Dynamic Deconvolution of Independent AR(1) Sources

Christian Gourieroux * and Joann Jasiak †

December 29, 2018

Abstract

We consider a multivariate system $Y_t = AX_t$, where the unobserved components (the sources) X_t are 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 the sources can be identified. The proof is constructive and it is used to introduce simple consistent estimators of all unknown scalar and functional parameters of the model. Economic applications to causal models with structural innovations are also discussed, such as the identification in error-in-variables models, the predictability puzzle and the impulse response functions.

Keywords: Identification, Independent Component Analysis, Blind Source Separation (BSS), Convolutive BSS, Deconvolution, Error-in-Variables Model, Causal Model, Predictability.

The authors thank D. Ben Moshe, M. Deistler, 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.

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 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 and their dynamics needs to be identified (convolutive mixtures) and/or the number of sources is larger than the number of observed outputs $K \geq L$ (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. In the undetermined convolutive system considered, the sources are autoregressive processes of order 1 (AR(1)).

For ease of exposition, let us first assume that there is a single output $L = 1$ [1] and that the observed univariate time series (Y_t) can be written as the following sum of K component series (sources):

$$Y_t = \sum_{k=1}^{K} \left(\frac{1}{1-\rho_k B} \epsilon_{kt} \right) \equiv \sum_{k=1}^{K} X_{kt}, \; 0 \leq \rho_k < 1, \; \forall k, \tag{1.1}$$

where the K sequences $(\epsilon_{kt}), k = 1, \ldots, K$, are strong white noises, which are mutually independent, with mean 0 and finite variances σ_k^2, $k = 1, ..., K$, and B is the lag operator.

We demonstrate that the following can be identified:

i) the autoregressive coefficients ρ_k, $k = 1, ..., K$;, and

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

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

Let us first introduce some simple identification restrictions. If, for example $\rho_1 = \rho_2$, we have:

[1] This assumption will later be relaxed.

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

where $\tilde{\epsilon}_t = \epsilon_{1t} + \epsilon_{2t}$, and we cannot distinguish the pair $(\epsilon_{1t}, \epsilon_{2t})$ from the pair $(\tilde{\epsilon}_t, 0)$.

Identification Assumption A.1

i) The distribution of ϵ_{kt} is not degenerate point mass at 0.

ii) The K autoregressive coefficients are distinct.

Another identification problem is to distinguish 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 an increasing order: $\rho_1 < \rho_2 < \ldots < \rho_K$.

The autoregressive coefficients can be of any sign and one of them can be equal to 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. We first discuss the identification of parameters $\rho_1, ..., \rho_K$, $\sigma_1^2, \ldots, \sigma_K^2$, and next, the identification of K sources distributions given the parameters.

The paper is organized as follows. In Sections 2 to 3, the one-output (univariate) model (1.1) is considered. In Section 2, we prove the second-order identification of parameters $\rho_k, \sigma_k^2,\ k = 1, ..., K$. 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 the pairwise cumulant generating functions (c.g.f) without assuming the non-Gaussianity of sources. In Section 4, the identification results are extended to systems with any number of observed outputs and sources. Section 5 introduces a simple non-parametric estimation method for the distributions of sources. Section 6 is focused on economic applications. We discuss the errors-in-variables model, the predictability puzzle as well as the analysis of structural

[2] See Section 6.2. for a more detailed discussion.

shocks, impulse response functions and nonlinear Granger causality. Section 7 concludes. The mathematical proofs are gathered in Appendices.

2 Identification of the autoregressive coefficients and sources variances

The parameters $\rho_k, \sigma_k, k = 1, ..., K$ 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:

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

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

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

Proof: Equation (2.1) is a partial fraction decomposition of the spectral density considered as a rational function of $exp(iw)$ (also called the transfer function). Then, the identification of parameters $\rho_k, \sigma_k^2, k = 1, ..., K$ 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)].
QED

This means that these parameters can be consistently estimated by maximizing a Gaussian Pseudo-Maximum Likelihood (PML). The Gaussian PML estimator can be obtained by writing the model in a state-space form and maximizing the Gaussian pseudo log-likelihood function by the Kalman filter. This procedure provides consistent estimates of parameters $\rho_k, \sigma_k^2,\ k = 1, ..., K$.

When K is not too large, an alternative approach is to apply the Box-Jenkins approach for univariate process with a single weak white noise, viewed as a weak ARMA(p,q). More precisely, we write:

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

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

The numerator is a sum of independent moving average processes of order $K-1$, the autoregressive polynomial is of order K, if $\rho_k \neq 0, \forall k$, and of order $K-1$, otherwise. The sum of these moving average processes 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) has uncorrelated components $u_t, u_{t-1}, ...,$ that are mutually 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. The parameters $\theta_j,\ j = 1, ..., K-1,\ , \rho_k, k = 1, ..., K$, and the variance $\sigma^2 = Var(u_t)$ can be estimated by applying the Box-Jenkins Gaussian likelihood-based estimation method for univariate ARMA(p,q) processes. Then, the estimators of $\sigma_k^2, \rho_k, k = 1, ..., K$ are derived 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 the linear filter is a consequence of the AR(1) assumption and of the uniqueness of a partial fraction decomposition. As such, it is also valid if additional strong ARMA(2,1) sources, like:

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

are introduced, where the roots of the AR lag polynomial are complex conjugates, or for strong AR(p) sources with multiple roots $(1-\rho B)^p X_t = \epsilon_t$. If the sources were finite-order MA processes, such as MA(1), the moving average coefficients would not be second-order identifiable. In such other cases, the coefficients can be identified under the additional assumption of sources being independent and non-Gaussian, by using the third order 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].

3 Identification of sources distributions from pairwise nonlinear dependence

Les us now consider the identification of sources distributions given the parameters ρ_k, σ_k^2, $k = 1, ..., K$. For this purpose, we consider the information contained in the pairwise distributions of $(Y_t, Y_{t-h}), h = 1, 2, ...$ This is the analogue of identification based on autocovariances $Cov(Y_t, Y_{t-h}), h = 1, 2...$, usually employed in second-order analysis. However, such second-order analysis does not suffice to identify the distributions of sources. Let us instead introduce a dynamic identification approach for any number K of sources.

First, let us recall the approach introduced in Bonhomme, Robin (2010), who write the bivariate vector $(Y_t, Y_{t-1})'$ as:

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

to link linearly the 2-dimensional output to the $2K$ independent inputs, and apply the static Independent Component Analysis (ICA). This approach developed in Bonhomme, Robin (2010) is valid for a number of inputs less than $\frac{L(L+1)}{2}$, where L is the number of outputs. In the one output case, this upper bound is equal to 3, and the approach can be used if $2K \leq 3$, i.e. $K \leq 1$, which is practically irrelevant.

To extend the identification analysis to the general case of any number of sources, further assumptions are introduced on the dynamics of sources $X_{k,t}$, or equivalently, on the marginal distributions of errors $\epsilon_{k,t}$:

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

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 a possibly complex argument. When u is pure imaginary, equation (3.3) defines the second characteristic function. When u is real, the existence of the real c.g.f. is required.

Assumption A.3: 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 $(X_{k,t}), k = 1, ..., K$.

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

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

It is easy to see that the c.g.f. b_k of the error ϵ_k is directly linked to the c.g.f c_k of the source X_k by (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 (resp. Assumptions A.1, A.2, A.3), the distributions of sources $(X_{k,t})$ and of errors $(\epsilon_{kt}),\ k = 1, ..., K$ are identifiable :

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

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

Proof: see Appendices A.3, A.4, A.5.

QED

This identification is obtained without the assumption of non-Gaussianity required in the static ICA framework [see, e.g. Comon (1994), Th.11]. This is due to the knowledge of the autoregressive dynamic structure. The fact that pairwise nonlinear dependence is sufficient to identify is not surprising, as it also holds in the static ICA [Comon (1994), Th 11]. The proof implies that the assumption of independence of the noises in model (1.1) can be replaced by a weaker assumption of subindependence [see, Shennach (2013)].

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) is identifiable, we easily derive the distributions of $\epsilon_{1t} = Y_t - \rho_1 Y_{t-1}$. This is the special case in which we identify not only the distribution of ϵ_{1t}, but also shocks ϵ_{1t} themselves. When there are more sources than observations, the relation between sources and observations cannot be inverted, in general[4], to find the values of sources without ambiguity .

The result in Proposition 2 can be interpreted in terms of deconvolution. Let us consider $K = 2$ for ease of exposition. The deconvolution concerns the bivariate density $f(y_t, y_{t-1}) = f_1(x_{1t}, x_{1,t-1}) * f_2(x_{2t}, x_{2,t-1})$, say. 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 to prove that the deconvolution is feasible.

4 Multivariate system

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

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

or

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

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

A given row of the system can be written:

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

where the $(\tilde{\epsilon}_{jk,t})$ are independent sequences of i.i.d. variables with mean 0. Thus, we can apply Proposition 2 for any fixed row 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

[4] except if the distributions of the $(\epsilon_{k,t})$ have sufficiently different supports [see the discussion in Yilmaz, Rickard (2003)].

$a_{l,k} \neq 0$). 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. To derive a general identification result, let us now introduce the assumption of saturated mixing matrix.

Assumption A.4: Saturated mixing matrix

The elements of A are different from 0.

Then we obtain :

Proposition 3 : Under Assumptions A.1, A.2, 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 $\rho_k,\ k = 1, ..., K$ are different from 0, each column of A is identifiable up to a scale effect. 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 $\alpha' Y_t$, say, 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 the linear combinations $\alpha' \eta_t$, for any α, the second result ii) follows.

QED

Corollary 1: Uniqueness:

Under the assumptions of Proposition 3, it is possible to identify the mixing matrix A, the autoregressive parameters $\rho_1, ..., \rho_K$, the distributions of the sources (X_{kt}) or (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.

As for the single output system, the above identification is proven under weaker assumptions than in the existing literature. It is valid for any number K of sources and without strong restrictions on the distributions of sources. The result in Proposition 3 can be compared to the standard conditions for identification of sources assumed serially independent. These conditions are:

a) at most one source distribution is Gaussian [Comon (1994), Hyvarinen, Karhunen, Oja (2001), Eriksson, Koivunen (2004) Th 5. iii), Chan, Ho, Tong (2006), Ben-Moshe (2018)], b) the moments exist up to order three, four, or six [Comon, De Lathauwer (2010)], c) the characteristic functions of the sources have no exponential factor [Eriksson, Koivunen (2004) Th 5, ii), iv)], d) the restrictions on the support or exclusion restrictions are imposed [Ben-Moshe (2018a), Williams (2018)]. Moreover the number of sources can be upper bounded [De Lathauwer (2008), Bonhomme, Robin (2010)], or Bayesian approaches can be used [5] [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 (2008), (2010), Ben Moshe (2018b), Section 3], the asymptotics are different as $N \to \infty$, T is fixed in their case (where N denotes the cross-sectional dimension of the panel), while L is fixed, $T \to \infty$ 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 [Schennach (2016), Section 3.1, Zinde-Walsh (2014)].

Our approach for proving the identifiability is different from those described in the literature on the BSS and ICA. Those approaches consider first the identifiability of the dynamic mixing coefficients (A and ρ_k's in our framework) 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, followed by the identification of sources distributions and of the instantaneous mixing parameters in A. 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 $a_{jk}^2, \sigma_{jk}^2,\ j = 1, ..., L,\ k = 1, ..., K$, thus the $|a_{jk}|$ under Assumption A.4, but not the signs of these mixing coefficients. To identify the signs, information on a pair of outputs is needed.

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

The assumptions 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. The example below is especially insightful in this respect.

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. We are interested in identifying the value of β, that characterizes the "causal" effect of the treatment, and has to be distinguished from the "association" effect from the common latent factor X_t. If the unobserved X_t follows an AR(1) process, the identification of β is obtained as follows : First, we write the associated reduced form :

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

Next, we consider different combinations of observed variables as in Szekely, Rao(2000): i) Proposition 2 is used to identify the distribution of u_{1t} from the first equation; ii) Proposition 2 is used to identify the distribution of $v_t = \beta u_{1t} + u_{2t}$ from the second equation; iii) Proposition 2 is used to identify the distribution of $w_t = (1+\beta)u_{1t} + u_{2t}$ from the equation for $D_t + Y_t$. iv) Then 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 applying the deconvolution of the distribution of v_t, with known distribution of βu_{1t}. It follows that this dynamic treatment effect is identified without introducing any external instrumental variable.

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

5 Nonparametric Estimation of Sources Distributions

As the scalar and functional parameters are identified (see Corollary 1), we can introduce consistent estimation methods. When the noise distributions are from parametric families, the associated parameters can be estimated by the Simulated Maximum Likelihood, for example, in order to avoid integrating out the latent sources. We can also consider non-parametric estimation methods to estimate the sources distributions [see e.g. Bonhomme, Robin (2010) in a similar framework]. In our special case of AR(1) sources, it is possible to get analytical formulas of some derivatives of the c.g.f. c_k , by following the proof of Proposition 2 in Appendix 1, and to use these formulas for estimation purpose. This is especially easy to do, if K=2, or 3. We describe in detail the estimators in these two cases. The extension to the case $K \geq 4$ requires the estimation of addditional parameters to recover a low degree polynomial component in the c.g.f (see Appendix 1).

5.1 Estimation for K=2

As shown in Appendix 1, A.5, 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: The 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:

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

that 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 easily consistently estimated by 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 joint real c.g.f. of (Y_t, Y_{t-1}) as well as for the joint characteristic function (with u, v replaced by iu, iv). Thus, the estimator of $\Psi(u,v)$ may be obtained either from the estimator of $E\cos(uY_t + vY_{t-1})$ and $E\sin(uY_t + vY_{t-1})$, or, if Assumption A.3. is satisfied, from the estimator of $Eexp(uY_t + vY_{t-1})$ with real u, v.

The above estimation method can be extended to include the estimation of the error density of ε_{1t}, for example. It 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 : $\hat{c}_1(iw) = \int_0^w \hat{c}_1'(iu)du$.

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

step 4 : 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 the joint real c.g.f. and the second characteristic function depends on the existence of these functions. The real c.g.f. is suitable for non-negative processes with negative arguments and the second characteristic function is suitable for real variables, 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 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 $w_1 = -\frac{\rho_3 - \rho_1}{\rho_2 - \rho_3} w,\ w_2 = w$, we get :

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

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

$$\frac{\partial^2 \Psi}{\partial u \partial v}(u, v) \quad =$$

$$\frac{E[Y_tY_{t-1}exp(uY_t+vY_{t-1})]}{E[exp(uY_t+vY_{t-1})]} - \frac{E[Y_{t-1}exp(uY_t+vY_{t-1})]}{E[exp(uY_t+vY_{t-1})]}\frac{E[Y_texp(uY_t+vY_{t-1})]}{E[exp(uY_t+vY_{t-1})]}$$

$$= Cov_{u,v}(Y_t, Y_{t-1}), \quad (5.6)$$

where $Cov_{u,v}$ is the covariance computed with respect to density 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 subsection, this type of derivative is consistently estimated by its empirical counterpart, by replacing the expectation by a sample average.

6 Implications for Structural Economic Modelling

Let us now discuss the economic application of the identification results in Propositions 2-3. We consider examples of structural economic models, also called causal economic models [see e.g. Pearl (2009), (2018)] that, by definition, require the assumption of shocks independence. This structural assumption allows us for (semi-) parametric identification. We explore these economic applications from the perspective of a recent debate on causal 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, that is equivalent to choosing between the Ordinary Least Squares (OLS) and the Principal Component Analysis (PCA). The basic specification [see e.g. Fuller (1987)] concerns two latent variables X_{1t}, X_{2t} measured with errors:

$$Y_{1t} = X_{1t} + \eta_{1t}, \quad (6.1)$$

$$Y_{2t} = X_{2t} + \eta_{2t}. \quad (6.2)$$

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

$$X_{2t} = aX_{1t}, \text{ say.} \quad (6.3)$$

Haavelmo (1944), p3, called X the theoretical variables to be distinguished from the observable variables Y, and defined relation (6.3) as the "hypothetical" model.

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

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

If X_{1t} is a sequence of i.i.d. non-Gaussian variables, 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]. 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, for deconvolution we do not need 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 [Delargle, Hall (2016)].

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

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

where $v_t = \eta_{2t} - a\eta_{1t}$. The lagged variable $Y_{1,t-1}$ can be used as an instrument for 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 the observables, it is not necessary to introduce external instruments for identification purpose [see Stock, Watson (2018) for a discussion] [6].

6.2 Scale of memory and predictability puzzle

The representation :

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

where the $X_{jt} = \rho_j X_{j,t-1} + \varepsilon_{j,t}, j = 1, \ldots, K$ are independent with $E\varepsilon_{j,t} = 0, V\varepsilon_{j,t} = 1-\rho_j^2$ [i.e. $VX_{j,t} = 1$] provides a decomposition of each series $(Y_{l,t})$ into components with different persistence. The larger the autoregressive coefficient $|\rho_j|$, the larger the memory scale is. When it exists, the decomposition (6.5) differs from a representation of process Y in the frequency domain (i.e. the Fourier representation), or from wavelets [see Bandi et al. (2018)]. These two latter representations always exist, but their components are uncorrelated, and dependent in general. These components are functions of a small number of strong noises,typically a single strong noise in the one-dimensional system ($L = K = 1$). As it is shown later in this section, the frequency and wavelet representations are not suitable for structural shocks and impulse response functions analysis.

Let us now examine how representation (6.5) can help rationalize some stylized facts highlighted in the predictability literature. This literature considers simple linear predictability from regressions of the averages of consecutive leads on the averages of consecutive lags of the observed variable. More precisely, let us consider a one-dimensional process (Z_t) and define:

$$Z_{t+1,t+h} = Z_{t+1} + \ldots + Z_{t+h}, Z_{t-k+1,t} = Z_t + Z_{t-1} + \ldots + Z_{t-k+1},$$

where $(t-k+1, t)$ is the formation period and $(t+1, t+k)$ the test period (see also the momentum literature [7]).

For given lead h and lag k, the theoretical simple linear regression is:

$$Y_{1,t+1,t+h} = \beta_{hk} Y_{2,t-k+1,t} + w_{kht}. \tag{6.6}$$

[6] 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.

[7] See e.g. De Bondt, Thaler (1987), Chan et al. (1996) for early papers in the momentum literature

Commonly, the dependent and explanatory variables are (consumption growth, dividend growth) [Bansal, Yaron (2004), Bonomo et al. (2015)], (asset return, dividend growth) [Campbell, Shiller (1988)], or (asset return, risk measure) [Bonomo et al. (2015)]. Such regressions can also be performed with two regressors, such as the dividend and the term spread [Fama, French (1989), Table 1], or a volatility measure and the dividend yields [Bandi, Perron (2008), Tables 11, 12], considered as potential predictors of future returns.

Proposition 6 : The theoretical regression coefficient is :

$$\beta_{hk} = \frac{\sum_{j=1}^{K} a_{1j} a_{2j} \gamma(h,k,\rho_j)}{\sum_{j=1}^{K} a_{2j}^2 \gamma(h,\rho_j)},$$

and the theoretical correlation between the dependent and explanatory variables is:

$$R_{hk} = \frac{\sum_{j=1}^{K} a_{1j} a_{2j} \gamma(h,k,\rho_j)}{\left[\sum_{j=1}^{K} a_{1j}^2 \gamma(h,\rho_j)\right]^{1/2} \left[\sum_{j=1}^{K} a_{2j}^2 \gamma(h,\rho_j)\right]^{1/2}},$$

where :

$$\gamma(h,k,\rho) = \frac{\rho(1-\rho^k)(1-\rho^h)}{(1-\rho)^2},$$

$$\gamma(h,\rho) = \frac{1+\rho}{1-\rho} h - \frac{2\rho}{(1-\rho)^2}(1-\rho^h).$$

Proof : These expressions are derived in Appendix 5 i), ii).

QED

We are interested in the pattern of β_{hk}, R_{kk}, when h, k vary. The corollary below follows directly from Proposition 6.

Corollary 1 : $\lim_{h,k\to\infty} \beta_{hk} = \lim_{h,k\to\infty} R_{hk} = 0$.

Proof : This is due to the term linear in h in $\gamma(h,\rho)$. In particular, the convergence to 0 is at a hyperbolic speed (not geometric), and therefore rather slow.

QED

As an illustration of the flexibility of a model combining independent AR(1) processes, let us first consider the simple case, where :

$$Y_t = X_{1t} + \sigma\varepsilon_{2t}, \tag{6.7}$$

where $X_{1t} = \rho X_{1t-1} + \varepsilon_{1t}, E\varepsilon_{1t} = 0, V\varepsilon_{1t} = 1-\rho^2$, and choose $Y_{1t} = Y_{2t} = Y_t, h = k$. Thus we perform autoregressions with the same duration for the formation and test periods for the series $Y_{1,t,t+h}$. Then we have :

$$\beta_{hh}(\rho,\sigma^2) = R_{hh}(\rho,\sigma^2) = \frac{\gamma(h,h,\rho)}{\gamma(h,\rho)+h\sigma^2}. \tag{6.8}$$

In particular :

$$R_{1,1}(\rho,\sigma^2) = \frac{1}{\rho+\sigma^2}, R_{2,2}(\rho,\sigma^2) = \frac{\rho(1+\rho)^2}{2(1+\rho)+2\sigma^2}.$$

Corollary 2 : Let us assume $\rho > 0$, then $R_{2,2}(\rho,\sigma^2) > R_{1,1}(\rho,\sigma^2)$, iff $(1+\rho)^2 \geq 2$ and $\sigma^2 \geq \dfrac{(1-\rho)^2}{(1+\rho)^2-2}$.

Proof : The condition on the correlations is :

$$\begin{aligned} & \frac{\rho(1+\rho)^2}{2(1+\rho)+2\sigma^2} > \frac{\rho}{1+\sigma^2} \\ \iff & (1+\rho)^2(1+\sigma^52) > 2(1+\rho)+2\sigma^2 \\ \iff & [(1+\rho)^2-2]\sigma^2 > (1+\rho)(1-\rho) > 0. \end{aligned}$$

The result follows.

QED

In brief, the sequence of correlations is first increasing in h, if ρ is sufficiently large $\rho > \sqrt{2}-1$, and σ^2 is also large. When ρ is close to 1, the constraint on σ^2 becomes : $\sigma^2 > 0$ and is always satisfied. Thus it seems important to analyze the behaviour of $R_{h,h}(\rho,\sigma^2)$ close to unit root.

Corollary 3 : Let us assume $\rho = 1-\delta/h$, where $\delta > 0$ is fixed. Then

$\lim_{h \to \infty} R_{h,h}(1 - \delta/h; \sigma^2) = \dfrac{[1 - \exp(-\delta)]^2}{2[\exp(-\delta) - 1 + \delta]} \equiv R_\infty(\delta)$.

The function $R_\infty(\delta)$ is a decreasing function of δ, from 1 for $\delta = 0$, to 0 for $\delta = \infty$.
Proof : See Appendix 5 iii).

Corollary 2 means that we can expect a hump-shaped pattern of the theoretical correlation. Corollary 3 suggests that the hump may arise for a rather large h and its size is determined by the parameter δ. The model in Corollary 3 resembles the dynamic model of consumption growth in Bansal, Yaron (2004), where X_{1t} is "a small persistent predictable component which determines the conditional expectation of consumption growth". The hump-shaped patterns are illustrated in Figure 1 below for parameters set equal to $\rho = 0.8, \sigma^2 = 10; \rho = 0.9, \sigma^2 = 20; \rho : 0.99, \sigma^2 = 100, \rho = 0.995, \sigma^2 = 10$.

[Insert Figure 1 : $\beta_{h,h} = R_{h,h}$ function of h]

Such patterns are evidenced in the empirical literature [see e.g. Bandi et al. (2018) for a similar curve, and Pandi, Perron (2008), Fig 2 for a curve for small h]. Predictability is not apparent in the short run (h small), as the small persistent component is hidden by the large noise. When h increases the effect of the noise diminishes. Hence, for very large h, function $R_{h,h}$ decreases slowly to zero due to the persistent component.

In general, the theoretical correlation R_{hk} depends on both k and h. This is illustrated in Figure 2 for model (6.7) with parameters set equal to the same values as in Figure 1. The contour plots in the top panels have been zoomed in for better visualisation.

[Insert Figure 2 : R_{hk} function of h and k]

Therefore, if we predict $Y_{t+1,t+h}$ from a linear regression on $Y_{t-k+1,t}$, there exists a value $k^* = h$ (in Figure 2) for which the associated theoretical R_{hk} attains its maximum. Instead of computing the optimal prediction of $Y_{t+h,t+1}$ given the past $Y_t, Y_{t-1}, \ldots$, we can improve the prediction method based on regression (6.6) along the following steps :

step 1 : Estimate ρ, σ^2 as in Section 2.

step 2 : Compute $\beta_{h,h}(\hat{\rho}_T, \hat{\sigma}^2_T) = R_{h,h}(\hat{\rho}_T, \hat{\sigma}^2_T)$, where $\hat{\rho}_T$, $\hat{\sigma}^2_T$ are the estimators from step 1.

step 3 : Predict $Y_{t+1,t+h}$ by $\beta_{hh}(\hat{\rho}_T, \hat{\sigma}^2_T) Y_{t-h+1,t} = R_{hh}(\hat{\rho}_T, \hat{\sigma}^2_T) Y_{t-h+1,t}, \, , h = 1 \ldots$.

The proposed approach above uses the data on Y_t to estimate ρ, σ^2 efficiently, without summing up the consecutive lags. This estimation approach differs from a much simpler approach that consists in regressing empirically $Y_{t+1,t+h}$ on $Y_{t-h+1,t}$. for $t = h, \ldots, T-h$, and provides estimators $\tilde{\beta}_{hh}, \tilde{R}_{hh}$, say. This latter approach has to be avoided if h is too large with respect to the total number of observations. Indeed, in such a case $\tilde{\beta}_{hh}$ and $\tilde{R}_{hh}$ can have different stochastic limits, and these limits can be very different from their theoretical counterparts [see Bandi, Perron (2008)] [8].

To illustrate the size of this bias, we consider a series of simulated data with $T = 400$ observations and parameters $\rho = 0.99, \sigma^2 = 10$. The simulated series is displayed in Figure 3.

[Insert Figure 3 : Simulated Series]

Despite the high value of the autoregressive coefficient, we do not observe a trend and the trajectory is rather erratic. The estimates of ρ, σ^2 in step 1 are : $\hat{\rho}_T = 0.904, \hat{\sigma}_T^2 = 10.23$. In Figure 4 we plot as a function of h the true values of R_{hh} (dashed line), the values of $R_{hh}(\hat{\rho}_T, \hat{\sigma}_T^2)$ (dotted line), as well as the values of $\tilde{R}_{h,h}$ (solid line). As expected, the $\tilde{R}_{h,h}$ line lies significantly above the true value of R_{hh} and is close to 1 for large lags h because of the overlapping and the decreasing number of observations from which it is calculated when lag h increases. The estimator $R_{hh}(\hat{\rho}_T, \hat{\sigma}_T^2)$ provides the hump shape, but underestimates the true value of the determination coefficient. This estimation bias is due to the near unit root persistence and has not been adjusted for in the example [9].

[Insert Figure 4 : True and Estimated Coefficients]

6.3 Noisy News and Impulse Response Functions

(a) The model

There is a renewed interest in the idea that business cycle could be driven by changes in the expectations about the future and that it is important to separate the actual changes in fundamentals (news) from those due to temporary measurement errors (noise) [see e.g. Barski, Sims (2012), Schmitt-Grohe, Uribe (2012), Blanchard et al. (2013), Forni et al. (2017)]. As a simple example, let us consider a model that illustrates that "the productivity is driven by two shocks : a permanent shock and a transitory shock" [Blanchard et al

[8] This is easily understood if for instance $T = 1000$ large and $h = k = 500$. The empirical regression is performed on a single point $\tilde{\beta}_{500,500} = \frac{Y_{501,1000}}{Y_{1,500}}$, and $\tilde{R}_{500,500} = 1$.

[9] This bias adjustment is out of the scope of this paper.

(2013)]. Formally it can be written as :

$$Y_t = X_t + \eta_t, \tag{6.9}$$

where Y_t denotes the productivity, X_t its permanent fundamental component and η_t is the transitory component. We assume that the fundamental component X_t has an AR(1) dynamics:

$$X_t = \rho X_{t-1} + \epsilon_t, \tag{6.10}$$

where ρ is close to 1, but less than 1.

The business cycle literature introduces various assumptions on the agent's information about the fundamental component X_t. It can be assumed that the agents have perfect information, that is, that the structural shocks ε_t are observed [Barski, Sims (2012)], or that agents observe noisy signals on the value of X_t [Forni et al (2017)], or that they just observe Y_t, which is another noisy observation of X_t. Below, for ease of exposition we consider that last case of imperfect information .

In Section 2, we showed that the observed process (Y_t) admits a weak ARMA(1,1) representation. More precisely, we have:

$$Y_t = \frac{\epsilon_t}{1-\rho B} + \eta_t = \frac{\epsilon_t + \eta_t - \rho\eta_{t-1}}{1-\rho B}.$$

The process $\epsilon_t + \eta_t - \rho\eta_{t-1}$ can be written as a weak MA(1) process $u_t - \theta u_{t-1}$. By comparing their second-order moments, we get:

$$\begin{cases} \sigma_\epsilon^2 + (1+\rho^2)\sigma_\eta^2 & = & (1+\theta^2)\sigma_u^2, \\ \rho\sigma_\eta^2 & = & \theta\sigma_u^2, \end{cases}$$

and θ is the solution with modulus less than 1 of the equation:

$$\frac{\theta}{1+\theta^2} = \frac{\rho\sigma_\eta^2}{\sigma_\epsilon^2 + (1+\rho^2)\sigma_\eta^2}. \tag{6.11}$$

(b) The expectations and innovations

Under imperfect information, it is impossible to exactly recover the two latent components X_t, η_t, or equivalently the news and noise : ε_t, η_t. The agent can only compute the expectations conditional on the available information $\underline{Y_{t-1}} = (Y_{t-1}, Y_{t-2}, \ldots)$. Since Y is observed, the conditional expectation $E(Y_t|\underline{Y_{t-1}})$ can be consistently estimated. In general it is not relevant from the economic point of view, while the prediction of X_t, defined by $E(X_t|\underline{Y_{t-1}})$ is. Proposition 2 shows that it can also be estimated nonparametrically. However prediction $E(X_t|\underline{Y_{t-1}})$ differs from $E[X_t|\underline{X_{t-1}}] = E[X_t|\underline{X_{t-1}}, \underline{\eta_{t-1}}]$, as the prediction error $\tilde{\varepsilon}_t = X_t - E(X_t|\underline{Y_{t-1}})$ differs from ε_t. Due to the lack of information, the knowledge of "reduced form" innovations $(\tilde{\varepsilon}_t)$ does not allow us to recover the structural innovations (ε_t). Thus, we cannot compute the values of structural innovations from the values of reduced form innovations. However the knowledge of the distribution of the reduced form innovation process allows us to recover the distributions of the structural innovations processes (see Corollary 1 in Section 5). This identification of shocks distributions is sufficient to evaluate the sensitivity of nonlinear causality measures to errors-in-variables [Anderson et al. (2018)], as well as the structural impulse response functions, discussed in this Section.

(c) Impulse response function

Even though the values of structural shocks ε_t, η_t cannot be recovered, the impulse response functions can be computed, as they depend on the distributions only.

From an economic point of view, the impulse response function (IRF) of economic interest concerns the fundamental variable X_t. Thus, we have to introduce a shock to the structural innovation ϵ_t, of size $\delta = q_\epsilon(95\%) - q_\epsilon(50\%)$, where $q_\epsilon(\alpha)$ is the α-quantile of the distribution of ϵ. Then, the IRF is:

$$\text{IRF}(h) = \rho^h[q_\epsilon(95\%) - q_\epsilon(50\%)]. \tag{6.12}$$

As the structural IRF involves the autoregressive parameter ρ and the distribution of the structural noise, it can be consistently estimated.

In practice the presence of additional noise η_t is often disregarded and a different IRF is computed from the weak ARMA(1,1) representation:

$$Y_t = \frac{(1-\theta B)}{1-\rho B} u_t.$$

This reduced form IRF is computed as:

$$\widetilde{\text{IRF}}(h) = |\rho^h - \theta\rho^{h-1}|[q_u(95\%) - q_u(50\%)]. \tag{6.13}$$

where the distribution of the reduced form innovation depends on the distributions of the structural innovations and parameters ρ, σ^2.

If $\sigma_\eta^2 = 0$, there is no noise η, $\theta = 0$, and the structural and reduced form IRF's are equal $\widetilde{IRF}(h) = IRF(h)$, $\forall h$. Otherwise, the two IRF's can significantly differ in the presence of noise η. In the extreme case of a large noise variance $\sigma_\eta^2 \to \infty$, we get $\theta = \rho$, and the reduced form IRF vanishes $\widetilde{IRF}(h) = 0$, $\forall h$.

The above discussion is related to the recent literature on the definition of shocks, and their impact and control. We can either fix the values of one specific innovation and consider that the other shocks are stochastic with their marginal distributions set equal to the marginal sample distribution (fixing), or condition the computation of IRF with respect to that specific innovation (conditioning) [Heckman, Pinto (2015), Pearl (2018)]. These methods provide different measures of the shocks impacts, i.e. different IRF's, except when the (structural) innovations are independent. In our framework, ε_t, η_t are independent, and the impact of shocks on ε_t, obtained from fixing and conditioning methods are identical that eliminates any ambiguity. Since the reduced form innovations $u_t's$ are uncorrelated, but dependent [10] the relevance of the reduced form $\widetilde{IRF}$ can be questioned.

6.4 Two-step 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

[10] except in the Gaussian case.

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 commonly used in linear factor models analysis in Finance [Sharpe (1964), Lintner (1965)] where these models underly the arbitrage pricing theory [Ross (1976)], and also in other fields [see e.g. Lin, Li, Liu (2007)].

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 are smooth approximations that create unreliable estimated patterns of the dynamics of X_{1t}, X_{2t} and their distributions.

Proposition 3 shows that it is possible to avoid these drawbacks by using the joint historical distributions 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)].

i) **The algorithm**

Let us consider the general form :

$$Y_t = AX_t + \eta_t, \tag{6.14}$$

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 $\underline{Y_t}$:

$$G_{t|t}(\mu) = E[\exp(i\mu' X_t)|\underline{Y_t}], \tag{6.15}$$

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

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

where $\psi_\eta(\lambda), \psi_u(\mu)$ are the characteristic functions of η_t and u_t, respectively. It follows that the joint characteristic function of (Y_{t+1}, X_{t+1}) given $\underline{Y_t}$ is :

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

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 :

$$G_{t+1|t+1}(\mu) = \frac{\int \psi_\eta(\lambda)\psi_u(A'\lambda+\mu)G_{t|t}[(diag\ \rho)(A'\lambda+\mu)]\exp(-i\lambda' y_{t+1})d\lambda}{\int \psi_\eta(\lambda)\psi_u(A'\lambda)G_{t|t}[(diag\ \rho)A'\lambda]\exp(-i\lambda' y_{t+1})d\lambda}. \tag{6.16}$$

This updating formula is easily implemented as long as the source dimension is small, such as $K \leq 4$.

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 error, while $Y_t = Y_t^* + \eta_t$ is the output measured with 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:

$$l^*(y_t^*|y_{t-1}^*) = \frac{1}{|detA|} g[A^{-1}y_t^*|A^{-1}y_{t-1}^*], \tag{6.17}$$

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. (2018).

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 independent AR(1) sources and left unspecified the distributions of the noises. We have shown how to identify and estimate nonparametrically the mixing matrix, the autoregressive coefficients, as well as the distributions of all underlying noises.

The importance of this identification result has been illustrated by considering various outstanding issues in Economics and Finance such as the identification in the errors-in-variables models (including Gaussian models), modelling of the scale of memory and the predictability puzzle as well as the computation of structural impulse response functions 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 established in this paper is an advancement in the econometrics of causal models, structural shocks and IRF's considered and debated in the recent economic literature [Heckman, Pinto (2015), Pearl (2018), Stock, Watson (2018)].

References

[1] Albataireh, Z. and F., Salem (2014): " A Robust ICA Based Algorithm for Blind Separation of Convolutive Mixtures", ArXiv.

[2] Anderson, B., Deistler, M., and J.M., Dufour (2018): "On the Sensitivity of Granger Causality to Errors-in-Variables, Linear Transformations and Subsampling", Journal of Time Series Analysis, forthcoming.

[3] Ansley, C. (1977): "On the Structure of Moving Average Processes", Journal of Econometrics, 6, 121-134.

[4] Babaie-Zadeh, M. (2002): "Darmois-Skitovich Theorem and its Proof", Sharif University of Technology, DP.

[5] Bandi, F., and B., Perron (2008) : "Long Run Risk-Return Trade Offs", Journal of Econometrics, 143, 349-374.

[6] Bandi,F. , Perron B., Tamoni, A., and C., Tebaldi (2018): "The Scale of Predictability", LSE, DP

[7] Bansal, R., and A., Yaron (2004) : "Risks for the Long Run : A Potential Resolution of Asset Pricing Puzzles", Journal of Finance, 59, 1481-1509.

[8] Barski, R., and E., Sims (2012) : "Information, Animal Spirits and the Meaning of Innovations in Consumer Confidence", American Economic Review, 102, 1343-1377.

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

[10] Bates, D. (2006) : "Maximum Likelihood Estimation of Latent Affine Processes", Review of Financial Studies, 19, 909-965.

[11] Bekker, P. (1986): "Comment on Identification in the Linear Errors-in-Variables Model", Econometrica, 54, 215-217.

[12] Ben-Moshe, D. (2018a): "Identification of Joint Distributions in Dependent Factor Models", Econometric Theory, 34, 134-165.

[13] Ben-Moshe, D. (2018b): "Linear Error-in-Variables and Dependent Factor Models", D.P. Hebrew University of Jerusalem.

[14] Berkson, J. (1950) : "Are There Two Regressions ?", JASA, 45, 164-180.

[15] Blanchard, O., l'Huillier, J-P., and G., Lorenzoni (2013) : "News, Noise and Fluctuations : An Empirical Exploration", American Economic Review, 103, 3045-3070.

[16] Bonhomme, S., and J.M., Robin (2010): "Generalized Nonparametric Deconvolution with an Application to Earning Dynamics", Review of Economic Studies, 77, 491-533.

[17] Bonomo, M., Garcia, R., Meddahi, N., and R., Tedongap (2015) : "The Long and the Short of the Risk Return Trade-Off", Journal of Econometrics, 187, 580-592.

[18] Bradley, W., and W., Cook (2012): "Two Proofs of the Existence and Uniqueness of Partial Fraction Decomposition", International Mathematical Forum, 7, 1517-1535.

[19] Bryc, W. (1995): "Normal Distributions: Characterizations with Applications", Lecture Note in Statistics, 100, Springer.

[20] Campbell, J., and R., Shiller (1988) : "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors", Review of Financial Studies, 1, 195-227.

[21] Cardoso, J., and B., Pham (2004): "Optimization Issues in Noisy Gaussian ICA", in Proceedings ICA, Grenada, Spain.

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

[23] Chan, L., Jegadesh, N., and J., Lakonishok (1996) : "Momentum Strategies", Journal of Finance, 51, 1681-1713.

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

[25] 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.

[26] Comon, P., and C., Jutten (2010): "Handbook of Blind Source Separation of Independent Component Analysis", New York Academic Press.

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

[28] Dagenais, M., and D., Dagenais (1997): "Higher Moment Estimators for Linear Regression Models with Errors in Variables", Journal of Econometrics, 76, 193-225.

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

[30] De Bondt, W., and R., Thaler (1987) : "Further Evidence on Investor Overreaction and Stock Market Seasonality", Journal of Finance, 42, 557-581.

[31] Delaigle, A., and P., Hall (2016): "Methodology for Non-Parametric Deconvolution when the Error Distribution is Unknown", JRSS B, 78, 231-252.

[32] De Lathauwer, L. (2008): "Blind Identification of Underdetermined Mixtures by Simultaneous Matrix Diagonalization", IEEE Transactions on Signal Processing, 56, 1096-1105.

[33] 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.

[34] 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.

[35] Eriksson, J., and V., Koivunen (2004): "Identifiability, Separability and Uniqueness of Linear ICA Models", IEEE Signal Process Lett. , 11, 601-604.

[36] Fama, E., and K., French (1988) : "Dividend Yields and Expected Stock Returns", Journal of Financial Economics, 22, 3-25.

[37] Fama, E., and K., French (1989) : "Business Conditions and Expected Returns on Stocks and Bonds", Journal of Financial Economics, 25, 23-49.

[38] Forni, M., Gambetti, L., Lippi, M., and L., Sala (2017) : "Noisy News in Business Cycle", American Economic Journal : Macroeconomics", 9, 122-152.

[39] Frisch, R. (1934): "Statistical Confluence Analysis be Means of Complete Regression Systems", Publication 5, University of Oslo.

[40] Fuller, W. (1987): "Measurement Error Models", New York, Wiley.

[41] Geary, R. (1942) : "Inherent Relations Between Random Variables", Proc. R. Irish Acad, 47,63-76.

[42] Haavelmo, T. (1944) : "The Probability Approach in Econometrics", Econometrica, 12, 1-115.

[43] Heckman, J., and R., Pinto (2015) : "Causal Analysis After Haavelmo", Econometric Theory, 31, 115-151.

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

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

[46] Kapteyn, A., and T., Wansbeek (1983): "Identification in the Linear Errors-in-Variables Model", Econometrica, 51, 1847-1849.

[47] Klepper, S., and E., Leamer (1984): "Consistent Sets of Estimators for Regressions with Errors in All Variables", Econometrica, 52, 163-183.

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

[49] Leonard, D. (2011): "Estimating a Bivariate Linear Relationship", Bayesian Analysis, 6, 727-754.

[50] 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.

[51] Linnik, Y. (1964): "Decomposition of Probability Distributions", Oliver and Boyd.

[52] 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.

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

[54] Mansour, A., Barros, K., and N., Ohnishi (1998): "Subspace Adaptive Algorithm for Blind Separation of Convolutive Mixtures by Conjugate Gradient Method", First International Conference and Exhibition Digital Signal Processing

[55] Marcinkiewicz, J. (1938): "Sur une propriete de la loi de Gauss", Math. Zeitschrift, 44, 612-618.

[56] 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.

[57] Pearl, J. (2009) : "Causality : Models, Reasoning and Inference", 2nd ed., Cambridge Univ.Press.

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

[59] 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.

[60] Poirier, D. (1998): "Revisiting Beliefs in Nonidentified Models", Econometric Theory, 14, 483-509.

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

[62] Ross, S. (1976): "The Arbitrage Theory of Capital Asset Pricing", Journal of Economic Theory, 17, 254-286.

[63] Samuelson, P. (1942): "A Note on Alternative Regressions", Econometrica, 10, 80-83.

[64] Schennach, S. (2013): "Convolution without Independence", DP. UCL

[65] Schennach, S. (2016) "Recent Advances in the Measurement Error Literature", Annual Review of Economics, 8, 341-377.

[66] Schennach, S., and A., Lewbel (2007): "Nonparametric Identification of the Classical Errors-in-Variables Model without Side Information", DP Brown University.

[67] Schmitt-Grohe, S., and M., Uribe (2012) : "What's News in Business Cycles", Econometrica, 80, 2733-2764.

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

[69] Shi, X. (2011): "Blind Signal Processing", Springer.

[70] Stefanski, L. (2000): "Measurement Error Models", JASA, 95, 1353-1358.

[71] Stock, J., and M., Watson (2018) : "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments", Economic Journal, 128, 917-948.

[72] Szekely, G., and C., Rao (2000): "Identifiability of Distributions of Independent Random Variables by Linear Combinations and Moments", Sankhya, A, 193-202.

[73] Thi, N., and C., Jutten (1995): "Blind Source Separation for Convolutive Mixtures", Signal Processing, 45, 209-229.

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

[75] Williams, B. (2018): "Identification of the Linear Factor Model", DP George Washington University.

[76] Wooley, E. (1941): "The Method of Minimized Errors as a Basis of Correlation Analysis", Econometrica, 9, 38-62.

[77] Yilmaz, A., and S., Rickard (2004): "Blind Separation of Speech Mixtures via Time Frequency Masking", IEEE Transactions on Signal Processing, 52, 1830-1847.

[78] 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.

We note :

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

We get :

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

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

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

Lemma 1:

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

Proof:

We have

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

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:

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

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)[11]

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 and functions $f_i,\ i = 1, ..., N$ are continuous. 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. The proof for the second characteristic function is similar with its real and imaginary components considered separately.

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

Due to the existence of first and second-order moments, the joint c.g.f. is differentiable. 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:

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

Then we can apply the change of arguments of appendix A.2 and write:

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

$c_1'(w_1) - \gamma_1'(w_1) + c_2'(w_2) - \gamma_2'(w_2) + \sum_{k=3}^{K}\{c_k'(\frac{1}{\rho_2-\rho_1}[(\rho_2-\rho_k)w_1 + (\rho_k-\rho_1)w_2]$
$-\gamma_k'(\frac{1}{\rho_2-\rho_1}[(\rho_2-\rho_k)w_1 + (\rho_k-\rho_1)w_2]\} = 0,$
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 $\gamma_k''(0) = c_k''(0) = \sigma_k^2/(1-\rho_k^2)$.

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

When $K \geq 4$, the function differs 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 $\gamma_k(0) - c_k(0) = 0, \ \forall k$.

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 based on the joint c.g.f. (a.2), we infer that the unknown $\delta_k, k = 1, ..., 4$ 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 have to also consider nonlinear serial dependence at higher lags.

i) If the $|\rho_k|$ are all different, that is, if there do 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

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

where $X_{kt} = \rho_k^2 X_{k,t-2} + \tilde{\epsilon}_{kt}$, with $\tilde{\epsilon}_{kt} = \epsilon_{kt} + \rho_k \epsilon_{k,t-1}$, 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 use 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 the ρ_k^3 are all different 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 $\gamma_k(u) - c_k(u) = \sum_{j=3}^{K-1} \delta_{kj} u^j$. 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 joint distribution of (Y_t, Y_{t-2}) or 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 $1+2+\ldots+(K-1) = 0.5K(K-1)$. Moreover if $\rho_k = \rho_l$, two columns of matrix C are equal, and then $det\ C = 0$. It follows that the polynomial C is divisible by $\rho_k - \rho_l$ for any pair $k \neq l$. Since there are $0.5K(K-1)$ 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 :

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

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

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

Therefore, by applying the Darmois' Lemma, we find that all second-order derivatives $c_j\text{"}(u),\ j = 1,2$, and $c\text{"}(u)$ 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:

$$\begin{aligned}\frac{\theta}{1+\theta^2} &= \frac{\sigma_1^2\theta_1 + \sigma_2^2\theta_2}{\sigma_1^2(1+\theta_1^2) + \sigma_2^2(1+\theta_2^2)} \\ &\iff \frac{\theta}{1+\theta^2} = \frac{\sigma_1^2(1+\theta_1^2)}{\sigma_1^2(1+\theta_1^2) + \sigma_2^2(1+\theta_2^2)}\frac{\theta_1}{1+\theta_1^2} + \frac{\sigma_2^2(1+\theta_2^2)}{\sigma_1^2(1+\theta_1^2) + \sigma_2^2(1+\theta_2^2)}\frac{\theta_2}{1+\theta_2^2},\end{aligned}$$

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, 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 possible simplifications when determining the asymptotic variance, let us consider the real c.g.f. with $K = 2$. In that case, the estimators are 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 four-wise distribution 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:

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

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

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

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 $\dfrac{\partial\psi(u,v)}{\partial v}$ 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 $M = [Ecos(uY_t + vY_{t-1})]]^2 + [Esin(uY_t + vY_{t-1})]^2$. We deduce 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

i) **Expression of** $\gamma(h,k;\rho)$

The covariance matrix between $(X_t, X_{t-1}, \ldots, X_{t-k+1})'$ and $(X_{t+1}, X_{t+2}, \ldots X_{t+h})$ is equal to :

$$
\Gamma(k,h) = \begin{pmatrix} \rho & \rho^2 & \ldots & \rho^h \\ \rho^2 & \rho^3 & & \rho^{h+1} \\ \vdots & & & \vdots \\ \rho^k & \rho^{k+1} & & \rho^{k+h} \end{pmatrix}.
$$

We have : $\gamma(h,k;\rho) = e_k'\Gamma(k,h)e_h$, where e_h is the h-dimensional vector with unitary components. Therefore :

$$\begin{aligned}\gamma(h,k;\rho) &= [\rho+\rho^2+\ldots+\rho^k]+\ldots+[\rho^h+\rho^{h+1}+\ldots+\rho^{h+k}]\\ &= [1+\rho+\ldots+\rho^{k-1}][\rho+\ldots+\rho^h]\\ &= \rho\frac{(1-\rho^k)(1-\rho^h)}{(1-\rho)^2}.\end{aligned}$$

ii) **Expression of** $\gamma(h;\rho)$

The variance-covariance matrix of $(X_{t+1},\ldots,X_{t+h})'$ is :

$$\Gamma(h)=\begin{bmatrix}1 & \rho & \ldots & \rho^{h-1}\\ \rho & \ddots & & \vdots\\ \vdots & & \ddots & \vdots\\ \rho^{h-1} & \ldots & & 1\end{bmatrix}.$$

Therefore :

$$\begin{aligned}\gamma(h;\rho) &= h+2\left\{(\rho+\ldots+\rho^{h-1})+(\rho+\ldots+\rho^{h-2})+\ldots+\rho\right\}\\ &= h+2\rho\left\{\frac{1-\rho^{h-1}}{1-\rho}+\frac{1-\rho^{h-2}}{1-\rho}+\ldots+\frac{1-\rho}{1-\rho}\right\}\\ &= h+\frac{2\rho}{1-\rho}\left[h-1-[\rho+\rho^2+\ldots+\rho^{h-1}]\right]\\ &= \frac{1+\rho}{1-\rho}h-\frac{2\rho}{1-\rho}\left[1+\frac{\rho}{1-\rho}(1-\rho^{h-1})\right]\\ &= \frac{1+\rho}{1-\rho}h-\frac{2\rho}{(1-\rho)^2}[1-\rho^h].\end{aligned}$$

iii) **Close to unit root behaviour**

When $\rho=1-\delta/h,\delta>0$, we get : $1-\rho^h=1-\exp[h\log(1-\delta/h)]\sim 1-\exp(-\delta)$, for h large. It follows from (6.8) that :

$$\begin{aligned} R(h, 1-\delta/h, \sigma^2) &\sim \frac{(1-\exp(-\delta))^2}{\delta^2} h^2 / \left[\frac{2h^2}{\delta} + h\sigma^2 - 2(1-\exp(-\delta)) \frac{h^2}{\delta^2} \right] \\ &\sim \frac{(1-\exp(-\delta))^2}{\delta^2} / \left\{ \frac{2}{\delta} - 2\frac{(1-\exp(-\delta))}{\delta^2} \right\} \\ &= \frac{(1-\exp(-\delta))^2}{2[\exp(-\delta) - 1 + \delta]} \equiv R_\infty(\delta), \ \text{for}\, h \to \infty. \end{aligned}$$

Lemma : The function $R_\infty(\delta)$ is a decreasing function of δ, such that:
$R_\infty(0) = 1, R_\infty(\infty) = 0.$

Proof : $\dfrac{dR_\infty(\delta)}{d\delta}$ has the same sign as the function :

$$\begin{aligned} a(\delta) &= 2(1-\exp(-\delta))[\exp(-\delta) - 1 + \delta] - (1-\exp(-\delta))^2[-\exp(-\delta) + 1] \\ &= (1-\exp(-\delta))\{2[\exp(-\delta) - 1 + \delta] - (1-\exp(-\delta))^2] \\ &= [1-\exp(-\delta)]2[\exp(-\delta) - 1 + \delta](1 - R_\infty(\delta)), \end{aligned}$$

and it is easy to see that the second factor on the right hand side is always negative.

QED

Figure 1 : $\beta_{hh} = R_{hh}$ function of h

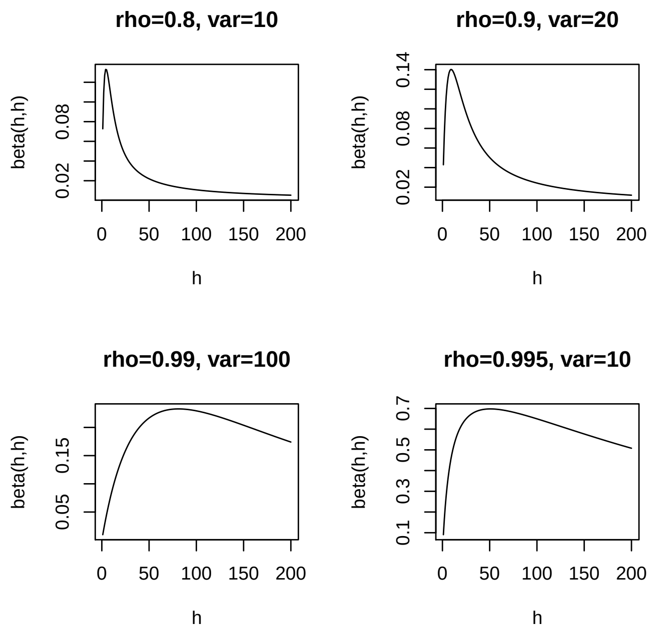

Figure 2 : $R_{h,k}$ function of h, k

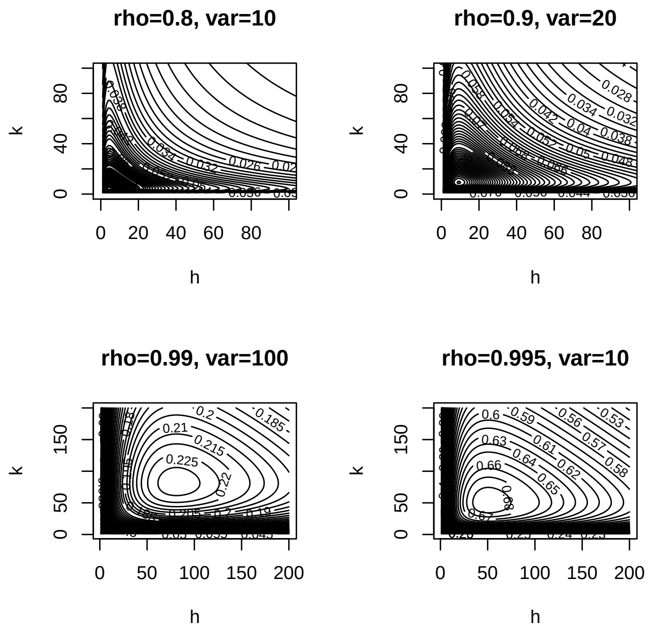

Figure 3 : Simulated Path

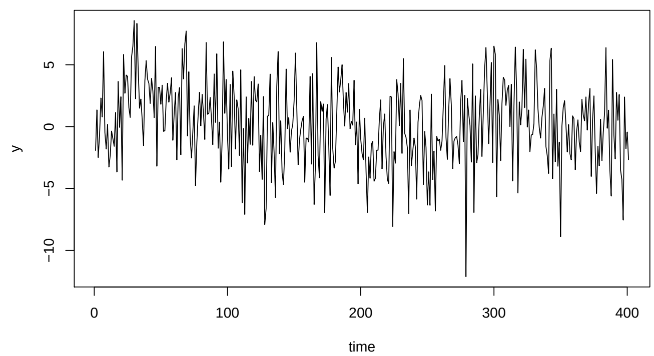

Figure 4 : True and Estimated Coefficients