$$

and that the semi-positive definite matrix $\partial\gamma_0(\theta'_0)/\partial\theta\,\partial\gamma_0(\theta_0)/\partial\theta'$ is consistently estimated by the matrix:

$$T\hat{\Omega}_T = T\partial\hat{\gamma}_T(\hat{\theta}'_T)'/\partial\theta\,\partial\hat{\gamma}_T(\hat{\theta}_T)/\partial\theta', \tag{4.3}$$

where γ, i.e. Γ, is replaced by its sample counterparts $\hat{\gamma}_T$, i.e. $\hat{\Gamma}_T$, and $\hat{\theta}_T$ is a GCov estimator in $\hat{\Theta}_T^*$. Then, we can estimate the rank by performing a spectral decomposition of $\hat{\Omega}_T$, which can be of a large dimension, and counting the number of significant eigenvalues along the lines of [Bai (2003), Fan et al. (2020)]. The analysis of the asymptotic properties of such an approach is out of the scope of the present paper.

An alternative approach can be based directly on the first and second-order derivatives of the objective function $\hat{\xi}_T(\theta)$ at $\hat{\theta}_T$.

4.4 An Asymptotic Linear Interpretation

To illustrate the identification method and to give some insights on the derivation of the asymptotic results (see Appendix 2 for the general proof), let us consider a stationary univariate time series defined by:

$$g(y_t, y_{t-1}; \theta) = u_t,$$

where g is invertible with respect to y_t and the errors u_t are i.i.d.. We do not apply any transformations to u_t and build the portmanteau statistic directly from the scalar variables u_t. Then the portmanteau measure is:

$$\xi(\theta) = \sum_{h=1}^{H} \rho_0^2(h, \theta),$$

where $\rho_0(h, \theta)$ is the true correlation between $g(y_t, y_{t-1}; \theta)$ and $g(y_{t-h}, y_{t-h-1}; \theta)$. Its sample counterpart is:

$$\hat{\xi}_T(\theta) = \sum_{h=1}^{H} \hat{\rho}_T^2(h, \theta).$$

The implied identified set is:

$$\Theta_0^*(f_0) = \{\theta : \theta = Argmin_\theta \sum_{h=1}^{H} \rho_0^2(h, \theta)\}.$$

and a consistent estimator of this set is:

$$\hat{\Theta}_T^* = \{\theta : \theta = Argmin_\theta \sum_{h=1}^{H} \hat{\rho}_T^2(h, \theta)\}.$$

It is important to note that

i) $\xi^* = Min_{\theta \in \Theta} \xi(\theta) = \xi(\theta_0^*)$ for any element θ_0^* in $\Theta_0^*(f_0)$.

and

ii) $\frac{\partial \xi(\theta_0^*)}{\partial \theta}$ is independent of $\theta_0^* \in \Theta_0^*(f_0)$.

A condition slightly stronger than ii) is : the derivatives $\frac{\partial \rho_0(h,\theta_0^*)}{\partial \theta}$, h=1,...,H are independent of $\theta_0^* \in \Theta_0^*(f_0)$.

Let us now consider the selection of $\hat{\theta}_T$ in $\hat{\Theta}_T^*$. Since $\hat{\Theta}_T^*$ tends to $\Theta_0^*(f_0)$, the sequence $\hat{\theta}_T$ tends to the set $\Theta_0^*(f_0)$, but in general not to a given point θ_0^* of this set. However, under the assumption of compact implied identified set, any Cauchy sequence of $\hat{\theta}_T$ converges pointwise as T tends to infinity. For ease of exposition, we consider $\hat{\theta}_T \to \theta_0^*$.

Then, if θ_0^* is in the interior of the implied identified set, we can perform an expansion of the first-order conditions. We have:

$$
\begin{aligned}
&\frac{\partial \hat{\xi}_T(\hat{\theta}_T)}{\partial \theta} = 0 \\
&\iff \quad \sum_{h=1}^{H} \frac{\partial}{\partial \theta}[\hat{\rho}_T^2(h, \hat{\theta}_T)] = 0 \\
&\iff \quad \sqrt{T} \sum_{h=1}^{H} \frac{\partial}{\partial \theta} \hat{\rho}_T^2(h, \theta_0^*) + \sum_{h=1}^{H} \frac{\partial^2}{\partial \theta \partial \theta'} \hat{\rho}_T^2(h, \theta_0^*) \sqrt{T}(\hat{\theta}_T - \theta_0^*) = o_p(1) \\
&\iff \quad \sum_{h=1}^{H} \left[\frac{\partial}{\partial \theta} \hat{\rho}_T(h, \theta_0^*) \sqrt{T} \hat{\rho}_T(h, \theta_0^*) \right] \\
&\qquad + \sum_{h=1}^{H} \left[\hat{\rho}_T(h, \theta_0^*) \frac{\partial^2}{\partial \theta \partial \theta'} \hat{\rho}_T(h, \theta_0^*) + \frac{\partial}{\partial \theta} \hat{\rho}_T(h, \theta_0^*) \frac{\partial}{\partial \theta'} \hat{\rho}_T(h, \theta_0^*) \right] \sqrt{T}(\hat{\theta}_T - \theta_0^*) = o_p(1).
\end{aligned}
$$

We know that $\sqrt{T}\hat{\rho}_T(h, \theta_0^*), h = 1, ..., H$ are the sample autocorrelations of the true white noise u_t. Therefore they are asymptotically independent standard Gaussian variables. Asymptotically, the estimator $\hat{\rho}_T$ can be replaced by ρ_0 in all other terms. Since $\rho_0(h, \theta_0^*) = 0$, we get:

$$
\sum_{h=1}^{H} \frac{\partial}{\partial \theta} \rho_0(h, \theta_0^*) [\sqrt{T} \hat{\rho}_T(h, \theta_0^*)] \approx \sum_{h=1}^{H} \left[\frac{\partial}{\partial \theta} \rho_0(h, \theta_0^*) \frac{\partial}{\partial \theta'} \rho_0(h, \theta_0^*) \right] \sqrt{T}(\hat{\theta}_T - \theta_0^*).
$$

Let us introduce $z_h = \frac{\partial}{\partial \theta} \rho_0(h, \theta_0^*)$, $v_h = -\sqrt{T} \hat{\rho}_T(h, \theta_0^*)$, $h = 1, ..., H$.

The above system can be equivalently written as:

$$
\sum_{h=1}^{H} z_h v_h \approx \left(\sum_{h=1}^{H} z_h z_h' \right) \sqrt{T}(\hat{\theta}_T - \theta_0^*),
$$

where z_h and the distribution of v_h do not depend on θ_0^* and on the selected sequence $\hat{\theta}_T$. Two cases can be distinguished:

i) If $Z'Z = \sum_{h=1}^H z_h z_h'$ is of full column rank, $RkZ'Z = \dim\theta$, then:

$$\sqrt{T}(\hat{\theta}_T - \theta_0^*) \approx (Z'Z)^{-1}Z'v,$$

can be seen as an OLS estimator obtained by regressing the independent Gaussian variables $-\sqrt{T}\hat{\rho}_T(h,\theta_0^*)$ on the explanatory variable $z_h = \frac{\partial}{\partial\theta}\rho_0(h,\theta_0^*)$.

ii) If $Z'Z$ is not invertible, $RkZ'Z < \dim\theta$, there is an identification issue with the degree of under-identification equal to $\dim\theta - Rk(Z'Z)$.

The above expansion shows that asymptotically in a neighbourhood of θ_0^* the manifold $\hat{\Theta}_T^*$ can be locally replaced by a linear subspace and the asymptotic results be based on this asymptotic linear interpretation (see Appendix 2 for this interpretation in the multivariate framework).

Let us now focus on the minimum value of the objective function, that is $\hat{\xi}_T^* = T\hat{\xi}_T(\hat{\theta}_T)$. We have:

$$\begin{aligned} T\hat{\xi}_T(\hat{\theta}_T) &\approx T\hat{\xi}_T(\theta_0) + \sqrt{T}\frac{\partial\hat{\xi}_T(\theta_0^*)}{\partial\theta'}\sqrt{T}(\hat{\theta}_T - \theta_0^*) \\ &= v'v - v'Z(Z'Z)^{+}Z'v \\ &= v'(Id - Z(Z'Z)^{+}Z')v, \end{aligned}$$

where $(.)^+$ denotes a generalized inverse. It follows that asymptotically $T\hat{\xi}_T(\hat{\theta}_T)$ has a chi-square distribution with the degree of freedom $H + \dim\Theta_0^*(f_0) - \dim\theta = H - Rk(Z'Z)$ for any choice of the sequence $\hat{\theta}_T$. The expression of the confidence set at level $1-\alpha$ of the implied identified set is obtained by inverting the test:

$$CS\hat{\Theta}_T^* = \{\theta : T\hat{\xi}_T(\theta) \leq \chi^2_{1-\alpha}(H - Rk(Z'Z)\}.$$

When H increases, $Z'Z$ increases (for the Loewner ordering on symmetric matrices [4]) as well as the statistic $T\hat{\xi}_T$ and the rank of $Z'Z$. It will be useful to evaluate this rank by performing for each H a spectral decomposition of $\hat{Z}'\hat{Z}$ with $\hat{z}_h = \frac{\partial}{\partial\theta}\hat{\rho}_T(h,\hat{\theta}_T),\ h = 1,...,H$ and estimating the rank from the number of significant eigenvalues of $\hat{Z}'\hat{Z}$.

This number of significant eigenvalues increases with H and is always less than $\dim\theta$. If it is equal to $\dim\theta$ for a large H, then θ is identifiable.

[4] For two symmetric matrices Σ, Σ^* of dimension $(n\times n)$, this ordering is defined by : $\Sigma \gg \Sigma^*$, if and only if $u'\Sigma u \geq u'\Sigma^* u$, for any $u \in I\!R^n$.

5 Illustration

Under partial identification, the dimension of the identified set is often rather large and the estimated confidence set is difficult to visualize. In order to provide interpretable figures we consider in this section illustrative applications, where the degree of under-identification is 0, 1, or 2.

5.1 Independent Component Analysis for Dimension 2

Let us consider the independent component analysis, discussed in Section 2.1, for dimension $K = 2$. We assume that both the errors and observations Y_i have mean 0 and an identity variance-covariance matrix. It is known that matrix C is not identifiable in general, but $C'C$ is identifiable. In fact, C is identifiable up to an orthonormal transformation [5]. Then, we can focus on the case:

$$QY_i = u_i, \; i = 1, ..., n, \iff Y_i = Q'u_i, \; i = 1, ..., n, \tag{4.4}$$

where matrix Q satisfies the condition $QQ' = Id$. In this framework, there are four parameters corresponding to the elements of matrix Q under three independent quadratic restrictions. Therefore the degree of under-identification is either 0, or 1. It is equal to 1 if the components of the error vectors follow Gaussian distributions. It is equal to 0, if at least one component is non-Gaussian. In the latter case, the dimension of the implied identified set can still be 1, if the moments introduced in the portmanteau statistic are not sufficiently informative (see Section 2.1). Due to these nonlinear restrictions on the parameter Q, the parametrization by an orthonormal matrix is not used [see Granziera, Moon, Schorfheide (2018) for such a parametrization in a SVAR model]. Alternative parametrizations by a skew-symmetric matrix:

$$B = \begin{pmatrix} 0 & -b \\ b & 0 \end{pmatrix}$$

with unconstrained parameter b have been introduced in the recent literature. These are:

i) the exponential transform: $Q = exp(B) = \sum_{j=0}^{\infty} \frac{B^j}{j!}$ [Magnus, Pijls, Sentana (2018)];

ii) the Cayley transform:

$$Q = (Id - B)^{-1}(Id + B) = \frac{1}{1 + b^2} \begin{pmatrix} 1 - b^2 & -2b \\ 2b & 1 - b^2 \end{pmatrix},$$

[5] and also up to signed scaling and permutation of rows. The other multivariate scaling effects are usually solved by pre-whitening the observations.

[Gourieroux, Jasiak (2023)]. We choose the Cayley transform in our illustration to avoid the numerical approximations in the exponential transform.

In the bivariate framework, any orthonormal matrix is a rotation matrix corresponding to an angle α, $\alpha \in [-\pi, \pi]$. Then, parameter b of the skew-symmetric matrix in the Cayley transform is $b = tan(\alpha/2)$, $\alpha \in (-\infty, \infty)$. Under the independence hypothesis, that matrix is identifiable up to a sign effect and permutation of u_1, u_2. Thus, there are still global identification issues that are solved by constraining α to $(0, \pi/2)$, or equivalently b to $(0, 1)$. Thus, our search for the confidence sets is carried out under the restriction $0 \leq b \leq 1$.

5.2 The Monte-Carlo Study

We consider different Data Generating Processes (DGP) by choosing:

a) the true value of the parameter: $b_0 = 0.2, 0.5$;

b) the true error distribution:

scheme 1: $u_1, u_2 \sim N(0, 1)$;

scheme 2: $u_1 \sim N(0, 1)$, $u_2 \sim U_{[-1,1]}$, rescaled to variance 1;

scheme 3: $u_1 \sim t(5)$, $u_2 \sim U_{[-1,1]}$, rescaled to variance 1;

c) the number of observations $n = 100, 200, 500$.

Next, we apply the GCov-based inference with the following types of nonlinear transformations:

portmanteau statistic 2 (st2) based on $Cov(u_1, u_2), Cov(u_1^2, u_2^2), Cov(u_1^2, u_2), Cov(u_1, u_2^2)$ with $d = 4$;

portmanteau statistic 3 (st3) based on $Cov\left[\begin{pmatrix} u_1 \\ u_1^2 \\ u_1^3 \end{pmatrix}, \begin{pmatrix} u_2 \\ u_2^2 \\ u_2^3 \end{pmatrix}\right]$ with $d = 9$.

For each simulated path, we compute the optimal value of the portmanteau statistic [6]and the estimated confidence set of the identified set at $1 - \alpha = 0.95$. For each case, we compute it with the degree of freedom d, equal to 4 and 9 and $d - 1$ equal to 3 and 8. The results are given in Tables 2-4.

Table 2: Confidence Set of Implied Identified Set: Gaussian Sources

[6]It is possible to apply standard optimization algorithms with a given number p of iterations to get values $\hat{b}_T(p)$. However, due to the partial identification issue, this value $\hat{b}_T(p)$ will not converge numerically, in general, when p tends to infinity.

$u_1, u_2 \sim N(0,1)$, T=100		
stat	b=0.2	b=0.5
st2 DF=3	(0,1))	(0,1)
st2 DF=4	(0,1)	(0,1)
st3 DF=8	(0,1)	(0,1)
st3 DF=9	(0,1)	(0,1)
$u_1, u_2 \sim N(0,1)$, T=200		
stat	b=0.2	b=0.5
st2 DF=3	(0,1)	(0,1)
st2 DF=4	(0,1)	(0,1)
st3 DF=8	(0,1)	[0.15, 0.99]
st3 DF=9	(0,1)	[0.14, 1)
$u_1, u_2 \sim N(0,1)$, T=500		
stat	b=0.2	b=0.5
st2 DF=3	(0,1)	(0,1)
st2 DF=4	(0,1)	(0,1)
st3 DF=8	(0,1)	(0,1)
st3 DF=9	(0,1)	(0,1)

Table 3: Confidence Set of Implied Identified Set: One Gaussian Source

$u_1 \sim N(0,1), u_2 \sim U[-1,1]$, T=100		
stat	b=0.2	b=0.5
st2 DF=3	(0,1)	(0, 1)
st2 DF=4	(0,1)	(0, 1)
st3 DF=8	[0.18, 0.22] ∪ [0.38, 0.90]	(0, 0.56] ∪ [0.87,1)
st3 DF=9	[0.18, 0.23] ∪ [0.37, 0.90]	(0, 0.56] ∪ [0.87,1)
$u_1 \sim N(0,1) u_2 \sim U[-1,1]$, T=200		
stat	b=0.2	b=0.5
st2 DF=3	[0.10, 1)	[0.19, 0.70]
st2 DF=4	[0.10, 1)	[0.15, 0.77]
st3 DF=8	[0.08, 0.33]	[0.24, 0.74]
st3 DF=9	[0.08, 0.35]	[0.22, 0.76]
$u_1 \sim N(0,1), u_2 \sim U[-1,1]$, T=500		
stat	b=0.2	b=0.5
st2 DF=3	[0.07, 0.44] ∪ [0.83, 1)	[0.13, 0.26] ∪ [0.62, 0.75]
st2 DF=4	[0.07, 0.46] ∪ [0.80, 1)	[0.10, 0.26] ∪ [0.62, 0.79]
st3 DF=8	(0, 0.03] ∪ [0.12, 0.19] ∪ [0.41, 0.45] ∪ [0.90,1)	[0.36, 0.44] ∪ [0.66, 0.72]
st3 DF=9	(0, 0.04] ∪ [0.12, 0.19] ∪ [0.41, 0.45] ∪ [0.88,1)	[0.34, 0.44] ∪ [0.66, 0.73]

Table 4: Confidence Set of Implied Identified Set: No Gaussian Sources

$u_1, u_2 \sim U[-1,1]$, T=100		
stat	b=0.2	b=0.5
st2 DF=3	[0.01, 1)	[0.36, 0.67]
st2 DF=4	[0.01, 1)	[0.32, 0.70]
st3 DF=8	[0.1, 0.29] ∪ [0.65, 0.79]	[0.42, 0.65]
st3 DF=9	[0.12, 0.28]	[0.40, 0.66]
$u_1, u_2 \sim U[-1,1]$, T=200		
stat	b=0.2	b=0.5
st2 DF=3	[0.11, 0.29]	[0.33, 0.77]
st2 DF=4	[0.08, 0.31]	[0.32, 0.79]
st3 DF=8	[0.13, 0.23]	[0.46, 0.61]
st3 DF=9	[0.12, 0.24]	[0.45, 0.61]
$u_1, u_2 \sim U[-1,1]$, T=500		
stat	b=0.2	b=0.5
st2 DF=3	[0.10, 0.32]	[0.37, 0.62]
st2 DF=4	[0.09, 0.33]	[0.37, 0.64]
st3 DF=8	[0.18, 0.23]	[0.46, 0.54]
st3 DF=9	[0.18, 0.23]	[0.46, 0.54]

The estimated confidence sets are intervals of the values of b. This corresponds to manifolds of dimension 1 even if b is identifiable. This shows that the dimension of Θ_0^* cannot be consistently estimated from the dimension of $\hat{\Theta}_T^*$.

When the two sources are Gaussian (Table 2), the confidence set is large and it is often equal to the maximum length interval [0,1]. When one source is Gaussian and the second one is uniformly distributed (Table 3), there is still an implied identification issue for statistic st2 (see, Table 1). It is partly solved when more moments are included and the "length" of the interval decreases when the number of observations increases. Note that the confidence sets contain the true value of parameter b.

Similar remarks can be done in the case of two non-Gaussian sources (Table 4).

We have not yet estimated the true degree of freedom of the asymptotic chi-square distribution. Let us now perform the spectral decomposition of the $\hat{\Omega}_T$ matrix defined in (4.3). An example of such spectral decomposition is given below for the case of one Gaussian source corresponding to Table 3, $b = 0.2, b = 0.5, T = 100, 200, 500$. In Tables 5 and 6, the eigenvalues of the spectral decomposition are arranged in an increasing order. In all cases one eigenvalue is much bigger that the other ones, which indicates that parameter b can possibly be identified.

Table 5: Eigenvalues

Statistic 2					
b=0.2			b=0.5		
T=100	T=200	T=500	T=100	T=200	T=500
7.1054274e-15	0.000000	0.0000000	0.0000000	0.0000000	0.0000000
84.477243	6.0987575	11.968264	171.48689	157.48586	74.523249
Statistic 3					
-6.5342252e-15	-1.1065710e-14	-3.5527137e-15	4.0993461e-15	6.4125221e-14	-4.5474735e-13
0.0000000	-2.1233290e-16	0.0000000	0.0000000	0.0000000	-7.1331829e-15
6.8870545e-15	0.0000000	0.0000000	2.5665165e-15	0.0000000	-7.1054274e-15
1.2754486e-13	4.5028131e-15	0.0000000	1.2897513e-14	1.7869588e-15	-3.1671026e-26
9.0949470e-13	1.2046207e-13	0.0000000	3.6121148e-14	2.2737368e-13	3.1671257e-26
6787.1522	1245.6222	2674.8814	1014.3788	1327.6723	2308.3416

6 Illustration

6.1 The Cayley transform

When the dimension is 3, the orthonormal matrix Q with eigenvalues different from 1 is obtained from the rotations of a skew-symmetric matrix around an axis. The vector (x, y, z) is a unit axis of rotation scaled by $tan(\theta/2)$, where θ is the angle. The rotations with angle π are excluded. More precisely, the orthonormal matrix is given by:

$$Q = \begin{bmatrix} w^2 + x^2 - y^2 - z^2 & 2(xy - wz) & 2(wy + xz) \\ 2(xy + wz) & w^2 - x^2 + y^2 - z^2 & 2(yz - wx) \\ 2(xz - wy) & 2(wx + yz) & w^2 - x^2 - y^2 + z^2 \end{bmatrix},$$

where $w^2 + x^2 + y^2 + z^2 = 1$. We use this normalization to be able to easily derive the conditions on x, y, z to solve the remaining up to permutation identification. Under this form, the permutations on the components of the sources are equivalent to permutations between x and y, x and z, y and z [7]. This allows us for $x < y < z$, say. Moreover, the sources are defined up to their sign. The last identification issue is solved by considering the positivity constraints:

$$0 \leq x \leq y \leq z, \quad x^2 + y^2 + z^2 \leq 1, \quad \text{with} \quad w = \sqrt{1 - x^2 - y^2 - z^2} \tag{4.5}$$

that will be taken into account in the optimization with respect to x, y, z [8].

6.2 Estimation results

We apply the GCov-based inference methods to a trivariate series of daily financial returns. The series of 750 daily returns on Tesla, Meta and Netflix are recorded over the period of 2019/01/02 until 2021/12/31 [9]. The series are displayed in Figure 1.

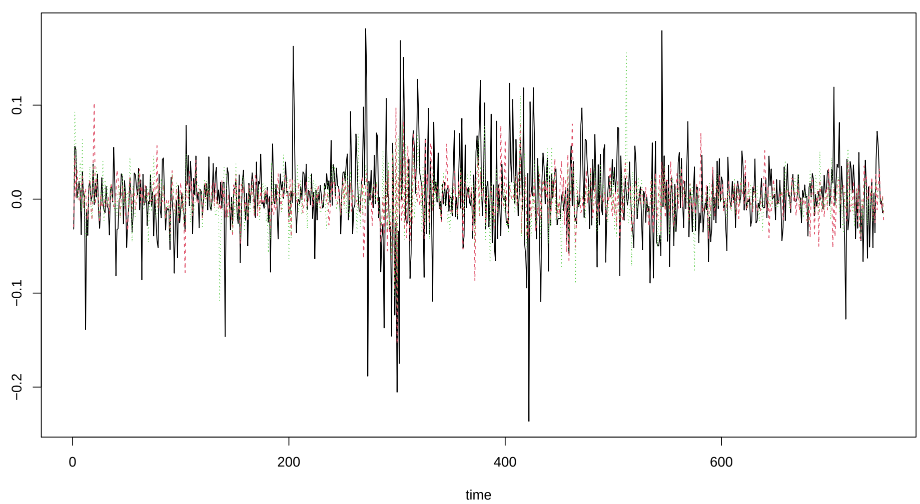

Figure 1: Daily Returns on Tesla, Meta and Netflix
black: Tesla, red: Meta, green: Netflix

[7] This is not the case for the normalization initially proposed by Cayley, in which the matrix is divided by $w^2 + x^2 + y^2 + z^2$ and then normalized with $w = 1$.

[8] This solves the identification on the interior of the set of x, y, z. There still exist remaining identification issues on the boundaries of the set. For example, it is easy to check that two permutation matrices are obtained for $x = y = z = 0$ and for $x = y = z = 0.5$.

[9] The returns are computed as the log differences of daily adjusted prices from Yahoo Finance Canada

The ICA is applied to a series of residuals obtained after pre-whitening of the return series. Our approach can be compared with the pre-whitening in macroeconomics, where the ICA is applied to the residuals of an estimated (structural) VAR model:

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

where C is a low-triangular matrix (Cholesky decomposition matrix) and Q is an orthogonal matrix. In a 3-dimensional VAR model, the number of the parameters characterizing the cross-sectional non-linear dependence is 9, with 6 parameters in matrix C and 3 parameters in matrix Q. This total number of parameters is less that the number of parameters in the variance-covariance matrix of the residuals $\Sigma = V(Y_t - \mu - \Phi Y_{t-1})$ which is 6, implying an identification problem. Hence, in macroeconomics, the ICA is applied to the residuals, which are linearly pre-whitened by the VAR model.

Our application shows that the ICA results depend on the quality of the pre-whitening, and a linear pre-whitening may not suffice. In order to highlight this effect, we proceed in two steps.

a) Linear pre-whitening

The ACF analysis shows that the returns on Tesla are serially uncorrelated, while the returns on Meta and Netflix have small but statistically significant autocorrelations at lag 1. Therefore, these two return series are filtered by autoregressions. Next, all three return series are transformed into a multivariate series with an identity variance-covariance matrix by using the Cholesky decomposition. This is a linear pre-whitening, routinely used in macroeconomics and finance.

The GCov-based method involving the autocovariances of the pre-whitened series and their squares for a total number of 12 cross-moments results in an empty confidence set of the identified set for matrix Q, i.e. parameters x, y, z. This implies that the hypothesis of the existence of independent sources which are strong white noises is strongly rejected. More precisely, the hypothesis is rejected because either the strong white noise condition, or the independence of sources condition does not hold.

b) VAR-ARCH pre-whitening

As the financial returns display time-varying volatility, we expect the rejection to be caused by the conditional heteroscedasticity, i.e. by the autocorrelation of squared returns. We estimate a VAR-ARCH model, and then transform the residuals by using a Choleski decomposition, in order to obtain standardized residuals with an identity variance-covariance matrix.

Figure 2 displays the 3-dimensional scatterplot describing the confidence set of the implied identified set, obtained on a grid with width 0.01. This scatterplot is the analogue of intervals and unions of intervals reported in Tables 2 and 3 for a univariate parameter. Figure 2 shows two subsets, suggesting a global identification issue. The pattern of the upper subset is interesting, as it is close to a manifold of dimension 2. It shows that the degree of (implied) under-identification of the orthogonal matrix Q is equal to one. The visual analysis of the confidence set provides information on the possibility to identify before the rank analysis of Section 4.2 is applied.

Note that this result is obtained with a standard VAR-ARCH model, without an adjustment for time varying risk premia (i.e. ARCH-in-mean), or for time-varying leverage effect. Nevertheless, the leverage effect is accounted for by the portmanteau statistic, which contains the autocovariances of the series and its squared values.

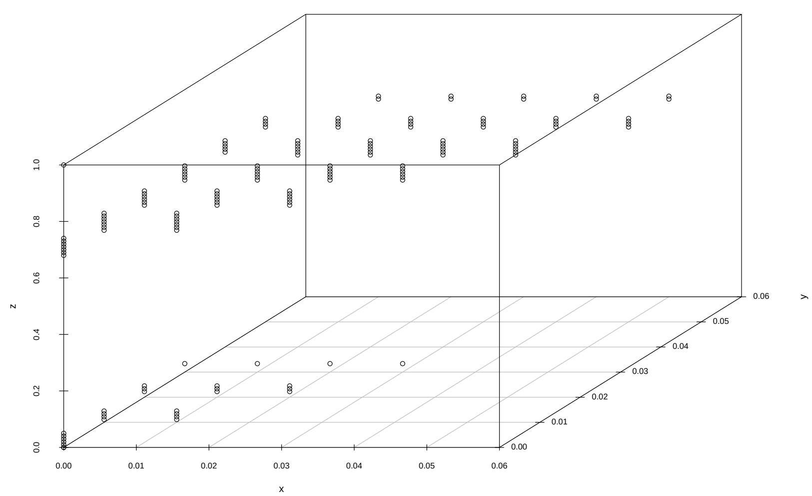

Figure 2: Confidence Set of Tesla, Meta and Netflix

7 Concluding Remarks

Independence restrictions are often used in static or dynamic models either directly for computing nonlinear impulse response functions, accommodating heavy tails of primitive shocks, or indirectly when deriving the asymptotic results. In all these cases, the restrictions provide information on the parameters of interest and facilitate its identification.

The independence restrictions can be written in terms of zero covariance restrictions on the transformed series and summarized by a portmanteau-type statistics. In this paper, we explained how the residual-based portmanteau statistics with the GCov parameter estimator can be used to define asymptotic confidence sets of the implied identified set with the exact asymptotic level and to test the pairwise independence hypotheses. We have also developed an approach for determining the degree of under-identification based on choosing a sequence of different designs for the portmanteau statistics. This is illustrated by examples of independent component analysis and an application to financial returns.

The application highlights the importance of selecting an appropriate pre-whitening approach to eliminate nonlinear serial dependence before applying the independent component analysis (ICA) approach to reveal independent latent sources. When the dimension n is strictly larger than 3, a partial visualization of the confidence set for Q can be obtained by considering cuts and/or projections of this set on spaces of dimension up to 3. The uncertainty on the orthonormal transformation has to be taken into account when functions depending on Q, such as the impulse response functions, are evaluated.

REFERENCES

Andrews, D., Berry, S., and P., Jia (2004): "Confidence Regions for Parameters in Discrete Games and Multiple Equilibria with an Application to Discount Chain Store Locations", D.P. Yale University.

Anderson, T. (1999): "Asymptotic Theory for Canonical Correlation Analysis", Journal of Multivariate Analysis, 70, 1-29.

Anderson, T. (2002): "Canonical Correlation Analysis and Reduced Rank Regression in Autoregressive Models", Annals of Statistics, 30, 1134-1154.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dimensions", Econometrica, 71, 135-171.

Black, F. (1976): "Studies of Stock Price Volatility Changes", Proceedings of the Business and Economic Section of American Statistical Association, 177-181.

Borkovec, M., and C., Kluppelberg (2001): "The Tail of the Stationary Distribution of an Autoregressive Process with ARCH(1) Errors", Annals of Applied Probability, 11, 1220-1241.

Box, G., and D., Pierce (1970): "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", JASA, 65, 1509-1526.

Canay, I., and A., Shaikh (2017): "Practical and Theoretical Advances in Inference for Partially Identified Models", Chapter 9 in Advances in Economics and Econometrics, Eleventh World Congress, ed. Honore, B., Pakes, A., Piazzezi, M., and I., Samuelson, Cambridge University Press.

Chernozhukov, V., Hong, H., and E., Tamer (2007): 'Estimation and Confidence Regions for Parameter Sets in Econometric Models", Econometrica, 75, 1243-1284.

Chitturi, R. (1974): "Distribution of Residual Autocorrelations in Multiple Autoregressive Schemes", JASA, 69, 928-934.

Chitturi, R. (1976): "Distribution of Multivariate White Noise Autocorrelations", JASA, 71, 223-226.

Chu, B. (2023): "A Distance Based Test of Independence Between Multivariate Time Series", Journal of Multivariate Analysis, 195, 105-151.

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

Davis, R., and S., Ng (2022): "Time Series Estimation of the Dynamic Effects of Disaster Type Shocks", D.P. Columbia University.

Davis, R., and P., Wan (2022): "Goodness of Fit Testing for Time Series Models with Distance Covariance", Journal of Econometrics, 227, 4-24.

De Gooijer, J. (2023): "On Portmanteau Type Tests for Nonlinear Multivariate Time Series", Journal of Multivariate Analysis, 195, 105-157.

Engle, R. ,Lilien, D., and R., Robbins (1987): "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Models", Econometrica, 55, 391-407.

Eriksson, J., and V., Koivunen (2004): "Identifiability, Separability and Uniqueness of Linear ICA Models", IEEE Signal Processing Letters, 11, 601-604.

Fan, J., Guo, J., and S., Zheng (2020): "Estimating Number of Factors by Adjusted Eigenvalue Thresholding", JASA, 117, 1-10.

Francq, C., Roy, R., and J.M., Zakoian (2005): "Diagnostic Checking in ARMA Models with Uncorrelated Errors", JASA, 100, 532-544.

Gourieroux, C., and J., Jasiak (2017): "Noncausal Vector Autoregressive Process: Representation, Identification and Semi-Parametric Estimation", Journal of Econometrics, 200, 118-134.

Gourieroux, C., and J., Jasiak (2022): "Generalized Covariance Estimator", Journal of Business and Economic Statistics, published online.

Gourieroux, C., and J., Jasiak (2023): "Statistical Inference for SVAR Model with Sign Restrictions", ArXiv.

Gourieroux, C., Monfort, A., and J.P., Renne (2017): "Statistical Inference for Independent Component Analysis: Application to Structural VAR Models", Journal of Econometrics, 196, 11-126.

Gourieroux, C., and J.M., Zakoian (2017): "Local Explosion Modelling by Noncausal Process" Journal of the Royal Statistical Society (JRSS), Series B, 79, 737-756.

Granziera, E., Moon, N., and F., Schorfheide (2018): "Inference for VAR's Identified with Sign Restrictions", Quantitative Economics, 9, 1087-1121.

Guay, A. (2021): "Identification of Structural Vector Autoregressions Through Higher Unconditional Moments", Journal of Econometrics, 225, 27-46.

Hosking, J. (1980): "The Multivariate Portmanteau Statistic", JASA, 75, 602-608.

Lanne, M., and J., Luoto (2021): "GMM Estimation of Non-Gaussian Structural Vector Autoregressions", Journal of Business and Economic Statistics", 39, 69-81.

Lanne, M., Meitz, M., and P., Saikkonen (2017): "Identification and Estimation of Non-Gaussian Structural Autoregressions", Journal of Econometrics, 196, 288-304.

Ling, S. (2007): "A Double AR(p) Model: Structure and Estimation", Statistica Sinica, 17,161-175.

Magnus, J., Pijls, H., and E., Sentana (2021): "The Jacobian of the Exponential Function", Journal of Economic Dynamics and Control", 127, 104-122.

Moon, H., and F., Schorfheide (2012): "Bayesian and Frequentist Inference in Partially Identified Models", Econometrica, 80, 755-782.

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

Shi, X., and M., Schum (2015): "Simple Two-Stage Inference for a Class of Partially Identified Models", Econometric Theory, 31, 493-520.

Szroeter, J. (1983): "Generalized Wald Methods for Testing Nonlinear Implicit and Overidentifying Restrictions", Econometrica, 51, 335-353.

Velasco, C. (2022): "Estimation of Time Series Models Using Residuals Dependence Measures", Annals of Statistics, 50, 3039-3063.

Velasco, C., and C., Lobato (2018): "Frequency Domain Minimum Distance Inference for Possible Noninvertible and Noncausal ARMA Models", Annals of Statistics, 46, 555-579.

Appendix 1

The Asymptotic Linear Approximations

Section 4.4 shows that for a univariate process the asymptotic results can be derived from an asymptotic virtual linear model. This possibility exists also in the multivariate framework [Gourieroux, Jasiak (2022), supplementary material].

i) The fact that the estimated $\sqrt{T}\hat{\rho}_T(h;\theta_0),\ h = 1, .., H$ are independent standard normal variables under the independence hypothesis in Section 4.4 has the following multivariate extension [Chitturi (1976)]:

Under the independence of the error terms, the estimated autocovariances $\sqrt{T}\hat{\Gamma}_T(h;\theta_0),\ h = 1, ..., H$ are asymptotically independent with identical distributions:

$$vec[\sqrt{T}\hat{\Gamma}_T(h;\theta_0)] \xrightarrow{d} N[0, \Gamma(0,\theta_0) \otimes \Gamma(0,\theta_0)], \tag{a.1}$$

where $\otimes$ is the Kronecker product.

ii) Then, the asymptotic quadratic expansion of the portmanteau dependence measure around a true value leads to solutions satisfying the limiting first-order conditions [see, Gourieroux, Jasiak (2022), eq. (a.8)-(a.9) in the supplemental material].

$$\begin{aligned} &\sum_{h=1}^{H} \left\{ \frac{\partial vec\,\Gamma(h;\theta_0)'}{\partial\theta} [\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}] \frac{\partial vec\,\Gamma(h;\theta_0)}{\partial\theta'} \right\} \sqrt{T}(\hat{\theta}_T - \theta_0) \\ \approx\ & \sum_{h=1}^{H} \left\{ \frac{\partial vec\,\Gamma(h;\theta_0)'}{\partial\theta} [\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}] vec\,[\sqrt{T}\hat{\Gamma}_T(h;\theta_0)] \right\}. \end{aligned} \tag{a.2}$$

Thus, asymptotically the elements of the estimated implied identified set are the least squares solutions in a virtual linear model:

$$v_h = z_h'\sqrt{T}(\theta - \theta_0) + w_h,\ h = 1, ..., H, \tag{a.3}$$

where $v_h = vec\,[\sqrt{T}\hat{\Gamma}_T(h;\theta_0)],\ z_h = \frac{\partial vec\,\Gamma(h;\theta_0)'}{\partial\theta}$ and the w_h are independent Gaussian vectors with mean zero and variance-covariance matrix $\Gamma(0;\theta_0) \otimes \Gamma(0;\theta_0)$.

The associated virtual matrices are:

$$V = \begin{pmatrix} v_1 \\ \vdots \\ v_H \end{pmatrix}, Z = \begin{pmatrix} z_1' \\ \vdots \\ z_H' \end{pmatrix}.$$

In general, the virtual matrix Z is not of full rank, which causes a problem of partial identification. More precisely, there exists an element $\hat{\theta}_T^*$ of $\hat{\Theta}_T^*$ that is asymptotically such that:

$$\sqrt{T}(\hat{\theta}_T^* - \theta_0) = [Z' diag[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}]Z]^+ Z' diag[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}]V, \quad \text{(a.4)}$$

where $\Sigma^{-1} \equiv diag[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}]$ is a diagonal matrix with identical blocks and $^+$ denotes the pseudo-inverse of a matrix.

Then, all other elements of $\sqrt{T}(\hat{\Theta}_T^* - \theta_0)$ are deduced by adding to this element, that plays the role of an origin, the null space of $Z' diag[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}]Z$. In other words asymptotically $\sqrt{T}(\hat{\Theta}_T^* - \theta_0)$ is a linear (stochastic) manifold.

We can also replace the asymptotic GLS expression (a.4) by an OLS expression, by replacing Z by $\tilde{Z}$, V by $\tilde{V}$, with $\tilde{z}_h = z_h[\Gamma(0;\theta_0)^{-1/2} \otimes \Gamma(0;\theta_0)^{-1/2}]$, $\tilde{v}_h = [\Gamma(0;\theta_0)^{-1/2} \otimes \Gamma(0;\theta_0)^{-1/2}]v_h$.

Appendix 2

Proof of Proposition 4

i) By using the notation of Appendix 1, we have:

$$V\hat{\theta}_{1T} = [Z_1'\Sigma_1^{-1}Z_1]^{-1}, \ \ V\hat{\theta}_{2T} = [Z_2'\Sigma_2^{-1}Z_2]^{-1},$$

where we can decompose Z_2, Σ_2 into:

$$Z_2 = \begin{pmatrix} Z_1 \\ \tilde{Z}_2 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Then, it follows from the block inversion of matrix Σ_2 that:

$$\begin{aligned} & Z_2'\Sigma_2^{-1}Z_2 - Z_1'\Sigma_1^{-1}Z_1 \\ = \ & (\tilde{Z}_2 - \Sigma_{21}\Sigma_1^{-1}Z_1)'(\Sigma_{22} - \Sigma_{21}\Sigma_1^{-1}\Sigma_{12})^{-1}(\tilde{Z}_2 - \Sigma_{21}\Sigma_1^{-1}Z_1) >> 0. \end{aligned} \quad \text{(a.5)}$$

The result follows.

ii) The result follows from the expansion (a.4) applied to the two designs and re-applying the above block-inversion.

Appendix 3

Proof of Proposition 5

Let us assume designs $\mathcal{D}_1 \subset \mathcal{D}_2$. Then, we have:

$$\begin{aligned} \sqrt{T}\hat{\xi}^*_{1T} &\sim \tilde{V}_1'[Id_1 - \tilde{Z}_1(\tilde{Z}_1'\tilde{Z}_1)^{-1}\tilde{Z}_1']\tilde{V}_1. && \text{(a.6)} \\ \sqrt{T}\hat{\xi}^*_{2T} &\sim \tilde{V}_2'[Id_2 - \tilde{Z}_2(\tilde{Z}_2'\tilde{Z}_2)^{-1}\tilde{Z}_2']\tilde{V}_2. && \text{(a.7)} \end{aligned}$$

There exist various ways of nesting the designs either by increasing the number H of lags, or by increasing the number of transformations a.

Let us consider the first case, where the set of transformations is the same for $\mathcal{D}_1$ and $\mathcal{D}_2$, but $H_1 < H_2$. When moving from $\mathcal{D}_1$ to $\mathcal{D}_2$, this is as if we increase the number of observations in the virtual model (a.3) by introducing new indexes $h,\ h = H_1 + 1, ..., H_2$. Therefore, we can write: $\tilde{V}_1 = (Id, 0)\tilde{V}_2$. It follows that:

$$\sqrt{T}\hat{\xi}^*_{1T} = \tilde{V}_2 \begin{pmatrix} Id_1 \\ 0 \end{pmatrix} [Id_1 - \tilde{Z}_1(\tilde{Z}_1'\tilde{Z}_1)^{-1}\tilde{Z}_1'](Id_1, 0)\tilde{V}_2, \qquad (a.8)$$

and similarly, we can write the difference between the optimal values in the objective functions as:

$$\sqrt{T}\hat{\xi}^*_{2T} - \sqrt{T}\hat{\xi}^*_{1T} = \tilde{V}_2'\Omega\tilde{V}_2,$$

where Ω is a symmetric matrix and $\tilde{V}_2$ is a standard Gaussian vector. It is known that such a quadratic form follows a chi-square distribution if and only if $\Omega^2 = \Omega$, that is if Ω is an orthogonal projector, and follows a mixture of chi-square distributions, otherwise.

In our framework it is easy to see that Ω does not satisfy the condition $\Omega^2 = \Omega$. Note however, that asymptotically $\sqrt{T}\hat{\xi}^*_{1T}$, $\sqrt{T}\hat{\xi}^*_{2T}$ are quadratic forms of the same Gaussian vector and their joint distribution can be easily approximated by simulations after replacing Z by their estimates.

A similar result is obtained when considering for example the difference $\sqrt{T}\hat{\xi}_{2T}(\hat{\theta}_{1T}) - \sqrt{T}\hat{\xi}_{1T}(\hat{\theta}_{1T})$ when θ is identified with design $\mathcal{D}_1$.

On-Line Appendix 1

Closed-Form Expressions of Power Covariances

Let us consider the framework of ICA with dimension 2 and the observations given by:

$$Y = Q'u,$$

where the rotation matrix is parametrized by the angle θ:

$$Q' = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Then, we have

$$Y_1 = u_1 \cos\theta - u_2 \sin\theta, \;\; Y_2 = u_1 \sin\theta + u_2 \cos\theta.$$

We assume that the sources u_1, u_2 are independent, with means zero $Eu_1 = Eu_2 = 0$ and unit variances: $E(u_1^2) = E(u_2^2) = 1$. Then it is possible to derive the closed-form expressions of the power covariances up to power 3. We have:

$$\begin{aligned}
Cov(Y_1, Y_2) &= 0, \\
Cov(Y_1^2, Y_2) &= \cos^2\theta \sin\theta E(u_1^3) + \cos\theta \sin^2\theta E(u_2^3), \\
Cov(Y_1^3, Y_2) &= \cos^3\theta \sin\theta [E(u_1^4) - 3] - \cos\theta \sin^3\theta [E(u_2^4) - 3], \\
Cov(Y_1^2, Y_2^2) &= \cos^2\theta \sin^2\theta [E(u_1^4) + E(u_2^4) - 6], \\
Cov(Y_1^2, Y_2^3) &= \cos^2\theta \sin^3\theta E(u_1^5) + \cos^3\theta \sin^2\theta E(u_2^5) \\
&\quad + E(u_1^3)[3\cos^4\theta \sin\theta + 6\cos^2\theta \sin^3\theta + \sin^5\theta - \sin^3\theta] \\
&\quad + E(u_2^3)[3\cos\theta \sin^4\theta - 6\cos^3\theta \sin^2\theta + \cos^5\theta - \cos^3\theta], \\
Cov(Y_1^3, Y_2^3) &= \cos^3\theta \sin^3\theta [E(u_1^6) - E(u_2^6)] \\
&\quad + [E(u_1^4) - E(u_2^4)][3\cos^5\theta \sin\theta - 9\cos^3\theta \sin^3\theta + 3\cos\theta \sin^5\theta] \\
&\quad + E(u_1^3)E(u_2^3)[9\cos^2\theta \sin^4\theta - 9\cos^4\theta \sin^2\theta] \\
&\quad - [E(u_1^3) - E(u_2^3)] \cos^3\theta \sin^3\theta.
\end{aligned}$$

The closed-form expressions can be used to characterize the distributions of sources u_1, u_2, for which a given power covariance is not-informative, i.e. equal to zero for any value of θ.

On-Line Appendix 2

Additional Assumptions

The additional assumptions given below complete Assumption A.1 and are sufficient to derive Propositions 1 and 2 by applying the results in Shi, Shum (2015).

Additional Assumptions AA:

i) The parameter space Θ is compact with a non-empty interior.

ii) The function $\xi(\theta)$ is twice continuously differentiable on the interior of the parameter space

iii) The closure of the interior of the implied identified set is such that:

$cl[int\Theta_0^*(f_0)] = \Theta_0^*(f_0)$.

iv) The rank of $\frac{\partial\gamma(\theta)}{\partial\theta'}$ is constant on the interior of the implied identified set.

Condition AA(ii) implies the first part of Assumption (4) in Theorem 2.1 of Shi, Shum (2015) and corresponds to the assumption in Andrews et al. (2004). Condition AA iv) is the second part of their Assumption (4).