Dynamic Deconvolution of (Sub)Independent Autoregressive Sources

Christian Gourieroux * and Joann Jasiak †

March 30, 2020

Abstract

We consider a multivariate system $Y_t = AX_t$, where the unobserved components (the sources) X_t are (sub)independent AR(1) processes and the number of sources is larger than the number of observed outputs (undetermined system). We demonstrate that the mixing matrix A, the autoregressive coefficients and the distributions of sources can all be identified, solving the deconvolution problem. The proof is constructive and allows us to introduce simple consistent estimators of all unknown scalar and functional parameters of the model. Next, the results are extended to a noisy deconvolution $Y_t = AX_t + \eta_t$, with the additional multivariate noise η_t. Applications to causal models with structural innovations are also discussed, such as the identification in error-in-variables models and causal mediation models.

Keywords: Independent Component Analysis (ICA), Blind Source Separation (BSS), Convolutive BSS, Deconvolution, Dynamic Subindependence, Learning Algortithm, Error-in-Variables Model, Causal Model.

The authors thank D. Ben Moshe, M. Deistler, A. Monfort, Y. Shin, A. Tamoni and V. Zinde-Walsh for helpful comments.

---

*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 chair ACPR/Risk Foundation: Regulation and Systemic Risks, the ECR DYSMOIA and the Natural Sciences and Engineering Council of Canada (NSERC).

1 INTRODUCTION

Let us consider a linear system $Y_t = AX_t$, where Y_t is a vector of observed outputs (sensors) of dimension L, X_t is a vector of unobserved (sub)independent components of dimension K (sources, inputs) and the mixing matrix A is unknown. There exists a large literature on the identification of the mixing matrix and sources, when $L = K$ and the sources are serially independent. However, the literature on identification is particularly sparse when the sources are serially dependent, their dynamics need to be identified (**convolutive mixtures**) and/or the number of sources is strictly larger than the number of observed outputs $K > L$ (**undetermined system**) [see Hyvarinen, Karhunen, Oja (2001), Pedersen et al. (2007), Comon, Jutten (2010), Shi (2011), Schennach (2016) for surveys].

The objective of this paper is to provide a solution to the identification problem in *multivariate undetermined convolutive systems* when the sources follow (sub)independent autoregressive processes of order 1 (AR(1)).

For illustration, we assume there is a single output $L = 1$ [1] The observed univariate time series (Y_t) can be written as the following sum of K component series (sources):

$$Y_t = \sum_{k=1}^{K} \left(\frac{1}{1 - \rho_k B} \epsilon_{kt} \right) \equiv \sum_{k=1}^{K} X_{kt}, \ |\rho_k| < 1, \ \forall k, \tag{1.1}$$

where the K sequences $(\epsilon_{kt}), k = 1, \ldots, K$, are strong white noises, which are mutually (sub)independent, with mean 0 and finite variances $\sigma_k^2, \ k = 1, ..., K$, and B is the lag operator. As discussed in Section 2, the independence assumption can be weakened and replaced by a sub-independence assumption [see, Appendix A.1 and Hamedani, Maadooliat (2015), Schennach (2019) for the definition of subindependence]. Henceforth, independence is assumed for expository purpose to relate the text with the standard ICA literature.

We demonstrate that:

i) the number of sources K, the autoregressive coefficients $\rho_k, \ k = 1, ..., K$, and

ii) the marginal distributions of the K strong white noise processes

can be identified. We also introduce consistent estimators of the identifiable scalar and functional parameters.

Let us introduce some mild identification restrictions. We observe that when $\rho_1 = \rho_2$, for example, we get:

[1] This assumption will be relaxed later.

$$\frac{1}{1-\rho_1 B}\epsilon_{1t} + \frac{1}{1-\rho_2 B}\epsilon_{2t} = \frac{1}{1-\rho_1 B}(\epsilon_{1t}+\epsilon_{2t}) = \frac{1}{1-\rho_1 B}\tilde{\epsilon}_t,$$

where $\tilde{\epsilon}_t = \epsilon_{1t} + \epsilon_{2t}$, and the pair $(\epsilon_{1t}, \epsilon_{2t})$ cannot be distinguished from the pair $(\tilde{\epsilon}_t, 0)$. Moreover, if one ϵ_{kt} is zero, then $X_{k,t} = 0$ and the number K of components cannot be identified. This leads us to the following assumption:

IDENTIFICATION ASSUMPTION A.1:

i) The K autoregressive coefficients are distinct.

ii) The marginal distribution of ϵ_{kt} is not degenerate with a point mass at 0, for all $k, k = 1, ..., K$.

Another identification problem is in distinguishing the decomposition $\frac{1}{1-\rho_1 B}\epsilon_{1t} + \frac{1}{1-\rho_2 B}\epsilon_{2t}$ from the decomposition $\frac{1}{1-\rho_2 B}\epsilon_{2t} + \frac{1}{1-\rho_1 B}\epsilon_{1t}$, because the unobserved sources are defined up to a permutation. To solve this second identification issue, we introduce the following assumption:

IDENTIFICATION ASSUMPTION A.2: The autoregressive coefficients are arranged in ascending order: $\rho_1 < \rho_2 < \ldots < \rho_K$.

The autoregressive coefficients can be of any sign and one of them can equal 0. Model (1.1) provides a decomposition of Y_t into components with different persistence (memory)[2].

Under the above Assumptions A.1-A.2, we demonstrate that all scalar and functional parameters are identifiable. First, we obtain the identification of parameters $K, \rho_1, ..., \rho_K$, $\sigma_1^2, \ldots, \sigma_K^2$, and next, the identification of the K distributions of sources given the parameters.

The paper is organized as follows. In Sections 2 and 3, the one-output (univariate) model (1.1) is considered. In Section 2, we prove the second-order identification of parameters $K, \rho_k, \sigma_k^2,\ k = 1, ..., K$. In Section 3, we show that the distributions of sources can be identified from pairwise distributions of (Y_t, Y_{t-1}), if $K \leq 3$, and of $(Y_t, Y_{t-1}), (Y_t, Y_{t-2}), (Y_t, Y_{t-3})$, if $K \geq 4$. This identification result is obtained by considering either the pairwise cumulant generating functions (c.g.f), or the second characteristic functions, without assuming the non-Gaussianity of sources. Thus, pairwise analysis is sufficient for identification, which explains why the condition of independence of sources

[2]See Section 6.2. for a more detailed discussion and Gourieroux, Jasiak (2020), Section 2.2.

can be weakened to a condition of sub-independence [See, Appendix A.1]. In Section 4, the identification results are extended to systems with any number of observed outputs and sources. Section 5 introduces simple non-parametric estimation methods for the distributions of sources. Section 6 focuses on applications, such as the errors-in-variables model, mediation models and filtering algorithms for factor models. Section 7 concludes. The mathematical proofs are gathered in the Appendices.

2 Identification of autoregressive coefficients and variances of sources

The parameters $K, \rho_k, \sigma_k, k = 1, ..., K$ can be identified by considering the second-order properties of process (Y_t). This second-order analysis is based on the spectral density of process (Y_t) given by:

$$\begin{aligned} \varphi(w) &= \sum_{k=1}^{K} \left[\frac{\sigma_k^2}{2\pi} \frac{1}{|1-\rho_k exp(iw)|^2} \right] \\ &= \sum_{k=1}^{K} \left[\frac{\sigma_k^2}{2\pi} \frac{1}{1+\rho_k^2 - 2\rho_k cos(w)} \right]. \end{aligned} \quad (2.1)$$

where w is the frequency and $i = \sqrt{-1}$ is the imaginary root of -1.

PROPOSITION 1: Parameters $K, \rho_k, \sigma_k^2, k = 1, ..., K$ are characterized by the spectral density.

PROOF: Equation (2.1) is a partial fraction decomposition of spectral density, considered as a rational function of $exp(iw)$ (also called the transfer function). The identification of parameters $K, \rho_k, \sigma_k^2, k = 1, ..., K$, follows from the uniqueness of a partial fraction decomposition [3] of a rational function when $\rho_1 < \rho_2 < ... < \rho_K$ [see, e.g. Bradley, Cook (2012)]. [4] QED

Proposition 1 implies that parameters $\rho_k, \sigma_k^2,\ k = 1, ..., K$ can be consistently estimated by the Gaussian Pseudo-Maximum Likelihood (PML) method for any given K. The Gaussian PML estimator can be obtained by writing model (1.1) in a state-space form with sources $X_{kt}, k = 1, ..., K$ as the state variables. Next, the Gaussian pseudo

[3] A consequence of the fundamental theorem of algebra, i.e. the D'Alembert-Gauss Theorem.

[4] This proof shows that our results can be extended to autoregressive sources of higher order (i.e. AR(p) sources), whenever the different lag polynomials are identifiable from the spectral density.

log-likelihood function is maximized by using the Kalman filter. This procedure provides consistent estimates of parameters $\rho_k, \sigma_k^2,\ k = 1, ..., K$.

When K is moderate, an alternative estimation approach is available. One can apply to (y_t) the Box-Jenkins estimation procedure for univariate ARMA processes with a single weak white noise. Indeed, process (y_t) can be viewed as a weak ARMA(p,q) and written as:

$$Y_t = \sum_{k=1}^{K} \frac{\epsilon_{kt}}{1-\rho_k B} = \frac{1}{\prod_{k=1}^{K}(1-\rho_k B)} \left\{ \sum_{k=1}^{K} \left[\left(\prod_{l \neq k} (1-\rho_l B) \right] \epsilon_{k,t} \right] \right\}.$$

The numerator is a sum of independent moving average processes of order $K-1$. The denominator is an autoregressive polynomial of order K, if $\rho_k \neq 0, \forall k$, and of order $K-1$, otherwise. The sum of the moving average processes in the numerator can also be written as a weak MA(K-1) process with a single noise [Ansley(1977)]:

$$Y_t = \frac{\left(1 - \sum_{j=1}^{K-1} \theta_j B^j\right) u_t}{\prod_{k=1}^{K}(1-\rho_k B)}, \tag{2.2}$$

where (u_t) is a weak white noise. The weak white noise $(u_t) = (u_t, u_{t-1}, ...,)$ is serially uncorrelated. However its terms are mutually (nonlinearly) serially dependent, except if all noises $\epsilon_{kt}, k = 1, ..., K$ are Gaussian (see the discussion in Appendix 3).

It follows that process (Y_t) has a weak ARMA (K-1,K) representation, if $\rho_k \neq 0,\ \forall k$, and a weak ARMA (K-1, K-1) representation, otherwise. Then, the moving average parameters $\theta_j,\ j = 1, ..., K-1$, autoregressive parameters $\rho_k, k = 1, ..., K$, and the variance of the weak white noise $\sigma^2 = Var(u_t)$ can be estimated by the Box-Jenkins procedure, i.e. by maximizing the Gaussian likelihood-function of a univariate ARMA(K-1,K) process. Next, the estimators of $\sigma_k^2, \rho_k, k = 1, ..., K$, are found by inverting the mapping, which defines the expressions of $\theta_j, j = 1, .., K-1, \sigma^2$ in terms of $\sigma_k^2,\ k = 1, ..., K$, for given $\rho_k,\ k = 1, ..., K$, for example.

The second-order identification of linear filter (2.2) is a consequence of the AR(1) assumption on the dynamics of sources and the uniqueness of partial fraction decomposition of spectral density. This remains valid if additional strong ARMA(2,1) sources are added, such as:

$$(1 - \phi_1 B - \phi_2 B^2) X_t = (1 - \theta B)\epsilon_t,$$

where the roots of the autoregressive lag polynomial are complex conjugates, or when the sources are strong AR(p) sources with multiple real roots $(1-\rho B)^p X_t = \epsilon_t$. If the sources are finite-order MA(q) processes, such as the MA(1) processes for example, the moving average coefficients θ_j are not second-order identifiable. In such a case, the moving average coefficients can be identified under the additional assumption of sources being (sub)independent and non-Gaussian, from the third and fourth order moments [see e.g. Thi, Jutten (1995) in a general framework, Reiersol (1950), Cragg (1997), Dagenais et al. (1997), Erickson, Whited (2002), Ben Moshe (2018b) for the errors-in-variables model]. Moreover, the moving average coefficients can be second-order identifiable under additional restrictions on those parameters [see Appendix 3 and e.g. Maravall (1979), Nowak (1985), (1989)].

3 Identification of source distributions from pairwise nonlinear dependence

Let us now consider the identification of source distributions given the parameters K, ρ_k, σ_k^2, $k = 1, ..., K$. For this purpose, we consider the information contained in the pairwise distributions of $(Y_t, Y_{t-h}), h = 1, 2, ...$ In this aspect, our approach resembles the identification based on autocovariances $Cov(Y_t, Y_{t-h}), h = 1, 2...$, which is commonly used in the time series analysis at second-order. However, the second-order analysis does not suffice to identify the distributions of sources. Therefore, we introduce a dynamic identification approach that is valid for any number K of sources.

Let us recall the approach of Bonhomme, Robin (2010) and discuss its limitations. The static Independent Component Analysis (ICA) is applied to the following linear model of a bivariate vector of outputs $(Y_t, Y_{t-1})'$:

$$\begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} \rho_1 & \dots & \rho_K \\ 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ \vdots \\ X_{K,t-1} \end{pmatrix} + \begin{pmatrix} 1 & \dots & 1 \\ 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{Kt} \end{pmatrix}. \quad (3.1)$$

The above system is equivalent to (1.1) with a linear relation between the 2-dimensional output and $2K$ independent inputs. The approach of Bonhomme, Robin (2010) is valid as long as the number of inputs $2K$ is less than or equal to $\frac{L(L+1)}{2}$, where L is the number of outputs [5]. In the case when $L = 2$, which is the dimension of output vector in (3.1),

[5]More precisely, this approach is based on the identification result in Szekely, Rao (2000), in which the number of sources is bounded. That result, in turn, follows from a version of Kotlarski's Lemma [see Evdokimov, White (2012)].

that upper bound is equal to 3. Hence, the approach of Bonhomme, Robin (2010) can be used if $2K \leq 3$, i.e. $K \leq 1$, which is rather limited for practical applications.

To obtain a dynamic deconvolution approach, valid for any number K of sources, an additional assumption is needed on the marginal distribution of the sources or, equivalently, on the distributions of errors $\epsilon_{k,t}$ in the autoregressive representations of sources:

$$X_{k,t} = \frac{\epsilon_{kt}}{1-\rho_k B},\ k = 1, ..., K. \tag{3.2}$$

Let us consider the cumulant generating function (c.g.f.) of the marginal (i.e. stationary) distribution of source $X_{k,t}$:

$$c_k(u) = \log E[\exp(uX_{k,t})],\ k = 1, ..., K, \tag{3.3}$$

where u is possibly complex. When u is pure imaginary, equation (3.3) defines the second characteristic function. This second characteristic function is assumed to be well-defined. In particular, a well-defined second characteristic function requires a non-vanishing characteristic function [6]. When u is real, the existence of the real c.g.f. is also required.

ASSUMPTION A.3:

i) The second characteristic function is well defined, i.e. the characteristic function does not vanish,

or ii) The real c.g.f.'s c_k, $k = 1, \ldots, K$, with real arguments u, exist in a neighborhood $U \subset R$ of 0 and these real c.g.f.'s on U characterize the marginal distributions of $(X_{k,t})$, $k = 1, ..., K$.

Assumption A.3 ii) implies the existence of all power moments of X_{kt}, and a summability condition on these power moments, i.e. the so-called Carleman condition [Carleman (1923)]. Hence, identification is obtained either from the second characteristic function, which requires Assumption A.3 i) only, or from the real c.g.f., which requires Assumption A.3 ii).

Similarly, we introduce the (complex) c.g.f. of the error:

$$b_k(u) = \log E[\exp(u\epsilon_{kt})],\ k = 1, ..., K. \tag{3.4}$$

[6] Identification results have also been derived in special cases under weaker conditions [see e.g. Sasvari (1986), Chapter 2 in Rao (1992), Evdokimov, White (2012)].

It is easy to see that the (complex) c.g.f. b_k of error (ϵ_k) is directly linked to the (complex) c.g.f c_k of source (X_k) as follows (see Appendix A.1): $b_k(u) = c_k(u) - c_k[\rho_k u],\ k = 1, ..., K$. Then, we get the following Proposition:

PROPOSITION 2: In model (1.1) under Assumptions A.1, A.2 and A.3 i) (non-vanishing characteristic functions) (resp. Assumptions A.1, A.2, A.3 ii)), the distributions of (sub)independent sources $(X_{k,t})$ and of errors $(\epsilon_{kt}),\ k = 1, ..., K$ are identifiable :

i) for $K \leq 3$, from the pairwise second characteristic functions (resp. joint real c.g.f.) of (Y_t, Y_{t-1}),

ii) for $K \geq 4$ from the pairwise second characteristic functions (resp. joint real c.g.f.) of (Y_t, Y_{t-1}) and (Y_t, Y_{t-2}), if all absolute values $|\rho_k|$ are distinct, and from the pairwise second characteristic functions (resp. joint real c.g.f) of (Y_t, Y_{t-1}) and (Y_t, Y_{t-3}), if some absolute values $|\rho_k|$ are equal.

PROOF: see Appendix 1. QED

The dynamic identification is obtained without the assumptions of non-Gaussianity and a determined system, which are required for identification in the static ICA framework [see, e.g. Comon (1994), Th.11]. This is because of the additional knowledge provided by the autoregressive dynamic structure. The pairwise nonlinear dependence being sufficient to identify is not surprising, as it is also the case in the static ICA [Comon (1994), Th 11]. The identification is achieved for any distributions of sources that can be continuous, discrete, or mixtures of discrete and continuous distributions [see e.g. Zinde-Walsh (2013), (2014) for the notion of generalized functions].

The identification of distributions for $K = 1$ (determined case) can be done directly. We have $Y_t = X_{1t} = \frac{\epsilon_{1t}}{1-\rho_1 B}$. Since the distribution of (Y_t) as well as parameter ρ_1 are identifiable, we can easily derive the distribution of $\epsilon_{1t} = Y_t - \rho_1 Y_{t-1}$. In this special case, we identify not only the distribution of ϵ_{1t}, but also the shocks ϵ_{1t} themselves. Generally, when there are strictly more sources than outputs, the relation between sources and outputs cannot be inverted [7], to find the values of sources without ambiguity.

When the sources are independent, the result of Proposition 2 can be interpreted in terms of deconvolution. Let us consider $K = 2$, for ease of exposition. The deconvo-

[7] except if the distributions of errors $(\epsilon_{k,t})$ have sufficiently different supports [see the discussion in Yilmaz, Rickard (2004)].

lution concerns the bivariate density $f(y_t, y_{t-1}) = f_1(x_{1t}, x_{1,t-1}) * f_2(x_{2t}, x_{2,t-1})$ [8]. This deconvolution is feasible as the joint densities f_1, f_2 are restricted : they depend only on parameters ρ_k and univariate functions $c_k, k = 1, 2$. These semi-parametric restrictions on the $f_k's$ are sufficient for the feasibity of deconvolution.

4 Multivariate system

Proposition 2 can be used for identifying the mixing matrix of coefficients A in an undetermined convolutive multivariate system:

$$Y_t = AX_t, \tag{4.1}$$

or with an additional noise:

$$Y_t = AX_t + \eta_t, \tag{4.2}$$

where the number of output series is $L \geq 1$, processes $X_t = (X_{1t}, ..., X_{Kt})'$ are the AR(1) processes of (sub)independent sources and (η_t) is a strong multivariate white noise (sub)independent of sources (X_t). The components of noise $(\eta_{jt}), j = 1, ..., L$ can be mutually (linearly or nonlinearly) dependent.

A given row of the system can be written as:

$$\begin{aligned} Y_{j,t} &= \sum_{k=1}^{K} a_{j,k} X_{kt} (+\eta_{jt}) && (4.3) \\ &= \sum_{k=1}^{K} \frac{a_{j,k}\epsilon_{k,t}}{1-\rho_k B} (+\eta_{jt}) && (4.4) \\ &= \sum_{k=1}^{K} \frac{\tilde{\epsilon}_{jk,t}}{1-\rho_k B} (+\eta_{jt}), && (4.5) \end{aligned}$$

where $(\tilde{\epsilon}_{jk,t})$ are (sub)independent sequences of variables with mean 0. Thus, Proposition 2 can be applied to any fixed row in order to identify the distribution of $\tilde{\epsilon}_{jk,t} = a_{j,k}\epsilon_{k,t}$. Next, by comparing the distributions of $\tilde{\epsilon}_{jk,t}$ and $\tilde{\epsilon}_{lk,t}$ we find the ratios $a_{j,k}/a_{l,k}$ (for $a_{l,k} \neq 0$). More precisely, these ratios are directly identified if the distribution of $\epsilon_{k,t}$ is not symmetric. Otherwise, the result is obtained by considering linear combinations of equations as shown in the example of a linear treatment effect given later in this Section.

[8] In general, under the weaker assumption of subindependence, the joint density of (y_t, y_{t-1}) cannot be characterized from the densities of $(x_{1t}, x_{1,t-1})$ and $(x_{2t}, x_{2,t-1})$ [Schennach (2019), Th. 4]

To derive a general identification result, let us now introduce the assumption of a saturated mixing matrix.

ASSUMPTION A.4: Saturated mixing matrix

The elements of mixing matrix A are different from 0.

Then, we obtain :

PROPOSITION 3 : Under Assumptions A.1, A.2, A.3 (i) or (ii), A.4:

i) In mixing model (4.1) each column of A is identifiable up to a scale factor.

ii) In the noisy mixing model (4.2), if the autoregressive coefficients $\rho_k,\ k = 1, ..., K$ are different from 0, each column of A is identifiable up to a scale factor. The joint distribution of η_t is identifiable too.

PROOF: By considering the jth row in model (4.2), we identify the marginal distribution of η_{jt}. We can also apply Proposition 2 to any linear combination $\alpha' Y_t$, and then identify the distribution of $\alpha' \eta_t$. As the knowledge of the joint distribution of η_t is equivalent to the knowledge of the distributions of linear combinations $\alpha' \eta_t$, for any α, the second result ii) follows. QED

The identification result of Proposition 3 i) shows that under Assumption A.4, it is always possible to assume all elements in the first row of matrix A are equal to 1 [see Assumption A.2 in Szekely, Rao (2000)]

COROLLARY 1: Uniqueness:

Under the assumptions of Proposition 3, it is possible to identify each column of the mixing matrix A, up to a scale factor, the autoregressive parameters $\rho_1, ..., \rho_K$, the distributions of sources (X_{kt}) and (ϵ_{kt}), and the joint distribution of the noise η_t.

Corollary 1 illustrates the so-called uniqueness property [Eriksson, Koivunen (2004), p. 602] that completes the discussion of identification for multivariate noisy systems.

For a single output system, the above identification is proven under weaker assumptions than in the existing literature, as it is valid for any number K of sources [9] and without strong restrictions on the distributions of sources. For more than one output, the result in Proposition 3 can be compared to the standard conditions for identification of sources that

[9] Identification of convolutive systems with less sources than outputs can be found in Albataineh, Salem (2014), for example.

are additionally assumed serially independent in the existing literature. Those conditions are:
a) at most one source distribution is Gaussian [see e.g. Comon (1994), Hyvarinen, Karhunen, Oja (2001), Eriksson, Koivunen (2004) Th 5. iii), Chan, Ho, Tong (2006), Ben-Moshe (2018a,b)],
b) the moments exist up to order three, four, or six [see e.g. Comon, De Lathauwer (2010), Erickson et al. (2014)],
c) the characteristic functions of sources have no exponential factor [Eriksson, Koivunen (2004) Th 5, ii), iv)],
d) restrictions on the support or exclusion restrictions are imposed [Ben-Moshe (2018a), Williams (2020)].
Moreover the number of sources can be upper bounded [De Lathauwer (2008), Bonhomme, Robin (2010), Zinde-Walsh(2013), Section 4.2.2.3 [10]], or Bayesian approaches can be used [11] [Klepper, Leamer (1984), Leonard (2011)].

In our approach, the identification is achieved by taking into account the AR(1) dynamics of sources. From a practical perspective, these dynamics will be detected, if a large number of observations T is available and the dimension L is fixed. As compared to the similar identification problem encountered in panel data [see e.g. Lewbel (1997), Bonhomme, Robin (2010), Ben Moshe (2018b), Section 3], the asymptotics are different as $N \to \infty$, T fixed is required in those cases (where N denotes the cross-sectional dimension of the panel), while L fixed $T \to \infty$ is assumed in our dynamic framework.

The identification in the presence of noise η_t with an unknown distribution is also obtained. Most ICA algorithms do not allow for an additional noise and assume $\eta_t = 0$, " hoping that the noise-free methods work well if the signal-to-noise ratio is high enough" [Cardoso, Pham (2011)]. Alternatively, the deconvolution literature assumes a given distribution of the additional noise [Meister (2009), Section 3.1, Zinde-Walsh (2013) p. 104, Schennach (2016)].

Our approach for proving the identifiability is different from those described in the literature on the BSS and ICA. Those methods consider first the identifiability of the

[10]as a consequence of the rank condition for matrix A assumed in Zinde-Walsh (2013), Lemma 4.1.

[11]See Poirier (1998) for the interpretation of Bayesian methods in nonidentified models.

mixing coefficients A (in our framework, additionally ρ_k's) and generally disregard the identification of the distribution of $\epsilon_{k,t}$. In our framework, we proceed in three steps : first the dynamic parameters ρ_k are identified, next the distributions of sources are identified, and finally, the instantaneous mixing parameters in A are identified. The rationale is the following: by applying the approach of Section 2 to each component of Y_t, it is possible to identify the values $a_{jk}^2, \sigma_{jk}^2,\ j = 1, ..., L,\ k = 1, ..., K$, and hence the $|a_{jk}|$ under Assumption A.4, but not the signs of these mixing coefficients. To identify the signs, combinations of pairs of outputs are needed. In the very special case of a determined system of convolutive mixtures, the direct approach of Miettinen et al.(2014) based on the second order moments could be used, although this approach is not valid for undetermined systems.

The assumptions stated in Corollary 1 are sufficient to derive the uniqueness property. Proposition 2 can also be used to derive the uniqueness, under well-chosen exclusion restrictions, as shown in the example below.

Example : A dynamic linear treatment effect.

Let us consider the model :

$$\begin{cases} D_t &= aX_t + u_{1t}, \\ Y_t &= \beta D_t + bX_t + u_{2t}, \end{cases}$$

where D_t is the continuous treatment (an economic policy, say), Y_t the continuous outcome, and the unobserved structural shocks u_{1t}, u_{2t} and latent variable X_t are independent of each other. This system is a linear analogue of the triangular binary model with factor structure considered in the recent literature on treatment effects [see e.g. Khan, Maurel, Zhan (2019) and the references therein]. We are interested in identifying the value of β, that characterizes the "causal" effect of the treatment and has to be distinguished from the confounding effect of the common latent factor X_t. If the unobserved confounding variable X_t follows an AR(1) process, the identification of β is obtained as follows: We write the associated reduced form

$$\begin{cases} D_t &= aX_t + u_{1t}, \\ Y_t &= (b + a\beta)X_t + \beta u_{1t} + u_{2t}, \end{cases}$$

and consider different combinations of observed variables as in Szekely, Rao (2000). At each step, Proposition 2 is used to i) identify the distribution of u_{1t} from the first equation;

ii) identify the distribution of $v_t = \beta u_{1t} + u_{2t}$ from the second equation; and iii) identify the distribution of $w_t = (1+\beta)u_{1t} + u_{2t}$ from the equation for $D_t + Y_t$. Next, iv) parameter β is identified as $\beta = \frac{1}{2}\left[\frac{Vw_t - Vv_t}{Vu_{1t}} - 1\right]$ and the distribution of u_{2t} is identified by the deconvolution of distribution of v_t, with a known distribution of βu_{1t}. It follows that the dynamic treatment effect is identified without introducing any external instrumental variables.

This identifiability property is an extension of a well-known identification result for simultaneous equation models. In the presence of explanatory variables, parameter β is identified if an explanatory variable that is present in the first equation is absent from the second equation. The example given above extends this identifiability property to unobservable shocks. Indeed, shock u_{1t}, which is present in the first equation, does not belong in the set of shocks of the second equation (i.e. it is independent of shocks in the second equation). This is analogous to the exclusion restrictions introduced in the errors-in-variables model in Ben Moshe (2018b).

5 Nonparametric Estimation of Distributions of Sources

Let us introduce consistent estimation methods for the identified scalar and functional parameters (see Corollary 1). When the noise distributions are from parametric families of distributions, these parameters can be estimated by the Simulated Maximum Likelihood, for example, to avoid integrating out the latent sources. One can also consider nonparametric estimation methods for the distributions of sources [see e.g. Bonhomme, Robin (2010) in a similar framework]. For AR(1) sources in our case, it is possible to use for estimation the analytical formulas of some derivatives of the c.g.f. c_k (or of the second characteristic function) given in the proof of Proposition 2 in Appendix 1. This can be easily done for K=2, or K=3. We describe in detail the estimators obtained in these two cases. The extension to the case $K \geq 4$ requires the estimation of additional parameters to recover a low degree polynomial component in the c.g.f. (see Appendices 1, 5).

5.1 Estimation for K=2

Let us denote the pairwise (complex) c.g.f of bivariate output vector (Y_t, Y_{t-1}) by $\Psi(u,v) = logE[exp(uY_t + vY_{t-1})]$. As shown in Appendix 1, A.4, we have:

$$\frac{\partial\psi}{\partial v}(u,v) = \sum_{k=1}^{2} c_k'(v + u\rho_k),$$

and upon the change of argument described in Appendix 1, A.2:

$$\frac{\partial\psi}{\partial v}\left[\frac{w_2 - w_1}{\rho_2 - \rho_1}, \frac{\rho_2 w_1 - \rho_1 w_2}{\rho_2 - \rho_1}\right] = c_1'(w_1) + c_2'(w_2). \tag{5.1}$$

PROPOSITION 4: For K=2, we have:

$$\begin{aligned} c_1'(w) &= \frac{\partial\psi}{\partial v}\left(\frac{-w}{\rho_2 - \rho_1}, \frac{\rho_2 w}{\rho_2 - \rho_1}\right), \\ c_2'(w) &= \frac{\partial\psi}{\partial v}\left(\frac{w}{\rho_2 - \rho_1}, \frac{-\rho_1 w}{\rho_2 - \rho_1}\right). \end{aligned}$$

PROOF: These formulas are obtained by setting $w_2 = 0$ (resp. $w_1 = 0$) in equation (5.1).

QED

When u is real, the derivative $\frac{\partial\psi}{\partial v}(u,v)$ is equal to:

$$\begin{aligned} \frac{\partial\psi}{\partial v}(u,v) &= \frac{\partial}{\partial v} logE[exp(uY_t + vY_{t-1})] \\ &= \frac{E[Y_{t-1}exp(uY_t + vY_{t-1})]}{Eexp(uY_t + vY_{t-1})}, \end{aligned} \tag{5.2}$$

which is the expectation of Y_{t-1} with respect to density $\frac{exp(uY_t+vY_{t-1})]}{Eexp(uY_t+vY_{t-1})} f(Y_t, Y_{t-1})$, where $f(Y_t, Y_{t-1})$ is the joint p.d.f. of Y_t, Y_{t-1}. This derivative is consistently estimated from its empirical counterpart, and so are the derivatives c_1', c_2', by applying the formulas in Proposition 4 (see Appendix 4 for the functional estimators and their asymptotic properties).

The closed-form formulas in Proposition 4 are valid for the pairwise real c.g.f. of (Y_t, Y_{t-1}) as well as for the pairwise second characteristic function (with u, v replaced by iu, iv). In that case, we write the pairwise functions Ψ and c_k in terms of imaginary arguments as follows:

$$\begin{aligned} \Psi(u,v) &= \log E\{exp[i(uY_t + vY_{t-1})]\} \\ &= log|Eexp[i(uY_t + vY_{t-1})]| + iArg\{Eexp[i(uY_t + vY_{t-1})]\}. \end{aligned}$$

The second characteristic function is not differentiable in u, v due to the presence of the modulus and Arg of the complex number in the polar representation. However, the relations between the "derivatives" are still valid with

$$\frac{\partial \Psi}{\partial v}(u,v) \equiv \frac{1}{\Phi(u,v)} \frac{\partial \Phi(u,v)}{\partial v},$$

where $\Phi(u,v) = Eexp[i(uY_t + vY_{t-1})]$ is the first characteristic function, by applying the theory of generalized functions [see e.g. Zinde-Walsh (2014)].

Thus, the estimator of $\Psi(u,v)$ may be obtained either from the estimators of $E\cos(uY_t + vY_{t-1})$ and $E\sin(uY_t + vY_{t-1})$ (under A.3 i)), or from the estimator of $Eexp(uY_t + vY_{t-1})$ with real u, v, (under A.3 ii)).

The above estimation method can be extended to provide the error density of ε_{1t}, for example, when ε_{1t} has a continuous distribution. This can be done along the following steps :

step 1 : Apply Proposition 4 to estimate $c_1'(iw)$ from its sample counterpart $\hat{c}_1'(iw)$, i.e. replace the expectation by the sample average.

step 2 : Find the estimator of the second characteristic function $c_1(iw)$ by : $\hat{c}_1(iw) = \int_0^w \hat{c}_1'(iu)du$, since $c_1(0) = 0$.

step 3 : Compute the estimator of the second characteristic function of ε_{1t} by : $\hat{b}_1(iw) = \hat{c}_1(iw) - \hat{c}_1[i\hat{\rho}_1 w]$.

The above three steps are valid for either continuous, or discrete variables.

step 4 : If ε_{1t} is a continuous variable, its density can be obtained through a regularized Fourier transform, in order to have a well-posed inverse problem. For example, one can use a kernel estimator of the density of ε_{1t}, such as [Ben-Moshe (2018a), eq.22 and p150] :

$$\hat{f}_1(u) = \frac{1}{2\pi} \int_{-1}^{1} \exp(-iuw) \exp[\hat{b}(iw)](1-u^2)^3 du.$$

In practice, the choice between using the pairwise real c.g.f. and the second characteristic function depends on their existence. The real c.g.f. is suitable for all processes with thin tails and the second characteristic function is suitable for all variables with finite second-order moments, including variables with fat-tailed distributions.

5.2 Estimation for K=3

Let us now consider the case $K = 3$. We get the following derivative of the joint c.g.f:

$$\frac{\partial\psi}{\partial v}\left[\frac{w_2-w_1}{\rho_2-\rho_1},\frac{\rho_2 w_1-\rho_1 w_2}{\rho_2-\rho_1}\right]=c_1'(w_1)+c_2'(w_2)+c_3'\left[\frac{(\rho_2-\rho_3)w_1}{\rho_2-\rho_1}+\frac{(\rho_3-\rho_1)w_2}{\rho_2-\rho_1}\right]. \quad (5.3)$$

Let us differentiate both sides of this equality with respect to w_1, say. We get:

$$\begin{aligned}
&- \frac{1}{\rho_2-\rho_1}\frac{\partial^2\Psi}{\partial u\partial v}\left[\frac{w_2-w_1}{\rho_2-\rho_1},\frac{\rho_2 w_1-\rho_1 w_2}{\rho_2-\rho_1}\right]+\frac{\rho_2}{\rho_2-\rho_1}\frac{\partial^2\Psi}{\partial^2 v^2}\left[\frac{w_2-w_1}{\rho_2-\rho_1},\frac{\rho_2 w_1-\rho_1 w_2}{\rho_2-\rho_1}\right]\\
&= c_1'(w_1)+\frac{\rho_2-\rho_3}{\rho_2-\rho_1}c_3''\left[\frac{\rho_2-\rho_3}{\rho_2-\rho_1}w_1+\frac{\rho_3-\rho_1}{\rho_2-\rho_1}w_2\right]. \quad (5.4)
\end{aligned}$$

In particular, by writing this relation for $w_1=0, w_2=w$ and for $w_1=-\frac{\rho_3-\rho_1}{\rho_2-\rho_3}w,\ w_2=w$ and combining, we get :

$$\begin{aligned}
-\frac{1}{\rho_2-\rho_1}\frac{\partial^2\Psi}{\partial u\partial v}\left[\frac{w}{\rho_2-\rho_1},\frac{-\rho_1 w}{\rho_2-\rho_1}\right] &+ \frac{\rho_2}{\rho_2-\rho_1}\frac{\partial^2\Psi}{\partial v^2}\left[\frac{w}{\rho_2-\rho_1},\frac{-\rho_1 w}{\rho_2-\rho_1}\right]\\
&= \sigma_1^2+\frac{\rho_2-\rho_3}{\rho_2-\rho_1}c_3''\left[\frac{\rho_3-\rho_1}{\rho_2-\rho_1}w\right]. \quad (5.5)
\end{aligned}$$

We deduce the closed-form expressions of the second-order derivatives of the c.g.f..

PROPOSITION 5:

For $K=3$, we have:

$$\begin{aligned}
c_3''(w) &= -\sigma_1^2\frac{\rho_2-\rho_1}{\rho_2-\rho_3}-\frac{1}{\rho_2-\rho_3}\frac{\partial^2\psi}{\partial u\partial v}\left[\frac{w}{\rho_3-\rho_1},\frac{-\rho_1 w}{\rho_3-\rho_1}\right]\\
&+ \frac{\rho_2}{\rho_2-\rho_3}\frac{\partial^2\psi}{\partial v^2}\left[\frac{w}{\rho_3-\rho_1},\frac{-\rho_1 w}{\rho_3-\rho_1}\right].
\end{aligned}$$

The second-order derivatives of the real joint c.g.f. have simple expressions. For example, from (5.2) it follows that:

$$\begin{aligned}
\frac{\partial^2\Psi}{\partial u\partial v}(u,v) &= \\
\frac{E[Y_tY_{t-1}exp(uY_t+vY_{t-1})]}{E[exp(uY_t+vY_{t-1})]} &- \frac{E[Y_{t-1}exp(uY_t+vY_{t-1})]}{E[exp(uY_t+vY_{t-1})]}\frac{E[Y_t exp(uY_t+vY_{t-1})]}{E[exp(uY_t+vY_{t-1})]}\\
&= Cov_{u,v}(Y_t,Y_{t-1}), \quad (5.6)
\end{aligned}$$

where $Cov_{u,v}$ is the covariance computed with respect to density $\{exp(uY_t+vY_{t-1})/E[exp\ (uY_t+vY_{t-1})]\}f(Y_t,Y_{t-1})$. As shown in the previous Section, this type of derivative is consistently estimated by its empirical counterpart, i.e. by replacing expectations by sample averages.

5.3 Estimation for $K \geq 4$

In structural economic models (see Section 6), the value of K is often small. Therefore, estimators introduced in Sections 5.1-5.2 are easily applicable. Similar estimation methods can also be developed for $K \geq 4$ from the constructive proof of Darmois Lemma. In contrast, K can be large in applications to operational research and engineering, such as facial or voice recognition, for example. Those applications require numerical algorithms for the implementation of dynamic deconvolution.

5.3.1 Derivation of the baseline c.g.f.

Let us introduce the differential operators:

$$D_k = \frac{\partial}{\partial u} - \rho_k \frac{\partial}{\partial v},\ k = 1, ..., K. \tag{5.7}$$

These operators commute, by Schwarz Lemma, and we have:

$$D_k[c_j'(v + u\rho_j)] = (\rho_j - \rho_k) c_j^{(2)}(v + u\rho_j),\ \forall j, k. \tag{5.8}$$

Let us now consider the relation:

$$\frac{\partial \Psi}{\partial v}(u, v) = \sum_{k=1}^{K} c_k'[v + u\rho_k],$$

and apply the operator $\prod_{j,j\neq k} D_j$ to both sides of the equality. We get:

$$\begin{aligned} \left[\prod_{j,j\neq k} D_j\right] \frac{\partial \Psi}{\partial v}(u, v) &= \left[\prod_{j,j\neq k} D_j\right] \left[\sum_{l=1}^{K} c_l'[v + u\rho_l]\right] \\ &= \prod_{j,j\neq k} (\rho_j - \rho_k) c_k^{(k)}(v + u\rho_k), \end{aligned}$$

whenever the c.g.f's are differentiable up to order K. Then, the k-th order derivative is obtained as:

$$c_k^{(k)}(v) = \left[\left(\prod_{j,j\neq k} D_j\right) \frac{\partial \Psi}{\partial v}(u, v)\right]_{u=0} \Big/ \left[\prod_{j,j\neq k} (\rho_j - \rho_k)\right]. \tag{5.9}$$

These derivatives have to be integrated out K times to recover the c_k functions. The constants of integration are obtained by using the pairwise c.g.f. of either (Y_t, Y_{t-2}), or (Y_t, Y_{t-3}) [see Appendix 5 for K=4].

5.3.2 Learning algorithm

This Section introduces an estimation method that does not require an assumption on the existence of the derivatives up to order K, but only the existence of variances. It is a learning algorithm, which is easy to implement.

Let us assume that all absolute values of autoregressive coefficients $|\rho_k|, k = 1, ..., K$ are distinct. This assumption can be tested as a null hypothesis from the first-step estimators $\hat{\rho}_k, k = 1, ..., K$. Then, functions $c'_k(u)$ are identified from the pairwise c.g.f.'s $\Psi(u, v) = logE[exp(uY_t + vY_{t-1})]$ and $\Psi_2(u, v) = logE[exp(uY_t + vY_{t-2})]$, and more precisely from their partial derivatives with respect to v that satisfy:

$$\frac{\partial \Psi(u, v)}{\partial v} = \sum_{k=1}^{K} c'_k(v + u\rho_k), \tag{5.10}$$

$$\frac{\partial \Psi_2(u, v)}{\partial v} = \sum_{k=1}^{K} c'_k(v + u\rho_k^2), \tag{5.11}$$

Let us denote the sample counterparts of these partial derivatives by $\frac{\partial \hat{\Psi}(u,v)}{\partial v}$, $\frac{\partial \hat{\Psi}_2(u,v)}{\partial v}$ and introduce a suitable basis of functions $g_j(u), j = 1, ..., J$, that will be used to approximate the $c'_k(u)$ functions as: $c'_k(u) \approx \sum_{j=1}^{J} a_{k,j} g_j(u)$. Then, nonparametric estimators of functions $c'_k(u)$ are the functions:

$$\hat{c}'_k(u) = \sum_{j=1}^{J} \hat{a}_{kj} g_j(u), \quad k = 1, ..., K, \tag{5.12}$$

where the coefficients $\hat{a}_{kj}, k = 1, ..., K, j = 1, ..., J$ minimize the following objective function:

$$\begin{aligned} \hat{a} &= Argmin_{a_{kj}} \int\int \left[\frac{\partial \hat{\Psi}(u, v)}{\partial v} - \sum_{k=1}^{K} \sum_{j=1}^{J} a_{kj} g_j(v + u\hat{\rho}_k) \right]^2 \pi(u, v) du dv \\ &+ \gamma \int\int \left[\frac{\partial \hat{\Psi}_2(u, v)}{\partial v} - \sum_{k=1}^{K} \sum_{j=1}^{J} a_{kj} g_j(v + u\hat{\rho}_k^2) \right]^2 \pi(u, v) du dv, \end{aligned} \tag{5.13}$$

where $\pi(u, v)$ and γ are given weights [12]. This is a least-squares optimization that is easily solved numerically. However, given the number KJ of unknown parameters, the least squares estimator formula requires the inversion of a $KJ \times KJ$ matrix, which can

[12] It is out of the scope of this paper to discuss the appropriate choices of basis $g_j(u), j = 1, ..., J$ and weights. These choices depend on whether the calibration relies on the real c.g.f., or on the second charasteristic function. For the second characteristic function, a basis of sine and cosine could be used, with weighting functions with support $[0, 2\pi]^2$.

be of a very large dimension. In such a case, the results may become costly and non-robust. As a solution, one can consider only the diagonal elements of this matrix, as it is commonly done in the machine learning methodology. Then, we need to implement a simple error correction algorithm where, at each step p, the values $\hat{a}_{kj}^{(p)}$ are updated by taking into account the prediction error and its sign. That algorithm is as follows:

$$\hat{a}_{kj}^{(p+1)} - \hat{a}_{kj}^{(p)} \tag{5.14}$$

$$= \int\int g_j(v+u\hat{\rho}_k)\left[\frac{\partial\hat{\Psi}(u,v)}{\partial v} - \sum_{k=1}^{K} c_k'^{(p)}(v+u\hat{\rho}_k)\right]\pi(u,v)dudv \,/ \int\int g_j(v+u\hat{\rho}_k)^2\pi(u,v)dudv$$

$$+ \quad \gamma\int\int g_j(v+u\hat{\rho}_k^2)\left[\frac{\partial\hat{\Psi}_2(u,v)}{\partial v} - \sum_{k=1}^{K} c_k'^{(p)}(v+u\hat{\rho}_k^2)\right]\pi(u,v)dudv \,/ \int\int g_j(v+u\hat{\rho}_k^2)^2\pi(u,v)dudv,$$

where $\hat{c}_k'^{(p)} = \sum_{j=1}^{J}(\hat{a}_{kj}^{(p)}g_j(v))$.

This simplified algorithm can be applied until a numerical convergence is reached to provide in the limit a numerical solution to optimization (5.10).

It would be recommended to check at each step of the algorithm the values of $\sum_{j=1}^{J}(\hat{a}_{kj}^{(p)}g_j(0))$, $k = 1, ..., K$, which need to be close to 0 in the limit (the condition of zero mean) as well as the values of $\sum_{j=1}^{J}(\hat{a}_{kj}^{(p)}g_j'(0))$, $k = 1, ..., K$, that have to be close to the first-step estimate $\hat{\sigma}_k^2$ in the limit.

6 Implications for Structural Modelling

The structural economic models, also called causal models [see e.g. Pearl (2009), (2018), Heckman, Pinto (2015)], by definition, require the assumption of shocks independence, which allows us for (semi-) parametric identification. The assumption of subindependence is not sufficient to define structural shocks and structural impulse response functions. Indeed, if $\epsilon_{1t} = \epsilon_{2t}$ are Cauchy distributed, the subindependence condition is satisfied, but ϵ_{1t} cannot be shocked alone without ϵ_{2t} being shocked too. Below, we discuss the error-in-variables model, mediation analysis and factor models.

6.1 The (structural) errors-in-variables model

The errors-in-variables models have been introduced very early in the econometric and statistical literature [Frisch (1934), Koopmans (1937), Wald(1940), Wooley (1941), Samuelson

(1942), Berkson (1950), Madansky (1959)] and triggered a debate on the choice of a linear versus orthogonal regression, which is equivalent to choosing between the Ordinary Least Squares (OLS) and the Principal Component Analysis (PCA). The basic specification [13][see e.g. Fuller (1987)] concerns two latent variables X_{1t}, X_{2t} measured with errors:

$$Y_{1t} = X_{1t} + \eta_{1t}, \tag{6.1}$$

$$Y_{2t} = X_{2t} + \eta_{2t}. \tag{6.2}$$

The two latent variables are assumed to satisfy a deterministic linear relationship:

$$X_{2t} = aX_{1t}, \text{ say.} \tag{6.3}$$

Haavelmo (1944), page 3, calls variables X the theoretical variables to be distinguished from the observable variables Y, and defines relation (6.3) as the "hypothetical" model.

In this example, $L = 2, K = 1$ and the structural model can be rewritten as:

$$\left\{ \begin{array}{rcl} Y_{1t} & = & X_{1t} + \eta_{1t} \\ Y_{2t} & = & aX_{1t} + \eta_{2t} \end{array} \right. \iff Y_t = \begin{pmatrix} 1 \\ a \end{pmatrix} X_{1t} + \eta_t. \tag{6.4}$$

It has been established in the literature that if X_{1t} is a sequence of i.i.d. non-Gaussian variables, then parameter a can be identified [Geary (1942), Reiersol (1950) for linear relationship between 2 variables, Kapteyn, Wansbeek (1983), Bekker (1986), Ben Moshe (2018b) for more than two variables]. If (X_{1t}) is a sequence of i.i.d. Gaussian variables, parameter a is not identifiable, but belongs in an identifiable convex hull [see, Klepper, Leamer (1984)]. Proposition 3 shows that if $X_{1t} = \rho_1 X_{1t-1} + \epsilon_{1t},\ \rho_1 \neq 0$, then parameter a is identifiable even if the distribution of X_{1t} is Gaussian. Moreover, we can identify the distributions of (ϵ_{1t}) and of $(\eta_t) = (\eta_{1t}, \eta_{2t})'$. Thus, the deconvolution is feasible without the assumption of i) a given fixed distribution of the noise [see Schennach (2016), Section 3.1], or 2) a Gaussian distribution of the noise [Ben Moshe (2018b), Assumption 2.3], or 3) the availability of additional data that make it feasible to estimate the noise distribution, or 4) the symmetry of the noise distribution and some irregularity in the distribution of X_1 [Delaigle, Hall (2016)].

The identifiability of parameter a can be interpreted in terms of instrumental variables as follows : Let us consider the equation:

[13]See Zinde-Walsh (2013), Table 4.1 for different types of measurement error models to which our results also apply.

$$Y_{2t} = aY_{1t} + v_t,$$

where $v_t = \eta_{2t} - a\eta_{1t}$. The lagged variable $Y_{1,t-1}$ can be used as an instrument for parameter a. Indeed, we get:

$$Cov(v_t, Y_{1,t-1}) = Cov(\eta_{2t} - a\eta_{1t}, X_{1,t-1} + \epsilon_{1,t-1}) = 0,$$

and

$$Cov(Y_{1t}, Y_{1,t-1}) = Cov(X_{1t} + \epsilon_{1t}, X_{1,t-1} + \epsilon_{1,t-1}) = \rho_1 \neq 0,$$

by assumption. Therefore, the assumption of AR(1) source allows us to use the lagged noisy observations as an internal instrument. Thus, even if there are more shocks than observed variables, external instruments are not necessary for identification [see Stock, Watson (2018) for a discussion] [14].

6.2 Causal mediation analysis

In the framework of linear causal mediation analysis, the identification result allows us to disentangle the effect of the mediator from the effect of the confounding variable [Holland (1988), Pearl (2009), Zhao, Luo (2019)]. The model below can be considered as an extension of the example of Section 4 with an unobservable confounding variable C_t and observable input X_t, mediator M_t and output [15] Y_t. We get:

$$\begin{aligned} X_t &= e_t, \\ M_t &= aX_t + C_t\alpha + e_{1t}, \\ Y_t &= cX_t + M_t b + C_t\beta + e_{2t}. \end{aligned}$$

This system is equivalently written as:

$$\begin{aligned} X_t &= e_t, \\ M_t &= ae_t + C_t\alpha + e_{1t}, \\ Y_t &= (ab + c)e_t + (b + \frac{\beta}{\alpha})\alpha C_t\beta + be_{1t} + e_{2t}. \end{aligned}$$

[14] This direct proof of identifiability of a achieved with an internal instrument is specific to the basic errors-in-variables models and cannot be extended to the general form introduced in Section 4.

[15] In the neural network methodology, this is a model with 3 layers (deep neural network) and intermediate neurons M_t.

which is a system with the following four sources: $X_t = e_t, C_t, e_{1t}, e_{2t}$. We are interested in the identification of the direct and indirect effects of X and Y that are c and $ab + c$, respectively, and more generally, in identifying all the parameters. If all the sources are independent Gaussian zero-mean variables, then six elements of the variance-covariance matrix of the observables are available, while 8 parameters of the structural causal model need to be found. The causal mediation model with confounding becomes identifiable if the sources are AR(1) processes with distinct autoregressive coefficients. Then, the identification proceeds as follows:

i) source e_t and its distribution are identifiable from the first equation of the system,

ii) next, from the second equation, parameter a, the distributions of e_{1t} and of αC_t are identifiable,

iii) from the third equation we obtain $ab + c$, $\beta(b + \beta/\alpha)$, b and the distribution of e_{2t}.

Thus, we can identify all parameters of interest without using any external instrumental variables [see e.g. Angrist et al. (1996)].

6.3 Filtering and updating algorithm for factor models

Let us consider a dynamic linear factor model:

$$Y_t = a_1 X_{1t} + a_2 X_{2t} + \eta_t,$$

where $dim Y_t = L, K = 2$ in the equation above, and X_{1t}, X_{2t} are serially dependent. In practice, the analysis of this model is often carried out in two steps as follows: In the first step, a static Singular Value Decomposition (i.e. a PCA adjusted for the presence of noise) is applied to the data on Y_t to find proxies of factors $\tilde{X}_{1t}, \tilde{X}_{2t}$, say. Next Y_t is regressed on the proxies to estimate the mixing coefficients a_1, a_2. In the last step, the dynamics of X_{1t}, X_{2t} is estimated from two separate AR(1) models for $\tilde{X}_{1t}$ and $\tilde{X}_{2t}$. This practice is common in linear factor model analysis in Finance [Sharpe (1964), Lintner (1965)], as these models underly the arbitrage pricing theory [Ross (1976)], and also in other fields [see e.g. Lin, Zhang (2018)].

As the above linear factor model is undetermined, the values $\tilde{X}_{1t}, \tilde{X}_{2t}$ cannot provide consistent approximations of the true values X_{1t}, X_{2t}. Moreover, these smooth approximations result in unreliable estimated dynamic patterns of X_{1t}, X_{2t} and their distributions.

Proposition 3 shows that it is possible to avoid these drawbacks by using the joint historical distribution of Y_t, Y_{t-1} instead of the marginal distribution of Y_t only (as in the PCA) and by taking into account the knowledge of the latent (identifiable) AR(1) dynamics of X_{1t}, X_{2t}. In particular, we can use the semi-affine form of process (Y_t) to derive exact filtration formulas based on the characteristic functions [see Bates (2006)], as described below [16]:

i) **The algorithm**

Let us consider the general form :

$$Y_t = AX_t + \eta_t, \tag{6.5}$$

where $X_{k,t} = \rho_k X_{k,t-1} + \varepsilon_{k,t}, k = 1, \dots, K$.

Below, we develop an algorithm that computes recursively the distribution of Y_{t+1}, X_{t+1} given the current and lagged observed values $\underline{Y_t}$ only. Let us denote by $G_{t|t}(.)$ the characteristic function of X_t given $\underline{Y_t}$:

$$G_{t|t}(\mu) = E[\exp(i\mu' X_t)|\underline{Y_t}], \tag{6.6}$$

and compute the joint conditional characteristic function of (Y_{t+1}, X_{t+1}) given $(\underline{Y_t}, \underline{X_t})$. We have :

$$\begin{aligned}
\Phi_{t+1|t}(\lambda, \mu) &= E[\exp(i\lambda' Y_{t+1} + i\mu' X_{t+1})|\underline{X_t}, \underline{Y_t}] \\
&= E(\exp[i(A'\lambda + \mu)' X_{t+1} + i\lambda' \eta_{t+1}]|\underline{X_t}, \underline{Y_t}] \\
&= E[\exp(i\lambda' \eta_{t+1})] E(\exp[i(A'\lambda + \mu)' X_{t+1}]|X_t) \\
&= \phi_\eta(\lambda) E| \exp[i(A'\lambda + \mu)'(diag\rho) X_t + i(A'\lambda + \mu)' \epsilon_t]|X_t] \\
&= \phi_\eta(\lambda) \phi_\epsilon(A'\lambda + \mu) \exp[i(A'\lambda + \mu)' \ (diag\rho) \ X_t],
\end{aligned}$$

where $\phi_\eta(\lambda), \phi_\epsilon(\mu)$ are the characteristic functions of η_t and $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{kt})'$, respectively. It follows the joint characteristic function of (Y_{t+1}, X_{t+1}) given $\underline{Y_t}$ is :

$$\begin{aligned}
\tilde{\Phi}_{t+1|t}(\lambda, \mu) &= E[\exp(i\lambda' Y_{t+1} + i\mu' X_{t+1})|\underline{Y_t}] \\
&= E[(\Phi_{t+1|t}(\lambda, \mu)|\underline{Y_t}] \ \text{(by iterated projection)} \\
&= \phi_\eta(\lambda) \phi_\epsilon(A'\lambda + \mu) G_{t|t}[(diag\ \rho)(A'\lambda + \mu)].
\end{aligned}$$

[16] This is an alternative to the recovering of latent sources by matching [Arellano, Bonhomme (2018)].

From the above joint characteristic function, we derive the characteristic function of X_{t+1} given $\underline{Y_{t+1}} = (Y_{t+1}, \underline{Y_t})$ by applying the Bartlett formula [Bartlett (1938)]. We get:

$$G_{t+1|t+1}(\mu) = \frac{\int \tilde{\Phi}_{t+1|t}(\lambda, \mu) \exp(-i\lambda' y_{t+1}) d\lambda}{\int \tilde{\Phi}_{t+1|t}(\lambda, 0) \exp(-i\lambda' y_{t+1}) d\lambda},$$

which provides the recursive updating formula of the $G_{t|t}$ function :

$$G_{t+1|t+1}(\mu) = \frac{\int \phi_\eta(\lambda)\phi_\epsilon(A'\lambda + \mu) G_{t|t}[(diag\ \rho)(A'\lambda + \mu)] \exp(-i\lambda' y_{t+1}) d\lambda}{\int \phi_\eta(\lambda)\phi_\epsilon(A'\lambda) G_{t|t}[(diag\ \rho)A'\lambda] \exp(-i\lambda' y_{t+1}) d\lambda}. \tag{6.7}$$

This updating formula is easily implemented as long as the number of sources is small, such as $K \leq 4$.

Remark: In the determined case when $L = K$ and $\eta_t = 0$, the filtering can be circumvented as follows. From the consistent estimator $\hat{A}_T$ of mixing matrix A, we obtain the filtered values of sources as $\hat{X}_t = \hat{A}_T^{-1} Y_t$, and the estimated distribution of X_t given Y as a point mass at $\hat{X}_t$. Hence, we get a degenerate distribution at $\hat{X}_t$.

ii) **Nonlinear causality measures**

The algorithm given above can be used to compare nonlinear causality measures between variables with and without measurement errors. Let us consider the case $L = K = 2$ of two outputs and two sources, and consider $Y_t^* = AX_t$, where A is invertible. Suppose that Y_t^* is the output measured without an error, while $Y_t = Y_t^* + \eta_t$ is the output measured with an error. Since all the underlying distributions are identifiable and consistently estimable, we can compute the conditional reduced form distributions $l(y_t|\underline{y_{t-1}})$ of Y_t given $\underline{Y_{t-1}}$ by using the algorithm given above, and also find the conditional structural distribution of Y_t^* given Y_{t-1}^*. Let $g(x_t|x_{t-1})$ denote the conditional distribution of X_t given X_{t-1}. Then, the conditional structural distribution of Y_t^* given Y_{t-1}^* is found from the linear transformation $X \to Y = AX$, which provides the conditional structural density:

$$l^*(y_t^*|y_{t-1}^*) = \frac{1}{|detA|} g[A^{-1}y_t^*|A^{-1}y_{t-1}^*], \tag{6.8}$$

that generally differs from $l(y_t|\underline{y_{t-1}})$. Then, it is possible to use the conditional density in (6.17) for nonlinear structural causality analysis. Moreover, after evaluating both conditional density functions, one can compare the reduced form non-causality from y_{1t} to

y_{2t} with the structural non-causality from y_{1t}^* to y_{2t}^*, say, in the spirit of Anderson et al. (2019).

7 Concluding Remarks

The aim of our paper was to solve the problem of deconvolution of sources in a linear dynamic multivariate system when the number of sources is larger than the number of outputs. For identification, we assumed (sub)independent AR(1) sources and left the distributions of noises unspecified. We have shown how to identify and estimate nonparametrically the mixing matrix, the autoregressive coefficients, as well as the distributions of all underlying noise processes.

The importance of this identification result has been illustrated by considering the identification in the errors-in-variables models and/or mediation models (including Gaussian models), and structural nonlinear causality measures. We have also provided a nonlinear filtering and prediction algorithm based on the characteristic function for dynamic factor models.

The identification result allows for disentangling the ultra long run component from noise in macro-economics and finance [see e.g. Gourieroux, Jasiak (2020) for the definition of a persistent stationary AR(1) process]. The approach for identification of the mixing matrix and the distribution of sources can also be extended to undetermined systems of spatial processes(i.e. random fields) under suitable assumptions on autoregressive patterns in the spatial dependencies(see e.g. Bachoc et al.(2020), Nordhausen et al.(2015) for blind source identification for spatial processes with a number of sources equal to the number of outputs,i.e. for determined system).

REFERENCES

ALBATAINEH, Z., AND F. SALEM (2014): " A Robust ICA Based Algorithm for Blind Separation of Convolutive Mixtures", *ArXiv*, 1408.

ANDERSON, B., DEISTLER, M., AND J.M. DUFOUR (2019): "On the Sensitivity of Granger Causality to Errors-in-Variables, Linear Transformations and Subsampling", *Journal of Time Series Analysis*, 40, 102-123.

ANGRIST, J., IMBENS, G., AND D. RUBIN (1995): "Identification of Causal Effects Using Instrumental Variables", *JASA*, 91, 444-455.

ANSLEY, C. (1977): "On the Structure of Moving Average Processes", *Journal of Econometrics*, 6, 121-134.

ARELLANO, M., AND S. BONHOMME (2018): "Recovering Latent Variables by Matching", CEMFI DP.

BABAIE-ZADEH, M. (2002): "Darmois-Skitovich Theorem and its Proof", Sharif University of Technology, DP.

BACHOC, F., GENTON, M.G., NORDHAUSEN, K. AND A. RUIZ-GAZEN (2020): "Spatial Blind Source Separation", *Biometrika*, 00, 1-20.

BARTLETT, M. (1938) : "The Characteristic Function of a Conditional Statistic", *The Journal of the London Mathematical Society*, 13, 63-67.

BATES, D. (2006) : "Maximum Likelihood Estimation of Latent Affine Processes", *Review of Financial Studies*, 19, 909-965.

BEKKER, P. (1986): "Comment on Identification in the Linear Errors-in-Variables Model", *Econometrica*, 54, 215-217.

BEN-MOSHE, D. (2018a): "Identification of Joint Distributions in Dependent Factor Models", *Econometric Theory*, 34, 134-165.

BEN-MOSHE, D. (2018b): "Linear Error-in-Variables and Dependent Factor Models", D.P. Hebrew University of Jerusalem.

BERKSON, J. (1950) : "Are There Two Regressions ?", *JASA*, 45, 164-180.

BONHOMME, S., AND J.M. ROBIN (2010): "Generalized Nonparametric Deconvolution with an Application to Earning Dynamics", *Review of Economic Studies*, 77, 491-533.

BRADLEY, W., AND W. COOK (2012): "Two Proofs of the Existence and Uniqueness of Partial Fraction Decomposition", *International Mathematical Forum*, 7, 1517-1535.

BRYC, W. (1995): "Normal Distributions: Characterizations with Applications", *Lecture Note in Statistics*, 100, Springer.

CARDOSO, J., AND B. PHAM (2004): "Optimization Issues in Noisy Gaussian ICA", in *Proceedings ICA*, Grenada, Spain.

CARLEMAN, T. (1923): "Sur les fonctions infiniment derivables", *C.R. Acad. Sci. Paris*, 177, 422-424.

CHAN, K., HO, L., AND H. TONG (2006): " A Note on Time Reversibility of Multivariate Linear Processes", *Biometrika*, 93, 221-227.

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

COMON, P., AND L. DE LATHAUWER (2010): "Algebraic Identification of Under-Determined Mixtures", Chapter 9 in Comon, P. and C., Jutten, eds., *Handbook of Blind Source Separation*, 235-272, Elsevier.

COMON, P., AND C. JUTTEN (2010): "Handbook of Blind Source Separation of Independent Component Analysis", New York Academic Press.

CRAGG, J. (1997): "Using Higher Moments to Estimate the Simple Errors-in-Variables Model", *Rand Journal of Economics*, 571-591.

DAGENAIS, M., AND D. DAGENAIS (1997): "Higher Moment Estimators for Linear Regression Models with Errors in Variables", *Journal of Econometrics*, 76, 193-225.

DARMOIS, G. (1953): "Analyse Generale des Liaisons Stochastiques: Etude Particuliere de l'Analyse Factorielle Lineaire", *Review of the International Statistical Institute*, 21, 2-8.

DELAIGLE, A., AND P. HALL (2016): "Methodology for Non-Parametric Deconvolution when the Error Distribution is Unknown", *JRSS B*, 78, 231-252.

DE LATHAUWER, L. (2008): "Blind Identification of Underdetermined Mixtures by Simultaneous Matrix Diagonalization", *IEEE Transactions on Signal Processing*, 56, 1096-1105.

ERICKSON, T., JIANG, C., AND T. WHITED (2014): "Minimum Distance Estimation of the Errors-in-Variables Model Using Linear Cumulant Equations", *Journal of Econometrics*, 183, 211-221.

ERICKSON, T., AND T. WHITED (2002): "Two-Step GMM Estimation of the Errors-in-Variables Model Using High-Order Moments", *Econometric Theory*, 18, 776-799.

ERIKSSON, J., AND V. KOIVUNEN (2004): "Identifiability, Separability and Uniqueness of Linear ICA Models", *IEEE Signal Process Lett.* , 11, 601-604.

EVDOKIMOV , K., AND H. WHITE (2012): "Some Extensions of a Lemma of Kotlarski", *Econometric Theory*, 28, 925-937.

FRISCH, R. (1934): "Statistical Confluence Analysis by Means of Complete Regression Systems", Publication 5, University of Oslo.

FULLER, W. (1987): "Measurement Error Models", New York, Wiley.

GEARY, R. (1942) : "Inherent Relations Between Random Variables", *Proc. R. Irish Acad*, 47,63-76.

GOURIEROUX, C., AND J. JASIAK (2020): "Inference for Noisy Long Run Component Processes", CREST DP.

HAAVELMO, T. (1944) : "The Probability Approach in Econometrics", *Econometrica*, 12, 1-115.

HAMEDANI, G., AND M. MAADOOLIAT (2015): "Sub-Independence: A Useful Concept", Computer Science, Technology and Applications, Hauppauge, New York.

HECKMAN, J., AND R. PINTO (2015) : "Causal Analysis After Haavelmo", *Econometric Theory*, 31, 115-151.

HOLLAND, P. (1988): "Causal Inference, Path Analysis and Recursive Structural Equation Models", Sociological Methodology, 18, 449-489.

HYVARINEN, A., KARHUNEN J., AND E. OJA (2001): "Independent Component Analysis", Wiley, New York.

KAGAN, A., LINNIK, Y., AND C. RAO (1973): "Characterization Problems in Mathematical Statistics", Wiley, New York.

KAPTEYN, A., AND T. WANSBEEK (1983): "Identification in the Linear Errors-in-Variables Model", *Econometrica*, 51, 1847-1849.

KHAN, S., MAUREL, A., AND Y. ZHANG (2019): "Informational Content of Factor Structures in Simultaneous Binary Response Models", Boston College D.P.

KLEPPER, S., AND E. LEAMER (1984): "Consistent Sets of Estimators for Regressions with Errors in All Variables", *Econometrica*, 52, 163-183.

KOOPMANS, J. (1937): "Linear Regression Analysis of Economic Time Series", Publication 20, Netherlands Economic Institute, Haarlem.

LEONARD, D. (2011): "Estimating a Bivariate Linear Relationship", Bayesian Analysis, 6, 727-754.

LEWBEL, A. (1997): "Constructing Instruments for Regression with Measurement Errors when no Additional Data are Available with an Application to Patents and R&D", *Econometrica*, 65, 1201-1213.

LIN, Z. AND H. ZHANG (2018): "Low-Rank Models in Visual Analysis", Academic Press

LINNIK, Y. (1964): "Decomposition of Probability Distributions", Oliver and Boyd.

LINTNER, J. (1965) : "Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolio and Capital Budget", *Review of Economics and Statistics*, 47, 13-37.

MADANSKY, A. (1959): "The Fitting of Straight Lines when Both Variables are Subject to Error", *JASA*, 54, 173-205.

MARAVALL, A. (1979): "Identification in Dynamic Shock-Error Models", Lecture Notes in Econometrics and Mathematical Systems, 165, Springer.

MARCINKIEWICZ, J. (1938): "Sur une propriete de la loi de Gauss", *Math. Zeitschrift*, 44, 612-618.

MEISTER, A. (2009): "Deconvolution Problems in Nonparametric Statistics", *Lecture Notes in Statistics*, Berlin, Springer.

MIETTINEN, J., NORDHAUSEN, K., OJA, H. AND S. TASKINEN (2014): "Deflation-Based Separation of Uncorrelated Stationary Time Series", *Journal of Multivariate Analysis*, 123, 214-27.

NORDHAUSEN, K, OJA, H, LILZMOSER, P. AND C. REIMANN (2015): "Blind Source Separation for Spatially Correlated Compositional Data", *Math. Geosci.*, 47, 753-70.

NOWAK, E. (1985): "Global Identification of the Dynamic Shock-Error Model", *Journal of Econometrics*, 17, 211-219.

NOWAK, E. (1989): "Identification of Simultaneous Equation Models with Measurement Errors Based on Time Series Structure", *Journal of Econometrics*, 40, 319-325.

PAVAN F., AND S. MIRANDA (2018): "On the Darmois- Skitovitch Theorem and Spatial Independence in Blind Source Separation", *Journal of Communication and Information Systems*, 33, 146-157.

PEARL, J. (2004): "Direct and Indirect Effects", in Proc. 17th Conf. Uncertainty in Artificial Intelligence, San Francisco.

PEARL, J. (2009) : "Causality : Models, Reasoning and Inference", 2nd ed., Cambridge Univ.Press.

PEARL, J. (2018) : "Reflections on Heckman and Pinto's "Causal Analysis After Haavelmo", DP UCLA.

PEDERSEN, M., LARSEN J., KJEMS, U., AND L. PARSA (2007): "A Survey of Convolutive Blind Source Separation Methods", in Springer Handbook of Speech Processing and Speech Communications, 1-34.

POIRIER, D. (1998): "Revisiting Beliefs in Nonidentified Models", *Econometric Theory*, 14, 483-509.

RAO, P. (1992): "Identifiability in Stochastic Models", Academic Press, Boston.

ROSS, S. (1976): "The Arbitrage Theory of Capital Asset Pricing", *Journal of Economic Theory*, 17, 254-286.

REIERSOL, O. (1950): " Identifiability of a Linear Relation Between Variables which are Subject to Error", *Econometrica*, 18, 375-389.

SASVARI, Z. (1986): "Characterizing the Distributions of the Random Variables X_1, X_2, X_3 by the Distributions of $(X_1 - X_3, X_2 - X_3)$", *Probability Theory and Related Fields*, 73, 43-49.

SAMUELSON, P. (1942): "A Note on Alternative Regressions", *Econometrica*, 10, 80-83.

SCHENNACH, S. (2016) "Recent Advances in the Measurement Error Literature", *Annual Review of Economics*, 8, 341-377.

SCHENNACH, S. (2019): "Convolution without Independence", *Journal of Econometrics*, 211, 308-318.

SCHENNACH, S., AND Y. HU (2013): "Nonparametric Identification and Semi-Parametric Estimation of Classical Measurement Error Models without Side Information", *JASA*, 108, 177-186.

SHARPE, W. (1964): "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk", *Journal of Finance*, 19, 425-447.

SHI, X. (2011): "Blind Signal Processing", Springer.

STEFANSKI, L. (2000): "Measurement Error Models", *JASA*, 95, 1353-1358.

STOCK, J., AND M. WATSON (2018) : "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments", *Economic Journal*, 128, 917-948.

SZEKELY, G., AND C. RAO (2000): "Identifiability of Distributions of Independent Random Variables by Linear Combinations and Moments", *Sankhya, A*, 193-202.

THI, N., AND C. JUTTEN (1995): "Blind Source Separation for Convolutive Mixtures", *Signal Processing*, 45, 209-229.

WALD, A. (1940): "The Fitting of Straight Lines when Both Variables are Subject to Error", *The Annals of Mathematical Statistics*, 12, 284-300.

WILLIAMS, B. (2020): "Identification of the Linear Factor Model", *Econometric Reviews*, 39, 92-109.

WOOLEY, E. (1941): "The Method of Minimized Errors as a Basis of Correlation Analysis", *Econometrica*, 9, 38-62.

YILMAZ, A., AND S. RICKARD (2004): "Blind Separation of Speech Mixtures via Time Frequency Masking", *IEEE Transactions on Signal Processing*, 52, 1830-1847.

ZHAO, Y., AND X. LUO (2019): "Granger Mediation Analysis of Multiple Time Series Models: An Application to SMRI", *Biometrics*, 75, 788-798.

ZINDE-WALSH, V. (2013): "Identification and Well-Posedness in Nonparametric Models with Independence Conditions", in *Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, eds. J. Racine and A. Ullah, Chapter 4, 97-126.

ZINDE-WALSH, V. (2014): "Measurement Error and Deconvolution in Spaces of Generalized Functions", *Econometric Theory*, 30, 1207-1246.

APPENDIX 1

Identification

A.1 The formula of pairwise c.g.f. under dynamic subindependence

Let us introduce the following notation:

$$\begin{aligned} c_k(u) &= logE[exp(uX_{k,t})], \\ b_k(u) &= logE[exp(u\epsilon_{k,t})],\ k=1,...,K. \end{aligned}$$

We get :

$$\begin{aligned} E[exp(uY_t + vY_{t-1})] &= E[exp(u\sum_{k=1}^{K} X_{k,t} + v\sum_{k=1}^{K} X_{k,t-1})] \\ &= \prod_{k=1}^{K} E[exp(uX_{k,t} + vX_{k,t-1})]. \end{aligned}$$

Hence, the joint c.g.f. of (Y_t, Y_{t-1}) is the sum of the joint c.g.f. of $(X_{k,t}, X_{k,t-1}), k = 1, \ldots, K$. Using straightforward notation, we get:

$$\Psi(u,v) = \sum_{k=1}^{K} \Psi_k(u,v).$$

The above decomposition is valid whenever

$$E[exp(u\sum_{k=1}^{K} X_{k,t} + v\sum_{k=1}^{K} X_{k,t-1})] = \prod_{k=1}^{K} E[exp(uX_{k,t} + vX_{k,t-1})],\ \forall u, v.$$

This condition is weaker than the condition of independence equivalent to

$$E[exp(\sum_{k=1}^{K} u_k X_{k,t} + \sum_{k=1}^{K} v_k X_{k,t-1})] = \prod_{k=1}^{K} E[exp(u_k X_{k,t} + v_k X_{k,t-1})],\ \forall u_k, v_k,\ k=1,...,K.$$

This is a condition of "dynamic subindependence" [see Hamedani, Maadooliat (2015), Schennach (2019) for the general definition of subindependence].

LEMMA 1:

$b_k(u) = c_k(u) - c_k(\rho_k u)$.

PROOF: We have

$$\begin{aligned} E[exp(uX_{k,t})|X_{k,t-1}] &= E[exp(u\rho_k X_{k,t-1} + u\epsilon_{k,t})|X_{k,t-1}] \\ &= exp[u\rho_k X_{k,t-1}]E[exp(u\epsilon_{k,t})]. \qquad (a.1) \end{aligned}$$

By taking expectations of each term, we get:

$$exp[c_k(u)] = exp[c_k(u\rho_k) + b_k(u)],$$

which implies that $b_k(u) = c_k(u) - c_k(u\rho_k)$.

LEMMA 2

$\Psi_k(u, v) = c_k(v + u\rho_k) + c_k(u) - c_k(u\rho_k)$.

PROOF: We have:

$$\begin{aligned}
E[exp(uX_{k,t} + vX_{k,t-1})] &= EE[exp(uX_{k,t} + vX_{k,t-1})|X_{k,t-1}] \\
&= E\{exp(vX_{k,t-1})E[exp(uX_{k,t})|X_{k,t-1}]\} \\
&= Eexp[(v + u\rho_k)X_{k,t-1} + b_k(u)], \\
&\quad \text{from equation (a.1)} \\
&= exp[c_k(v + u\rho_k) + c_k(u) - c_k(u\rho_k)] \\
&\quad \text{by Lemma 1.}
\end{aligned}$$

QED

It follows that:

$$\Psi(u, v) = \sum_{k=1}^{K} [c_k(v + u\rho_k) + c_k(u) - c_k(u\rho_k)]. \qquad (a.2)$$

A.2 A Change of Argument

At this point, we need to introduce the following change of arguments:

$$w_1 = v + \rho_1 u, \quad w_2 = v + \rho_2 u.$$

This change of arguments is valid for $K \geq 2$.

As $\rho_1 \neq \rho_2$, this mapping is bijective. We have:

$$u = \frac{w_2 - w_1}{\rho_2 - \rho_1} \quad v = \frac{\rho_2 w_1 - \rho_1 w_2}{\rho_2 - \rho_1},$$

and in general :

$$v + \rho u = \frac{1}{\rho_2 - \rho_1}[(\rho_2 - \rho)w_1 + (\rho - \rho_1)w_2], \text{ for any } \rho.$$

This change of arguments will be applied later on to the first-order derivative of the pairwise c.g.f..

A.3 Darmois' Lemma [Darmois (1953), p.7, Kagan et al. (1973), p.89]

The proof of Proposition 2 is based on the Darmois' Lemma. This Lemma has been initially introduced to prove the Darmois-Skitovich Theorem, and is generally employed to analyze the identification in static ICA [see e.g. Comon (1994), Lemma 20, Pavan, Miranda (2018), Lemma 1].

LEMMA 3 (Darmois) [17]

Let us assume that the following condition is satisfied :

$$\sum_{i=1}^{N} f_i(a_i u + b_i v) = g_1(u) + g_2(v), \;\; \forall u, v, \in U,$$

where $U \subset R$ is an open set including 0, functions $f_i,\ i = 1, ..., N$ are continuous and a_j, b_j are real numbers. Then, if $a_i \neq 0,\ i = 1, ..., N$ and $a_i b_j - a_j b_i \neq 0, \forall i, j, i \neq j$, the functions $f_i,\ i = 1, ..., N, g_1, g_2$ are necessarily polynomials of degree less or equal to N.

As the Darmois' Lemma is written for real arguments, we use it below either for the real c.g.f, or for the second characteristic function by distinguishing its real and imaginary components.

In the next part of Appendix 1, the identification is proven for real c.g.f under Assumption A.3 ii). The proof for the second characteristic function would be similar, except for the real and imaginary components separately considered.

A.4 The identification for $K = 2, 3$

Due to the existence of first and second-order moments, the pairwise joint c.g.f. of Y_t, Y_{t-1} is differentiable [18]. Its first-order derivative with respect to v is:

$$\frac{\partial \Psi}{\partial v}(u, v) = \sum_{k=1}^{K} c_k'(v + u\rho_k), \; u, v, \subset U \in R.$$

Let us assume that $\gamma_k,\ k = 1, ..., K$, are other candidates for the c.g.f. of $(X_{kt}), k = 1, ..., K$. As function Ψ and its partial derivative are identifiable, we get:

$$\sum_{k=1}^{K} c_k'(v + u\rho_k) = \sum_{k=1}^{K} \gamma_k'(v + u\rho_k) \;\; \forall u, v \in U \subset R.$$

[17] A proof of the Darmois Lemma under differentiability conditions can be found in Babaieh, Zadeh (2002), or in Comon, De Lathauwer (2010).

[18] As mentioned in the main text, the approach below is also valid with the second characteristic function, where by definition $\frac{\partial \Psi}{\partial v}(u, v) \equiv \frac{1}{\Phi(u,v)} \frac{\partial \Phi(u,v)}{\partial v}$, with $\Phi(u, v)$ as the first characteristic function, and similar definitions for the derivatives of the c_k functions.

Then we can apply the change of arguments of Appendix A.2 and write:
$c_1'(w_1) - \gamma_1'(w_1) + c_2'(w_2) - \gamma_2'(w_2) + \sum_{k=3}^K \{c_k'(\frac{1}{\rho_2-\rho_1}[(\rho_2-\rho_k)w_1 + (\rho_k-\rho_1)w_2]$
$-\gamma_k'(\frac{1}{\rho_2-\rho_1}[(\rho_2-\rho_k)w_1 + (\rho_k-\rho_1)w_2]\} = 0,$
where the last sum disappears if $K = 2$.

The conditions of the Darmois Lemma are satisfied and then the differences $c_k' - \gamma_k'$ are polynomials of degree less or equal to $N = K - 2$.

i) If $K = 2$, $c_k'(u) - \gamma_k'(u) = \alpha$, where α is a constant, and by integration $\gamma_k(u) = c_k(u) + \alpha u + \beta$, say.

However $\alpha = \beta = 0$, since $\gamma_k(0) = c_k(0) = \gamma_k'(0) = c_k'(0) = 0$ (the condition on the first-order derivative is equivalent to the constraint of zero mean).

ii) If $K = 3$, $c_k'(u) - \gamma_k'(u)$ is a polynomial of degree 1, and by integration $\gamma_k(u) = c_k(u) + \alpha u^2 + \beta u + \delta$, say. But $\alpha = \beta = \delta = 0$, since $\gamma_k(0) = c_k(0) = \gamma_k'(0) = c_k'(0) = 0$ and $\gamma_k''(0) = c_k''(0) = \sigma_k^2/(1-\rho_k^2)$ is identifiable.

A.5 The identification for $K \geq 4$.

When $K \geq 4$, the functions c_k, γ_k differ by a polynomial of degree higher or equal to 3, and the knowledge of the first and second order derivatives at 0 is not sufficient to prove that $\gamma_k(u) - c_k(u) = 0, \ \forall k, u$.

When $K = 4$, the same reasoning as in Appendix A.4 shows that the difference $\gamma_k - c_k$ is of the form:

$$\gamma_k(u) - c_k(u) = \delta_k u^3, \text{ say,}$$

where $\delta_k, \ k = 1, ..., 4$ are scalars. By considering the identification restriction applied to the joint c.g.f. (a.2), we infer that the unknowns $\delta_k, k = 1, ..., 4$ are such that:

$$\sum_{k=1}^4 \{\delta_k[(v + u\rho_k)^3 + u^3 - u^3\rho_k^3]\} = 0, \ \forall u, v \in U \subset R. \qquad (a.3)$$

That leads to a homogeneous system of 3 restrictions:

$$\sum_{k=1}^4 \delta_k = 0, \ \sum_{k=1}^4 \delta_k\rho_k = 0, \ \sum_{k=1}^4 \delta_k\rho_k^2 = 0. \qquad (a.4)$$

This is insufficient to prove that the coefficients $\delta_k, k = 1, ..., 4$ are 0.

To get additional restrictions, we also have to consider nonlinear serial dependence at higher lags.

i) If all $|\rho_k|$ are distinct, that is, if does not exist j, k such that $\rho_j = -\rho_k$, then we can use the pairwise dependence based on (Y_t, Y_{t-2}). Since

$$Y_t = \sum_{k=1}^{K} X_{k,t},$$

where $X_{k,t} = \rho_k^2 X_{k,t-2} + \tilde{\epsilon}_{k,t}$, with $\tilde{\epsilon}_{k,t} = \epsilon_{k,t} + \rho_k \epsilon_{k,t-1}$, we get a condition similar to condition (a.3) in which ρ_k is replaced by ρ_k^2.

We deduce another set of restrictions similar to (a.4), that are:

$$\sum_{k=1}^{4} \delta_k = 0, \; \sum_{k=1}^{4} \delta_k^2 \rho_k^2 = 0, \; \sum_{k=1}^{4} \delta_k \rho_k^4 = 0, \qquad (a.5)$$

that is an additional restriction $\sum_{k=1}^{4} \delta_k \rho_k^4 = 0$. Since the matrix:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ \rho_1 & \rho_2 & \rho_3 & \rho_4 \\ \rho_1^2 & \rho_2^2 & \rho_3^2 & \rho_4^2 \\ \rho_1^4 & \rho_2^4 & \rho_3^4 & \rho_4^4 \end{pmatrix}$$

is of full rank (since the $|\rho_k|$ are distinct), we deduce $\delta_k = 0, k = 1, ..., 4$, and the identification follows.

ii) If two $|\rho_k|$ are equal, we cannot base the result on the Darmois Lemma applied to the pairwise distribution of (Y_t, Y_{t-2}). However, by considering the pairwise dependence based on (Y_t, Y_{t-3}), we know that all ρ_k^3 are distinct and get an additional restriction $\sum_{k=1}^{4} \delta_k \rho_k^3 = 0$. The homogenous system is of full rank (see Appendix 2) and we get again $\delta_k = 0, \; k = 1, ..., 4$.

In general, if $K > 4$, the same reasoning shows that the differences $\gamma_k(u) - c_k(u)$ are necessarily of the form $\gamma_k(u) - c_k(u) = \sum_{j=3}^{K-1} \delta_{kj} u^j$. By applying the identification based on the joint c.g.f. (a.1), we derive that for each $j, j = 3, ..., K-1$, the information on the distribution of (Y_t, Y_{t-1}) provides $K-1$ restrictions on the $\delta_{kj}, k = 1, ..., K$. An additional restriction is obtained as for $K = 4$ by considering either the pairwise distribution of (Y_t, Y_{t-2}) or the pairwise distribution of (Y_t, Y_{t-3}).

APPENDIX 2
Full Rank of the System

Let us consider the (K,K) matrix:

$$C = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \rho_1 & \rho_2 & \dots & \rho_K \\ \vdots & \vdots & \vdots & \vdots \\ \rho_1^{K-1} & \rho_2^{K-1} & \dots & \rho_K^{K-1} \end{pmatrix}.$$

Lemma: The matrix C is invertible if and only if $\rho_l \neq \rho_k$ for any $k \neq l$.

Proof: The determinant of C is a polynomial in $\rho_1, ..., \rho_K$ of degree $1+2+\ldots+(K-1) = K(K-1)/2$. Moreover, if $\rho_k = \rho_l$, two columns of matrix C are equal and $det\ C = 0$. It follows that the polynomial determinant of C is divisible by $\rho_k - \rho_l$ for any pair $k \neq l$. Since there are $K(K-1)/2$ such pairs, $det\ C$ is equal to $\prod_{k<l}(\rho_k - \rho_l)$ up to a multiplicative factor, by the fundamental theorem of algebra. Therefore, it is different from 0, if and only if $\rho_k \neq \rho_l$ for any $k \neq l$.

QED

APPENDIX 3

Sum of MA(1) Processes

Let us consider two independent MA(1) processes: $X_{jt} = \epsilon_{jt} + \theta_j \epsilon_{j,t-1},\ |\theta_j| < 1$, $\theta_j \neq 0, j = 1, 2$, where (ϵ_{jt}) is a sequence of i.i.d. variables. As noted in Section 2, their sum $X_t = \sum_{j=1}^{2} X_{jt}$ can be written as another MA(1) process : $X_t = \epsilon_t + \theta \epsilon_{t-1},\ |\theta| < 1$, where (ϵ_t) is a weak white noise with variance σ^2. The following Lemma characterizes the case where (ϵ_t) is a sequence of i.i.d. variables:

Lemma: Let us assume $\theta_1 < \theta_2$ and the existence of second-order moments . The (ϵ_t) is a sequence of i.i.d. variables, if and only if, all noises $(\epsilon_{jt}),\ j = 1, 2$ are Gaussian. Then (ϵ_t) is also Gaussian.

Proof:

Let us denote the c.g.f. of $\epsilon_{jt},\ j = 1, 2$ (resp. of ϵ_t) by $c_j(u),\ j = 1, 2$ (resp. $c(u)$). The joint c.g.f. of (X_t, X_{t-1}) is :

$$\begin{aligned} \Psi(u, v) &= \log E exp(uX_t + vX_{t-1}) \\ &= c(u) + c(u\theta + v) + c(v\theta) \\ &= \sum_{j=1}^{2} [c_j(u) + c_j(u\theta_j + v) + c_j(v\theta_j)],\ \forall u, v. \end{aligned}$$

By differentiating the last equality first with respect to u and then with respect to v, we get:

$$\theta c"(u\theta + v) = \sum_{j=1}^{2} c_j"(u\theta_j + v), \forall u, v.$$

Therefore, by applying the Darmois' Lemma, we find that all second-order derivatives $c_j"(u)$, $j = 1, 2$, and $c"(u)$ are polynomials of degree less or equal to 2, whenever $\theta \neq \theta_1, \theta \neq \theta_2$. Then, the c.g.f.'s are polynomials (of degree less or equal to 4), which implies that the distributions are Gaussian [Marcinkiewicz Theorem (Marcinkiewicz (1938), Bryc (1995), Th 2.5.3)]. The fact that θ is both different from θ_1 and θ_2 follows from the formulas of $VX_t, Cov(X_t, X_{t-1})$ used to derive θ:

$$\begin{aligned}\frac{Cov(X_t, X_{t-1})}{Var(X_t)} \quad &= \quad \frac{\theta}{1+\theta^2} = \frac{\sigma_1^2\theta_1 + \sigma_2^2\theta_2}{\sigma_1^2(1+\theta_1^2) + \sigma_2^2(1+\theta_2^2)} \\ \iff \quad & \frac{\theta}{1+\theta^2} = \frac{\sigma_1^2(1+\theta_1^2)}{\sigma_1^2(1+\theta_1^2) + \sigma_2^2(1+\theta_2^2)}\frac{\theta_1}{1+\theta_1^2} + \frac{\sigma_2^2(1+\theta_2^2)}{\sigma_1^2(1+\theta_1^2) + \sigma_2^2(1+\theta_2^2)}\frac{\theta_2}{1+\theta_2^2},\end{aligned}$$

which shows that $\frac{\theta}{1+\theta^2} \in (\frac{\theta_1}{1+\theta_1^2}, \frac{\theta_2}{1+\theta_2^2})$. Therefore, $\theta \in (\theta_1, \theta_2)$, since $\theta \to \frac{\theta}{1+\theta^2}$ is strictly increasing on [-1, 1]. QED

APPENDIX 4
Asymptotic Theory

Since the nonparametric estimators of the c.g.f.'s of the sources have closed form expressions involving the estimated pairwise c.g.f. of (Y_t, Y_{t-h}), their asymptotic properties can be derived from the asymptotic properties of the empirical counterpart of these pairwise c.g.f. by applying the δ-method. In particular, if the sources are independent [19], they are pointwise convergent at speed $1/\sqrt{T}$ and asymptotically normal as the assumption of stationary AR(1) dynamics of sources implies the geometric ergodicity of the observable process.

i) Estimation of the real c.g.f., $K = 2$

For illustration and to point out possible simplifications in the formula of asymptotic variance, let us consider the real c.g.f. with $K = 2$. In this case, the estimators are

[19] Under the weaker condition of subindependence, additional regularity conditions for asymptotic inference have to be introduced.

computed from the empirical partial derivatives of the pairwise c.g.f. with respect to v. We have:

$$\frac{\partial \hat{\Psi}}{\partial v}(u,v) = \frac{\sum_t [Y_{t-1} exp(uY_t + vY_{t-1})]}{\sum_t exp(uY_t + vY_{t-1})} = \frac{\sum_t A_t}{\sum_t D_t}, \text{ say.}$$

By the δ-method, we find that:

$$\sqrt{T}\left[\frac{\partial \hat{\Psi}}{\partial v}(u,v) - \frac{\partial \Psi}{\partial v}(u,v)\right] \to N(0, \sigma^2(u,v)),$$

where

$$\sigma^2(u,v) = \left[\frac{1}{ED}, -\frac{EA}{(ED)^2}\right] \sum_h \Gamma(h) \left[\frac{1}{ED}, -\frac{EA}{(ED)^2}\right]',$$

$$\Gamma(h) = Cov\left[\begin{pmatrix} A_t \\ D_t \end{pmatrix}, \begin{pmatrix} A_{t-h} \\ D_{t-h} \end{pmatrix}\right] = E\left[\begin{pmatrix} A_t \\ D_t \end{pmatrix}(A_{t-h}, D_{t-h})\right] - \begin{pmatrix} EA \\ ED \end{pmatrix}(EA, ED).$$

It follows directly from the expression of $\sigma^2(u,v)$ that we can disregard the outer product of expectations in the computation of $\sigma^2(u,v)$. Thus, $\Gamma(h)$ can be replaced by $E\left[\begin{pmatrix} A_t \\ D_t \end{pmatrix}(A_{t-h}, D_{t-h})\right]$ only.

The term associated with $\Gamma(h)$ involves the quadruple distributions of $Y_t, Y_{t-1}, Y_{t-h}, Y_{t-h-1}$, for each h.

Let us derive the expression for $\Gamma(0)$. The part of $\sigma_0^2(u,v)$ corresponding to the term $\Gamma(0)$ is:

$$\sigma_0^2(u,v) = \left[\frac{1}{ED}, -\frac{EA}{(ED)^2}\right] \begin{pmatrix} E(A^2) & E(AD) \\ E(AD) & E(D^2) \end{pmatrix} \left[\frac{1}{ED}, -\frac{EA}{(ED)^2}\right]',$$

in which the time index is omitted due to stationarity. Let us now denote the (real) moment generating function by $\Phi(u,v) = E[exp(uY_t + vY_{t-1})]] = exp[\Psi(u,v)]$. We have:

$$EA = \frac{\partial \Phi(u,v)}{\partial v},\ ED = \Phi(u,v),\ E(A^2) = \frac{\partial^2 \Phi(2u,2v)}{\partial v^2},$$

$$E(AD) = \frac{\partial \Phi(2u,2v)}{\partial v},\ E(D^2) = \Phi(2u,2v).$$

where $\frac{\partial^2 \Phi(2u,2v)}{\partial v^2}$, (resp. $\frac{\partial \Phi(2u,2v)}{\partial v}$) denotes $\frac{\partial^2 \Phi(u,v)}{\partial v^2}|_{2u,2v}$ (resp. $\frac{\partial \Phi(u,v)}{\partial v}|_{2u,2v}$).

We deduce:

$$\sigma_0^2(u,v) = \frac{1}{[\Phi(u,v)]^2}\left\{\frac{\partial^2 \Phi(2u,2v)}{\partial v^2} - 2\frac{\partial \Phi(u,v)}{\partial v}[\Phi(u,v)]^{-1}\frac{\partial \Phi(2u,2v)}{\partial v} + \left[\frac{\partial \Phi(u,v)}{\partial v}[\Phi(u,v)]^{-1}\right]^2 \Phi(2u,2v)\right\}.$$

Since :

$$\frac{1}{\Phi^2}\frac{\partial^2 \Phi}{\partial v^2} = \frac{\partial^2 \log \Phi}{\partial v^2} + \left(\frac{\partial \log \Phi}{\partial v}\right)^2 = \frac{\partial^2 \Psi}{\partial v^2} + \left(\frac{\partial \Psi}{\partial v}\right)^2,$$

we can rewrite the expression of $\sigma_0^2(u,v)$ in terms of the pairwise c.g.f and its partial derivative. We get:

$$\begin{aligned}
\sigma_0^2(u,v) &= \exp[\Psi(2u,2v) - 2\Psi(u,v)] \\
&\quad \left\{\frac{\partial^2\Psi(2u,2v)}{\partial v^2} + \left[\frac{\partial\Psi(2u,2v)}{\partial v}\right]^2 - 2\frac{\partial\Psi(u,v)}{\partial v}\frac{\partial\Psi(2u,2v)}{\partial v} + \left[\frac{\partial\Psi(u,v)}{\partial v}\right]^2\right\} \\
&= \exp[\Psi(2u,2v) - 2\Psi(u,v)]\left\{\frac{\partial^2\Psi(2u,2v)}{\partial v^2} + \left[\frac{\partial\Psi(2u,2v)}{\partial v} - \frac{\partial\Psi(u,v)}{\partial v}\right]^2\right\}.
\end{aligned}$$

Each component in the formula above can be consistently estimated by its sample counterpart.

ii) Estimation of the second characteristic function $K = 2$

A similar derivation can be performed for the estimated second characteristic function. However, while the computation of derivatives can be done numerically in the complex space, real space computations are necessary to derive the asymptotic distribution. Below, we explain how to write explicitly the real and imaginary components of $\frac{\partial \psi(u,v)}{\partial v}$ in terms of moments. Below, i denotes the imaginary root of -1. We have :

$$\begin{aligned}
\frac{\partial \psi}{\partial v}(u,v) &= \frac{\partial}{\partial v}\log E[\exp(iuY_t + ivY_{t-1})] \\
&= \frac{E[iY_{t-1}\exp(iuY_t + ivY_{t-1})]}{E[\exp(iuY_t + ivY_{t-1})]} \\
&= \frac{iE[Y_{t-1}\cos(uY_t + vY_{t-1})] - E(Y_{t-1}\sin(uY_t + vY_{t-1})]}{E[\cos(uY_t + vY_{t-1})] + iE[\sin(uY_t + vY_{t-1})]} \\
&= \tfrac{1}{M}\{iE[Y_{t-1}\cos(uY_t + vY_{t-1})] - E[Y_{t-1}\sin(uY_t + vY_{t-1})]\} \\
&\quad \{E[\cos(uY_t + vY_{t-1})] - iE[\sin(uY_t + vY_{t-1})]\}.
\end{aligned}$$

where $M = [Ecos(uY_t + vY_{t-1})]]^2 + [Esin(uY_t + vY_{t-1})]^2$. It follows that :

$$\begin{aligned} Re\,\frac{\partial\psi}{\partial v}(u,v) &= \tfrac{1}{M}\{E[Y_{t-1}\cos(uY_t+vY_{t-1})]E[\sin(uY_t+vY_{t-1})] \\ &- E[Y_{t-1}\sin(uY_t+vY_{t-1})]E[\cos(uY_t+vY_{t-1})]\}, \end{aligned}$$

$$\begin{aligned} \mathcal{I}m\,\frac{\partial\psi}{\partial v}(u,v) &= \tfrac{1}{M}\{E[Y_{t-1}\cos(uY_t+vY_{t-1})]E[\cos(uY_t+vY_{t-1})] \\ &+ E[Y_{t-1}\sin(uY_t+vY_{t-1})]E[\sin(uY_t+vY_{t-1})]\}. \end{aligned}$$

These real and imaginary components can be estimated by replacing each theoretical moment by its sample counterpart.

APPENDIX 5

Direct solution for $K = 4$

When the coefficients $|\rho_k|$ are distinct and the moments exist up to order 3, the technique introduced in Section 5.2 can be extended as follows:

i) Let us extend eq. (5.4) for $K = 4$. We get:

$$\begin{aligned} A(w_1,w_2) = c_1'(w_1) &+ \frac{\rho_2-\rho_3}{\rho_2-\rho_1}c_3''[\frac{\rho_2-\rho_3}{\rho_2-\rho_1}w_1+\frac{\rho_3-\rho_1}{\rho_2-\rho_1}w_2] \\ &+ \frac{\rho_2-\rho_4}{\rho_2-\rho_1}c_4''[\frac{\rho_2-\rho_4}{\rho_2-\rho_1}w_1+\frac{\rho_4-\rho_1}{\rho_2-\rho_1}w_2], \end{aligned}$$

where $A(w_1, w_2)$ denotes the left-hand side of eq. (5.4). Then, by considering the partial derivative with respect to w_2, we obtain:

$$\begin{aligned} \frac{\partial A(w_1,w_2)}{\partial w_1} &= \frac{(\rho_2-\rho_3)(\rho_3-\rho_1)}{(\rho_2-\rho_1)^2}c_3^{(3)}[\frac{\rho_2-\rho_3}{\rho_2-\rho_1}w_1+\frac{\rho_3-\rho_1}{\rho_2-\rho_1}w_2] \\ &+ \frac{(\rho_2-\rho_4)(\rho_4-\rho_1)}{(\rho_2-\rho_1)^2}c_4^{(3)}[\frac{\rho_2-\rho_4}{\rho_2-\rho_1}w_1+\frac{\rho_4-\rho_1}{\rho_2-\rho_1}w_2], \end{aligned}$$

where $c_k^{(3)}$ denotes the third-order derivative of c_k. By choosing carefully the pairs w_1, w_2, we get $c_3^{(3)}(u), c_4^{(3)}(u)$ as functions of coefficients ρ_k, of the associated values of partial derivatives and of $c_3^{(3)}(0), c_4^{(3)}(0)$.

ii) The last step consists in finding the values of $c_3^{(3)}(0), c_4^{(3)}(0)$.
For $w_1 = w_2 = 0$, we get:

$$\frac{\partial A(0,0)}{\partial w_1} = \frac{(\rho_2 - \rho_3)(\rho_3 - \rho_1)}{(\rho_2 - \rho_1)^2} c_3^{(3)}(0) + \frac{(\rho_2 - \rho_4)(\rho_4 - \rho_1)}{(\rho_2 - \rho_1)^2} c_4^{(3)}(0). \tag{A.1}$$

A similar equation can be written for the pairwise distribution of Y_t, Y_{t-2}. With obvious notation, we get:

$$\frac{\partial A_2(0,0)}{\partial w_1} = \frac{(\rho_2^2 - \rho_3^2)(\rho_3^2 - \rho_1^2)}{(\rho_2^2 - \rho_1^2)^2} c_3^{(3)}(0) + \frac{(\rho_2^2 - \rho_4^2)(\rho_4^2 - \rho_1^2)}{(\rho_2^2 - \rho_1^2)^2} c_4^{(3)}(0). \tag{A.2}$$

The values $c_3^{(3)}(0), c_4^{(3)}(0)$ are the solutions of equations (A.1) and (A.2).