Identification of Mixed Causal-Noncausal Models : How Fat Should We Go?

Maastricht University

Alain Hecq, Lenard Lieb,
Sean Telg

Identification of Mixed Causal-Noncausal Models: How Fat Should We Go?

RM/15/035

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O Box 616
NL- 6200 MD Maastricht
The Netherlands

Identification of Mixed Causal-Noncausal Models: How Fat Should We Go?

Alain Hecq[*] **Lenard Lieb** **Sean Telg**

Maastricht University

September 29, 2015

Abstract

Gouriéroux and Zakoïan (2013) propose to use noncausal models to parsimoniously capture nonlinear features observed in financial time series and in particular bubble phenomena. In order to distinguish causal autoregressive processes from purely noncausal or mixed causal-noncausal ones, one has to depart from the Gaussianity assumption on the error distribution. This paper investigates by means of simulation how fat the tails of the distribution of the error process have to be such that those models can be identified in practice. We compare the performance of the MLE, assuming a $t-$distribution, with those of the LAD estimator that we propose in this paper. Similar to Davis, Knight and Liu (1992) we find that for infinite variance autoregressive processes both the MLE and LAD estimator converge faster. We further specify the general asymptotic normality results obtained in Andrews, Breidt and Davis (2006) for the case of t-distributed and Laplacian distributed error terms. We first illustrate our analysis by estimating mixed causal-noncausal autoregressions to model the demand for solar panels in Belgium over the last decade. Then we look at the presence of potential noncausal components in daily realized volatility series for 21 equity indexes. The presence of a noncausal component is confirmed in both empirical illustrations.

Keywords: noncausal models, non-Gaussian distributions, realized volatilities, bubbles.

[*]Corresponding author: Alain Hecq, Maastricht University, School of Business and Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: a.hecq@maastrichtuniversity.nl; URL: http://researchers-sbe.unimaas.nl/alainhecq/

1 Introduction

Several time series textbooks (e.g., Brockwell and Davis, 1991, 2002) advocate the use of noncausal (or similarly, forward-looking) models. The reason for this is that these models offer the possibility to rewrite a process with explosive roots into a process in reverse time with roots outside the unit circle. Also in the applied econometric literature there is a growing interest in noncausal models. We see at least three reasons explaining this interest. First, several authors (see e.g., Lanne, Luoto and Saikkonen, 2012; Lanne, Nyberg and Saarinen, 2012) have shown that noncausal models, which have the feature that they encompass information about the future path of stationary time series, might improve forecast performances obtained with purely causal autoregressive models. Most papers look at inflation in their evaluations, but also commodity prices (Lof and Nyberg, 2015) as well as other macroeconomic variables (Lanne et al., 2012). Second, noncausal models are together with non-invertible models, special cases where shocks are non-fundamental (see Lippi and Reichlin, 1993a, 1993b; Alessi, Barigozzi and Capasso, 2011; Beaudry, Fève, Guay and Portier, 2015). This issue is crucial in macroeconometrics (Hansen and Sargent, 1991) and is interpreted as the evidence that the econometrician has a smaller information set at his disposal than the economic agents have. Third, and this will be the angle taken in our paper, forward-looking autoregressive models with non-Gaussian fat tails disturbances are able to replicate features that previously could only be obtained by highly nonlinear and complex models. Gouriéroux and Zakoïan (2013) have shown that a simple noncausal AR(1), i.e., $y_t = \rho y_{t+1} + \varepsilon_t$, with an *iid* standard Cauchy distributed error term, is able to mimic processes similar to bubbles observed in economic and financial variables. Without entering into definition details (see e.g., Scherbina, 2013 or Stiglitz, 1990 for a survey), the term bubble roughly refers to as a rapid acceleration of prices (assets, real estate, commodity, etc.), over a natural level (often called the intrinsic value) given the fundamentals of the economy and discounted present values of expected returns. In general, this is driven by mechanisms such as speculations or anticipations. Because it is difficult to observe intrinsic values in real time, bubbles are often conclusively observed and identified in retrospect, when a sudden drop in prices appears. Such a drop is known as a crash or a bubble burst. However, standard causal autoregressive models were found to be incapable of capturing the behavior of series that increase and then suddenly burst. Non-linear processes with a mixture of unit and explosive roots are often considered in the literature to model this bubble feature (see Phillips, Wu and Yu, 2011 or the survey in Homm and Breitung, 2012). We also argue that noncausal models can provide a simple and convenient alternative representation. Moreover, they can be exploited to forecast turning points (crashes say) for different horizons and consequently noncausal models yield helpful risk measurements.

This paper provides several elements that would facilitate the interpretation and the estimation of mixed-causal and noncausal models for practitioners. The few existing applied econometric papers[1] on purely noncausal and mixed causal-noncausal models are often not very specific on how to implement the estimation and the construction of inference (in a very practical, applied sense). Also the existing literature lacks a small sample assessment of the (relative) performance of different approaches. We want to guide the practitioner on the following issues that we emphasize in this paper: (*i*) We show that the fatter the tails of the distribution are, the faster the convergence of the estimator is achieved (and the more accurate the identification of the model is). This means that estimating causal and noncausal models for volatile financial time series will be, in theory, relatively easy; (*ii*) We compare the performance of a Student's t Maximum Likelihood Estimation (MLE) approach (advocated by Breidt, Davis, Lii and Rosenblatt, 1991 or Lanne and Saikkonen, 2011) with an alternative Least Absolute Deviation (LAD) estimator that we propose in this paper. We show by simulation that this latter approach has appealing properties for odd distributions such as asymmetric ones or mixture of distributions

[1] Actually most of the papers (plus references therein) can be found at Markku Lanne's (http://blogs.helsinki.fi/lanne/) and Joann Jasiak's (http://dept.econ.yorku.ca/jasiakj/) webpages.

that are commonly encountered in practice; (*iii*) We also investigate how fast the causal and noncausal parameters (by MLE or LAD estimator) approximate their asymptotic normal distribution; (*iv*) As a first illustration, we provide the details of the necessary steps to implement those tools for the demand of solar panels in Belgium. While that application might be affected by the small sample that is used (about 75 observations), we provide a second application where we look at mixed causal-noncausal models on daily realized volatilities for 21 equity indexes measured over 2000-2014 and hence on more than 3000 observations; (*v*) We have developed a Matlab toolbox that easily estimates and identifies those models.[2]

The rest of this paper is as follows. Section 2 recalls the elements needed to understand the literature on noncausal models. In Section 3 we elaborate on estimation and how to construct inference on mixed causal-noncausal models. The results from the Monte Carlo simulations are collected in Section 4. Section 5 details the two applications. Section 6 concludes. Appendices collect additional material.

2 The Mixed Causal-Noncausal Autoregressive Model

The univariate mixed causal-noncausal autoregressive model, denoted MAR(r, s) (see inter alia Lanne and Saikkonen, 2011; Gouriéroux and Jasiak, 2015) is usually written as

$$(1 - \phi_1 L - ... - \phi_r L^r)(1 - \varphi_1 L^{-1} - ... - \varphi_s L^{-s})y_t = \varepsilon_t, \quad (1)$$

$$\phi(L)\varphi(L^{-1})y_t = \varepsilon_t, \quad (2)$$

with L the backshift operator, i.e., $Ly_t = y_{t-1}$ gives lags and $L^{-1}y_t = y_{t+1}$ produces leads. When $\varphi_1 = ... = \varphi_s = 0$, the process y_t is a purely causal autoregressive process, denoted AR$(r,0)$:

$$\phi(L)y_t = \varepsilon_t. \quad (3)$$

[2] Available upon request.

Model specification (3) can be seen as the standard backward-looking AR process, with y_t being regressed on y_{t-1} up to y_{t-r}. The process in (2) is a purely noncausal AR$(0, s)$ model

$$\varphi(L^{-1})y_t = \varepsilon_t, \tag{4}$$

when $\phi_1 = ... = \phi_r = 0$. Model specification (4) is the counterpart of (3), since it is a purely forward-looking AR process. That is, y_t is not regressed on its past values, but rather on its future values y_{t+1} up to y_{t+s}. Models of the form (2) that contains both lags and leads of the dependent variable, are called mixed causal-noncausal models.

The roots of both the causal and noncausal polynomial are assumed to lie outside the unit circle, that is $\phi(z) = 0$ and $\varphi(z) = 0$ for $|z| > 1$ respectively. These conditions imply that the series y_t admits a two-sided moving average (MA) representation $y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$, such that $\psi_j = 0$ for all $j < 0$ implies a purely causal process x_t (w.r.t. ε_t) and a purely noncausal model when $\psi_j = 0$ for all $j > 0$ (Lanne and Saikkonen, 2011). Error terms ε_t are assumed *iid* (and not only weak white noise) non-Gaussian with $E(|\varepsilon_t|^{\delta}) < \infty$, for some $\delta \in (0, 1)$.[3] Following Gouriéroux and Jasiak (2015), we define the unobserved causal and noncausal components of the process y_t as follows:

$$u_t \equiv \phi(L)y_t \leftrightarrow \varphi(L^{-1})u_t = \varepsilon_t,$$

and

$$v_t \equiv \varphi(L^{-1})y_t \leftrightarrow \phi(L)v_t = \varepsilon_t.$$

The specification of these filtered values is very useful in simulating, estimating and forecasting mixed causal-noncausal processes.

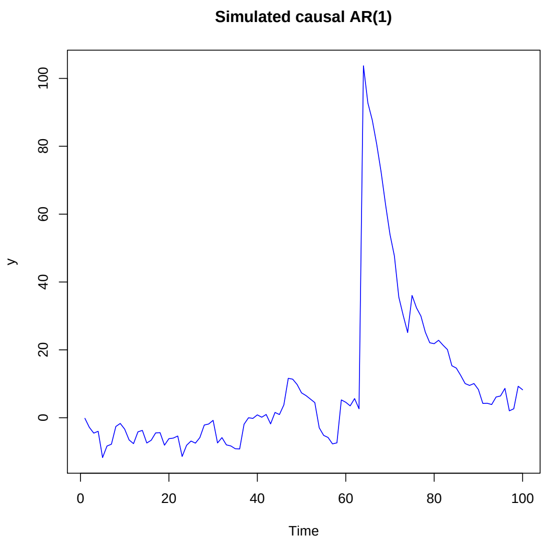
[3]Proposition 2.1 in Gouriéroux and Zakoïan (2013) shows that assuming the summability condition $\sum_{j=-\infty}^{\infty} |\psi_j|^{\delta} < \infty$ for some $\delta \in (0, 1)$ ensures absolute summability of the MA coefficients, i.e., the convergence of the series y_t with probability 1. Then the process y_t is well defined for any *iid* sequence ε_t such that $E(|\varepsilon_t|^{\delta}) < \infty$ for some $\delta \in (0, 1)$. This extension is extremely useful in the context where $E(|\varepsilon_t|)$ is not defined, like in the Cauchy case.

The non-Gaussianity assumption ensures the identifiability of the causal and the noncausal part (Breidt et al., 1991). Most papers by Lanne, Saikkonen and coauthors use $t_\nu-$distributions with $\nu \geq 3$ as an alternative to the Gaussian distribution while Gouriéroux and coauthors rely on the Cauchy or a mixture of Cauchy and Normal distributions. However, for most applications it emerges that the Cauchy has too strong fat tails features and many series would have (if estimated by a Student's t MLE) a degree of freedom between 1.5 and 2.5.

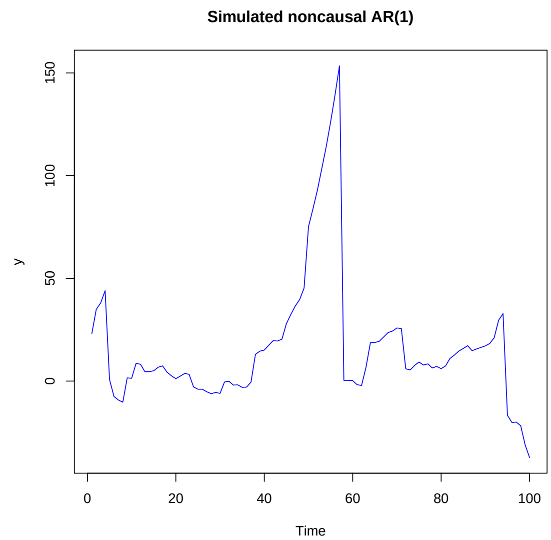
Figure 1 characterizes how purely causal and purely noncausal AR processes of the same order can create a different structure when a non-Gaussian error distribution is assumed (for $T = 100$). The first two graphs (1a,1b) show simulated paths[4] for both a causal and noncausal AR(1) with an autoregressive parameter value of 0.9. In both cases, ε_t is assumed to be standard Cauchy, i.e., a t_1-distribution. The pattern of the peaks in these graphs reveal the main difference: (i) in the causal case, a jump is followed by an exponential decrease and consequently cannot mimic the typical bubble pattern, while (ii) in the noncausal case, one observes a process of exponential growth followed by a drop. The two remaining graphs (1c,1d) show processes simulated from a model consisting of both a forward- and backward-looking component, namely the mixed causal-noncausal MAR(2,2) model. It can be seen how the combination of a causal and noncausal component can create asymmetric bubbles which encompass the features of both pure models discussed previously. This proves helpful for macroeconomic time series to describe asymmetric business cycles, for instance when recession and recovery phases have different amplitude. By means of choosing parameter values ϕ (respectively φ) close to unity it is possible to increase the causal (respectively noncausal) polynomial as driving force in the process.

Any estimation method based solely on second-order properties of the system will be unable to distinguish among causal and noncausal models. If y_t is a noncausal AR(p) driven by an *iid*

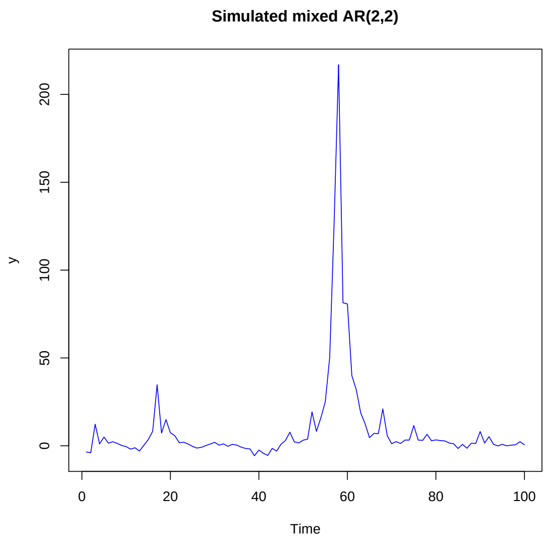
[4] MAR(r, s) models are simulated as follows. First generate a sequence of ε_t. Assume terminal values for u_t and construct the rest of the sequence according to $u_t = \varphi_1 u_{t+1} + ... + \varphi_s u_{t+s} + \varepsilon_t$. Subsequently, y_t can be constructed, since $y_t = \phi_1 y_{t-1} + ... + \phi_r y_{t-r} + u_t$ and only starting values for y_t are required. A burn-in period for u_t and y_t is recommended to remove the effects of the initial values.



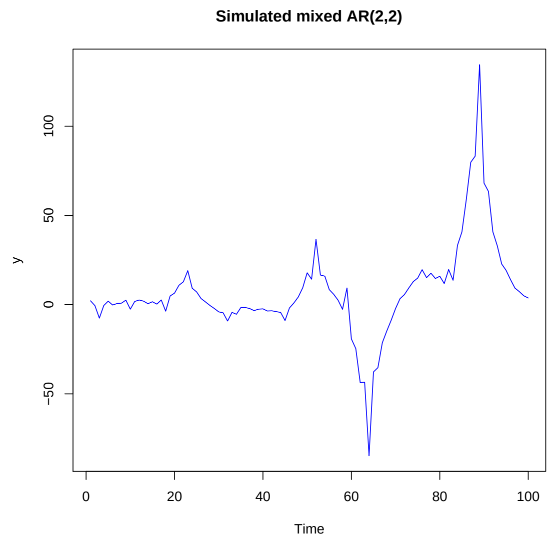
(a) Simulated causal MAR(1,0) process ($\phi = 0.9$) with standard Cauchy errors



(b) Simulated noncausal MAR(0,1) process ($\varphi = 0.9$) with standard Cauchy errors



(c) Simulated MAR(2,2) process ($\phi = (0.2, 0.3), \varphi = (0.2, 0.1)$) with standard Cauchy errors



(d) Simulated MAR(2,2) process ($\phi = (0.3, 0.3), \varphi = (0.4, 0.4)$) with standard Cauchy errors

Figure 1: Simulated processes from various MAR(r, s) specifications, $T = 100$

sequence of disturbances with mean zero and variance σ^2, then y_t can also be expressed as a causal AR(p) driven by a newly created white noise sequence with mean zero and variance $\tilde{\sigma}^2$. Both representations however yield exactly the same autocovariance function. In Appendix B it is shown that in the non-Gaussian case, this new sequence of innovations is generally not a strong white noise. Consequently, it is of interest to observe the potential misleading implications of estimating a pseudo-causal model when the data generating process is actually a MAR(r, s) (see Gouriéroux and Jasiak, 2014). Indeed, y_t in (1) also has a pseudo-causal AR(p) representation (see Appendix A) of order $p = r + s$ given by

$$a(L)y_t = \varepsilon_t^*. \tag{5}$$

Although uncorrelated, the error term ε_t^* in (5) is generally not independent anymore (see Brockwell and Davis, 1991 or Appendix B for a direct proof). Gouriéroux and Zakoïan (2013) exploit this property to discriminate between purely causal and purely noncausal AR(1) processes. Lanne and Saikkonen (2011) propose to use usual information criteria to first determine the pseudo-causal (Gaussian) AR(p). Hencic and Gouriéroux (2014) propose to look at the empirical ACF which is consistent even in the infinite variance case to determine p.

3 Estimation

Let us assume that we have rejected the Gaussianity null hypothesis and that we observe some non $iid-$ness when estimating a (pseudo-)causal autoregressive model. This deviation from Gaussianity can be interpreted as evidence that investigating a noncausal or mixed causal-noncausal specification is justified. In this section, we outline the estimation strategy of mixed causal-noncausal models as specified in Section 2, with the small modification that these models are augmented with an intercept α. If the error process ε_t possesses a density $f_\varepsilon(\varepsilon_t)$, the joint density of the data y_t can be approximated by

$$\prod_{r+1}^{T-s} f_\varepsilon(\varepsilon_t(y_t)|\phi, \varphi, \alpha, \lambda),$$

where the vector λ collects distributional parameters (e.g., scale parameters, degrees of freedom, etc.) necessary to characterize the density. Thus, the associated approximate likelihood function[5] is given by

$$\prod_{r+1}^{T-s} f_\varepsilon(\phi, \varphi, \alpha, \lambda|y_t).$$

Properties of the maximum likelihood estimators (MLEs) have been studied in Andrews, Breidt and Davis (2006) and Lanne and Saikkonen (2011). Both papers provide regularity conditions under which the MLE is consistent and asymptotically normal. The latter property requires in particular that $E(|\varepsilon_t|^\delta) < \infty$ for some $\delta \geq 2$. For consistency, the condition on δ is less restrictive. In particular, consistency when $\delta \in (0,1)$ has been shown in Gouriéroux and Zakoïan (2013). Footnote 3 of this paper summarizes this result.

In the econometric literature the Student's t-distribution (and the Cauchy as a particular case) has almost exclusively been used in applications, since it is a straightforward choice which allows to incorporate heavy tail behavior, often observed in financial data. Relatively little is known about the properties of the MLE of mixed causal-noncausal models if $f_\varepsilon(\varepsilon_t)$ is not assumed to be known. Properties of the quasi-MLE (QMLE) are derived in Huang and Pawitan (2000) and Wu and Davis (2010) for the specific case of a Laplacian objective function, resulting in the least absolute deviation estimator (LAD). Under suitable assumptions (restrictions) on the underlying noise sequence ε_t, the LAD estimator is consistent and asymptotically normal. In the sequel we present and assess the Student's t MLE and the LAD estimator in more detail.

[5]The term 'approximate' stems from the fact that the sample used in the likelihood contains only $T-(r+s)$ terms. As shown in Breidt et al. (1991), this quantity is only an approximation of the true joint density of the data vector $y = (y_1, ..., y_T)$.

3.1 Student's t Maximum Likelihood Estimation

If ε_t is a sequence of *iid* zero mean random variables with probability density function

$$f_\varepsilon(\varepsilon_t|\sigma,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}\left(1+\frac{1}{\nu}\left(\frac{\varepsilon_t}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}},$$

the corresponding (approximate) log-likelihood function, conditional on the observed data $(y_1,...,y_T)$ can be formulated as

$$\begin{aligned} l_y(\phi,\varphi,\alpha,\sigma,\nu|y_1,...,y_T) &= (T-p)\left[\ln(\Gamma((\nu+1)/2)) - \ln(\sqrt{\nu\pi}) - \ln(\Gamma(\nu/2)) - \ln(\sigma)\right] \\ &\quad -(\nu+1)/2\sum_{t=r+1}^{T-s}\ln(1+((\phi(L)\varphi(L^{-1})y_t-\alpha)/\sigma)^2/\nu), \end{aligned} \tag{6}$$

where we still have that $p = r+s$. The scale parameter is denoted by σ, $\nu > 0$ are degrees of freedom, and $\Gamma(\cdot)$ is the gamma function. Thus, the MLE corresponds to the solution of the problem:

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} l_y(\theta|y_1,..,y_T),$$

with $\theta = [\phi,\varphi,\alpha,\sigma,\nu]'$ and Θ is a permissible parameter space containing the true value of θ, say θ_0, as an interior point. Since an analytical solution of the score function is not directly available, gradient based (numerical) procedures like the Berndt-Hall-Hall-Hausman (BHHH) and Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithms can be used to find $\hat{\theta}_{ML}$.

If $\nu > 2$, and hence $E(|\varepsilon_t|^2) < \infty$, the MLE is $\sqrt{T}$-consistent and asymptotically normal. In particular, based on the unobserved causal and noncausal components discussed above, it is straightforward to show (see e.g., Fonseca, Ferreira, and Mignon, 2008) that in case of the mixed causal-noncausal model

$$\sqrt{T}(\hat{\phi}_{ML}-\phi_0) \sim \mathcal{N}(0,(\nu+3)/(\nu+1)\sigma^2\Upsilon_\varphi^{-1}), \tag{7}$$

$$\sqrt{T}(\hat{\varphi}_{ML} - \varphi_0) \sim \mathcal{N}(0, (\nu+3)/(\nu+1)\sigma^2\Upsilon_{\phi}^{-1}), \tag{8}$$

holds. We use the notations $v_t(\varphi) = \varphi(L^{-1})y_t$, $u_t(\phi) = \phi(L)y_t$ and then $\Upsilon_{\varphi} = E[v_t^2(\varphi)]$ and similarly $\Upsilon_{\phi} = E[u_t^2(\phi)]$. The variance-covariance matrices Υ_{ϕ} and Υ_{φ} can be consistently estimated by $\sum_{t=r+1}^{T-s} u_t^2(\hat{\phi}_{ML})/(T-p)$ and $\sum_{t=r+1}^{T-s} v_t^2(\hat{\varphi}_{ML})/(T-p)$, respectively.

For large ν, i.e., $\nu \to \infty$, l_y approaches the Gaussian (log)-likelihood, and the model parameters cannot be consistently estimated anymore. In Section 4 we further explore the properties of the MLE by means of simulations.

3.2 Least Absolute Deviation Estimation

The LAD criterion is derived via a likelihood approximation assuming that the underlying noise ε_t is mean zero Laplacian, i.e., $f_{\varepsilon}(\varepsilon_t|\sigma) = \frac{1}{2\sigma}\exp(-|\varepsilon_t|/\sigma)$, with σ denoting the scale parameter. Then the approximate log-likelihood function is

$$l_y(\phi, \varphi, \alpha, \sigma | y_1, ..., y_T) = -(T-p)\ln(2\sigma) - \frac{1}{\sigma}\sum_{t=r+1}^{T-s} |\phi(L)\varphi(L^{-1})y_t - \alpha|. \tag{9}$$

By maximizing (9) with respect to σ, we obtain the maximizer

$$\hat{\sigma} = (T-p)^{-1}\sum_{t=r+1}^{T-s} |\phi(L)\varphi(L^{-1})y_t - \alpha|,$$

which we can use to concentrate out σ in the same equation (9) to obtain the objective function

$$-(T-p)\left(1+\log(2)\right) - (T-p)^{-1}\ln\left(2(T-p)^{-1}\right) - \ln\left(\sum_{t=r+1}^{T-s} |\phi(L)\varphi(L^{-1})y_t - \alpha|\right),$$

which is (given T) maximized if the criterion $\sum_{t=r+1}^{T-s} |\phi(L)\varphi(L^{-1})y_t - \alpha|$ is minimized. Minimizing this criterion with respect to ϕ, φ, and α corresponds to a nonlinear program, which can be solved numerically with various approaches. We find the BFGS algorithm mentioned earlier to be a fast and stable variant.

Wu and Davis (2010) have shown[6] that if ε_t is a sequence of *iid* random variables with mean zero, median zero, finite variance and probability density function f_ε that is continuous in a neighborhood of zero, the LAD estimators (minimizing the modified objective function above) are $\sqrt{T}$-consistent and asymptotically normal. Again, based on the formulation of the unobserved causal and noncausal components and by using standard results in the literature (e.g., Knight, 1998) we find that

$$\sqrt{T}(\hat{\phi}_{LAD} - \phi_0) \sim \mathcal{N}\left(0, \frac{1}{4f_\varepsilon^2(0)}\Upsilon_\varphi^{-1}\right) \tag{10}$$

$$\sqrt{T}(\hat{\varphi}_{LAD} - \varphi_0) \sim \mathcal{N}\left(0, \frac{1}{4f_\varepsilon^2(0)}\Upsilon_\phi^{-1}\right). \tag{11}$$

In order to use these two expressions for inference, we need to estimate the density of the errors at zero. We can use a logistic kernel $K(x) = \exp(-x)/\left[(1+\exp(-x)\right]^2$ and estimate $f_\varepsilon(0)$ as

$$\hat{f}_\varepsilon(0) = \frac{1}{b_n(T-p)} \sum_{t=r+1}^{T-s} \frac{\hat{\phi}_{LAD}(L)\hat{\varphi}_{LAD}(L^{-1})y_t - \hat{\alpha}_{LAD}}{b_n}, \tag{12}$$

where we choose the bandwidth $b_n = 0.9A(T-p)^{-1/5}$, with $A = \min\left[std(\hat{\varepsilon}_t), iqr(\hat{\varepsilon}_t)/1.34\right]$, with *std* and *iqr* being the standard deviation and the interquantile range of the residuals, respectively (see Silverman, 1986).

4 Simulation Results

In this section, we want to assess how the asymptotic results derived in (7)-(8) and (10)-(11) behave in finite samples by means of Monte Carlo experiments. In particular, we are interested in how the estimators behave in "boundary" cases where critical assumptions are violated. This

[6]However, based on a slightly different formulation of the initial mixed causal-noncausal model (see also Lanne and Saikkonen (2011, section 3) for the equivalence of both representations).

includes for example the case when the error distribution is 'very close' to a Gaussian one, or has infinite variance. The data are generated by a mixed causal-noncausal model of lag and lead order one:

$$(1 - 0.3L)(1 - 0.7L^{-1})y_t = \varepsilon_t, \tag{13}$$

under varying assumptions on the error distribution and different sample sizes ($T = 200$ and $T = 800$). The number of replications is 10,000 in all simulation experiments. More specifically, we investigate the finite sample properties of the estimators under the assumption that the errors ε_t follow

1. a $t-$distribution with 10 degrees of freedom.
2. a $t-$distribution with 3 degrees of freedom. The skewness and kurtosis are consequently undefined.
3. a $t-$distribution with 1.5 degrees of freedom. In this case the variance, the skewness and the kurtosis are undefined.
4. a normal mixture $\varepsilon_t = p_1x_1 + p_2x_2 + p_3x_3$ with equal probabilities $p_j = 1/3$, $j = 1, 2, 3$ and $x_1 \sim \mathcal{N}(-1, 2)$, $x_2 \sim \mathcal{N}(0, 0.5)$, $x_3 \sim \mathcal{N}(1, 2)$

To get a first impression of the behavior of $\hat{\phi}$ and $\hat{\varphi}$ we plot in Figure 2 for $T = 200$ and in Figure 3 for $T = 800$ the histograms of the estimators obtained in all simulation repetitions. The true parameter values are set to $\phi = 0.3$ and $\varphi = 0.7$. We can observe that the more the underlying distribution of the errors is approaching the normal distribution the less precise the resulting estimates. Moreover, the heavier the tails the faster the estimator seems to converge. This result is known for infinite variance autoregressions (see Davis et al., 1992). The LAD seems to be slightly more accurate than the MLE, in particular when $T = 200$ for the case 4 (asymmetric distribution). This is a very interesting result because identification of the mixed causal-noncausal model can be done for relatively small samples, for very fat tails distributions

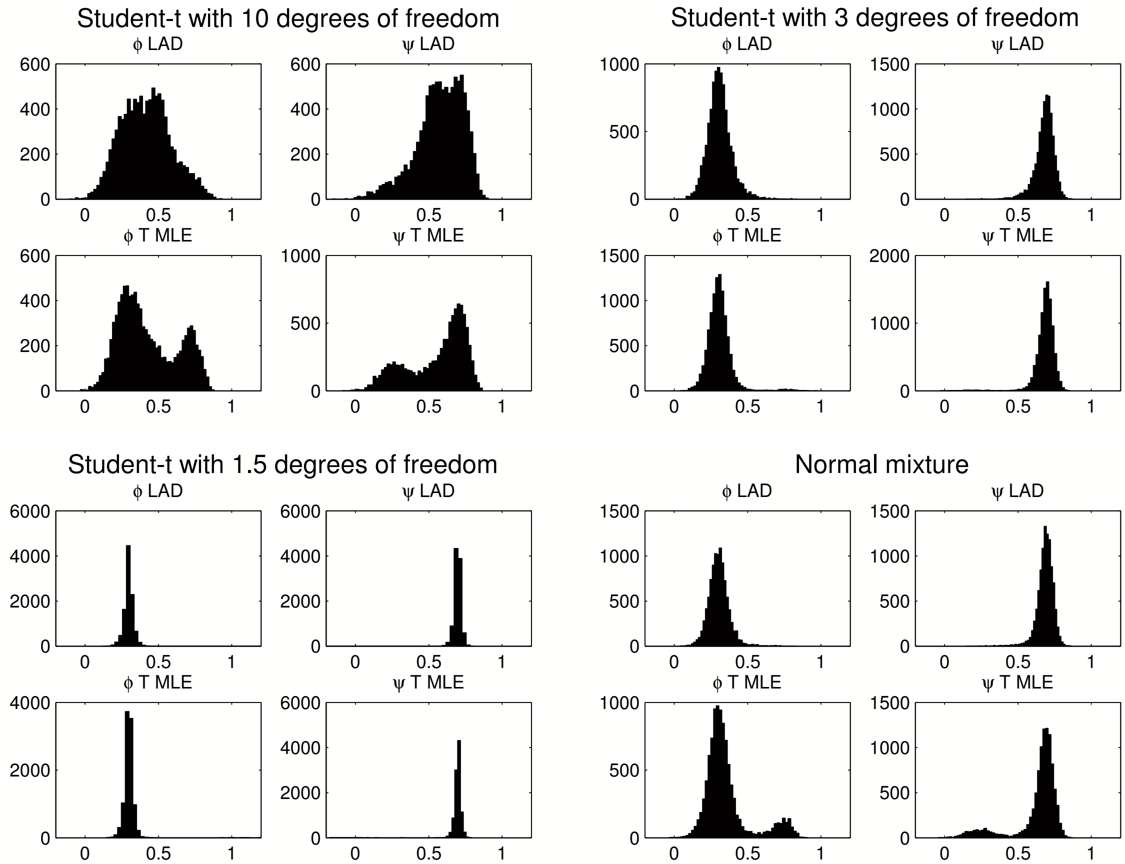


Figure 2: Histograms of estimators based on $T = 200$ observations

and without a reference to a well specified parametric distribution in a MLE approach such as a $t-$distribution.

Table 1 summarizes the results of the simulation study in terms of means and standard deviations (SD) of the estimated parameters over all replications. The entry AD corresponds to the asymptotic standard deviation computed as the average over the 10,000 replications. We can see that for cases 2, 3 and 4 (i.e., all our specifications but the Student's t_{10}) point estimates are very close to the ones in the DGP, namely $\phi = 0.3$ and $\varphi = 0.7$. The only exception is the MLE for the mixture in small samples. Both MLE and LAD present some bias for the $t-$distribution with 10 degrees of freedom. To the question: how fat must the the tails of the

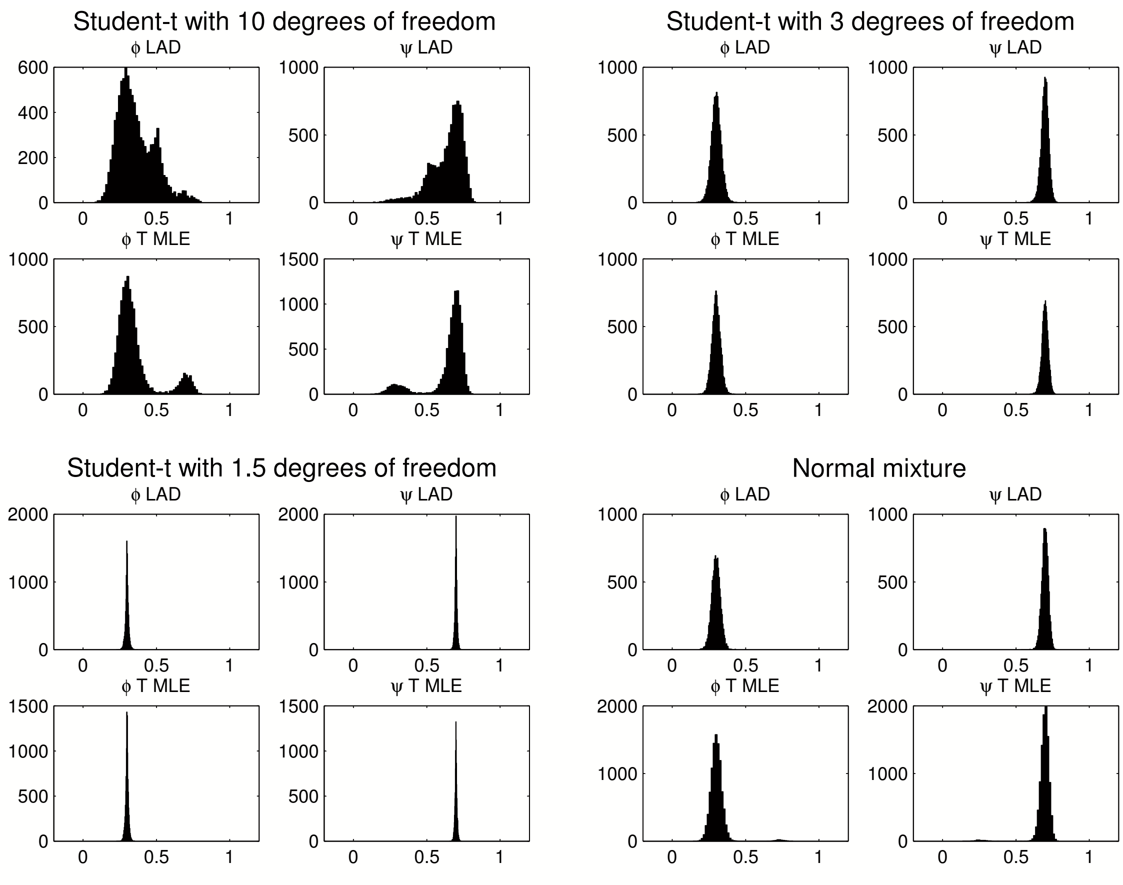


Figure 3: Histograms of estimators based on $T = 800$ observations

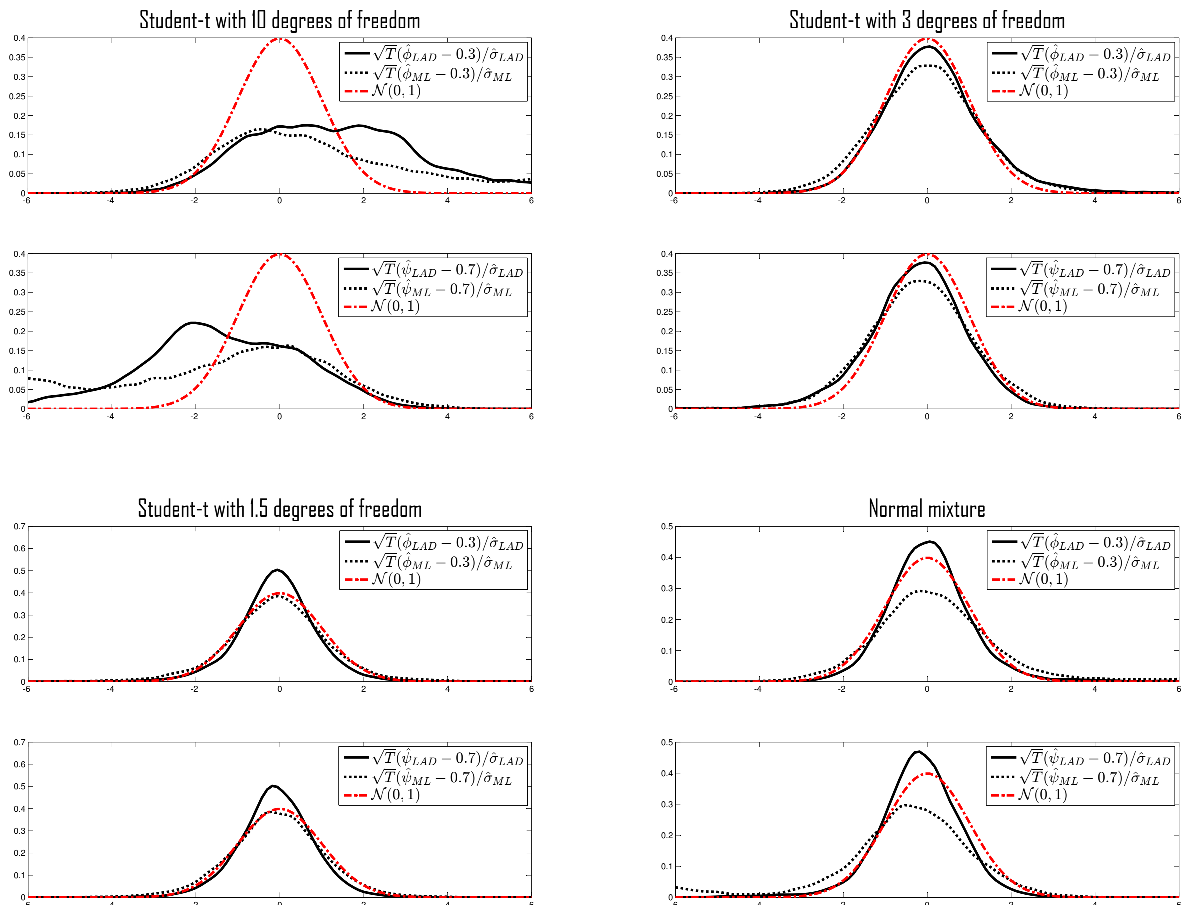


Figure 4: Density plots of the estimators's sampling distribution and their asymptotic limit based on $T = 200$ observations

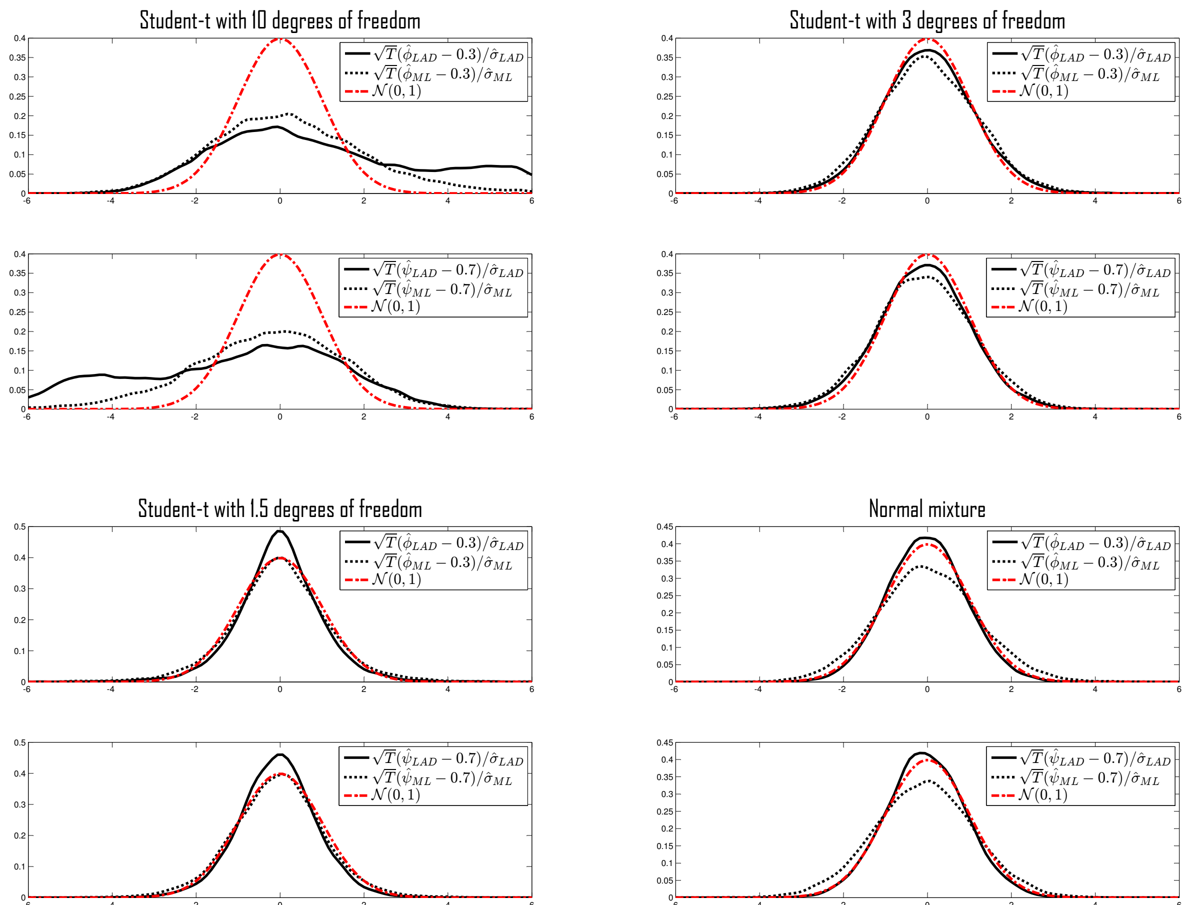


Figure 5: Density plots of the estimators's sampling distribution and their asymptotic limit based on $T = 800$ observations

distribution of the error process be in practice, we can conclude that the heavier the tails are, the more accurate the estimation of the parameters is. This is particularly true when using the LAD but also the Student's t MLE performs well. This is in line with results in Davis et al. (1992), who find that both the MLE and LAD estimator converge faster in the case of infinite variance autoregressive processes. When comparing the estimated standard error (SD) to the asymptotic one (AD) it emerges that the approximation is poor for t_{10} but very accurate for more leptokurtic and asymmetric distributions and in particular when using the LAD estimator. This is illustrated in Figures 4 and 5 where it is shown how fast the asymptotic approximations in (7)-(8) and (10)-(11) approach their distributional limit.

		t_{10}		t_3		$t_{1.5}$		Mixture	
		ϕ_{LAD}	φ_{LAD}	ϕ_{LAD}	φ_{LAD}	ϕ_{LAD}	φ_{LAD}	ϕ_{LAD}	φ_{LAD}
Mean	$T=200$	0.410	0.575	0.309	0.681	0.299	0.696	0.304	0.686
	$T=800$	0.363	0.649	0.300	0.697	0.299	0.699	0.299	0.698
SD	$T=200$	0.169	0.160	0.082	0.067	0.044	0.033	0.071	0.057
	$T=800$	0.125	0.111	0.033	0.025	0.011	0.008	0.030	0.023
AD	$T=200$	0.083	0.074	0.069	0.052	0.033	0.026	0.069	0.055
	$T=800$	0.041	0.034	0.031	0.024	0.011	0.008	0.031	0.023
		ϕ_{ML}	φ_{ML}	ϕ_{ML}	φ_{ML}	ϕ_{ML}	φ_{ML}	ϕ_{ML}	φ_{ML}
Mean	$T=200$	0.429	0.554	0.306	0.683	0.301	0.696	0.355	0.631
	$T=800$	0.353	0.644	0.299	0.698	0.299	0.699	0.306	0.691
SD	$T=200$	0.199	0.198	0.084	0.077	0.055	0.042	0.155	0.155
	$T=800$	0.138	0.135	0.029	0.022	0.010	0.007	0.057	0.053
AD	$T=200$	0.060	0.055	0.050	0.038	0.022	0.017	0.056	0.047
	$T=800$	0.032	0.026	0.025	0.018	0.009	0.006	0.029	0.022

Table 1: Means and standard deviations of the MLE and LAD estimators

The next section illustrates our main findings in two empirical applications on both monthly variables and 5-min return series.

5 Empirical Illustration 1: Solar Panel Bubbles

5.1 The Data

We first investigate the existence of a bubble observed in the solar panels market in Belgium (more exactly the Walloon Region, i.e., the southern part of Belgium). Photovoltaic systems are technologies that use solar panels to directly convert the solar energy into electricity. This technology has been promoted in many countries, in Europe in particular, in order to reduce carbon dioxide emissions. While solar panels are nowadays relatively cheap due to the massive imports from China, this was not the case until roughly 2010, say. Thus, policy makers, for instance to meet the 1997 Kyoto Protocol that commits parties to reduce greenhouse gases emissions, have started to subsidize installation of such photovoltaic systems at the end of the 90's. This financial assistance has indeed helped to make this technology popular but has also created some side effects. Solar panels prices paid by final consumers did not tend to decrease as fast as expected in the short-run because some companies have maintained high prices, incorporating those financial helps in their bills to consumers.

Figure 6 displays the series we use on the number of solar panels installed (Nombre d'UPD, denoted $Units_t$ hereafter) and the power (Puissances, denoted $Power_t$ hereafter) they generate. Monthly variables spanning 2008:01-2014:08 are released by the *Commission Wallone pour l'Energie* (CWaPE), the official organization that regulates the electricity and gas markets.[7] The familiar bubble and burst pattern is obvious in both series. To some extent, the bubble could have been foreseen for several reasons. First, regional policy makers did not adapt the subsidies (greenhouse certificates) to the new market conditions (price decreased on the world market). Thus the internal return provided by the solar panels was much above what the financial market had to offer. The second problem was due to the financial crisis hitting Europe after 2008. This has imposed budget cuts such that the financial assistance had to stop. This development has

[7] See http://www.cwape.be/

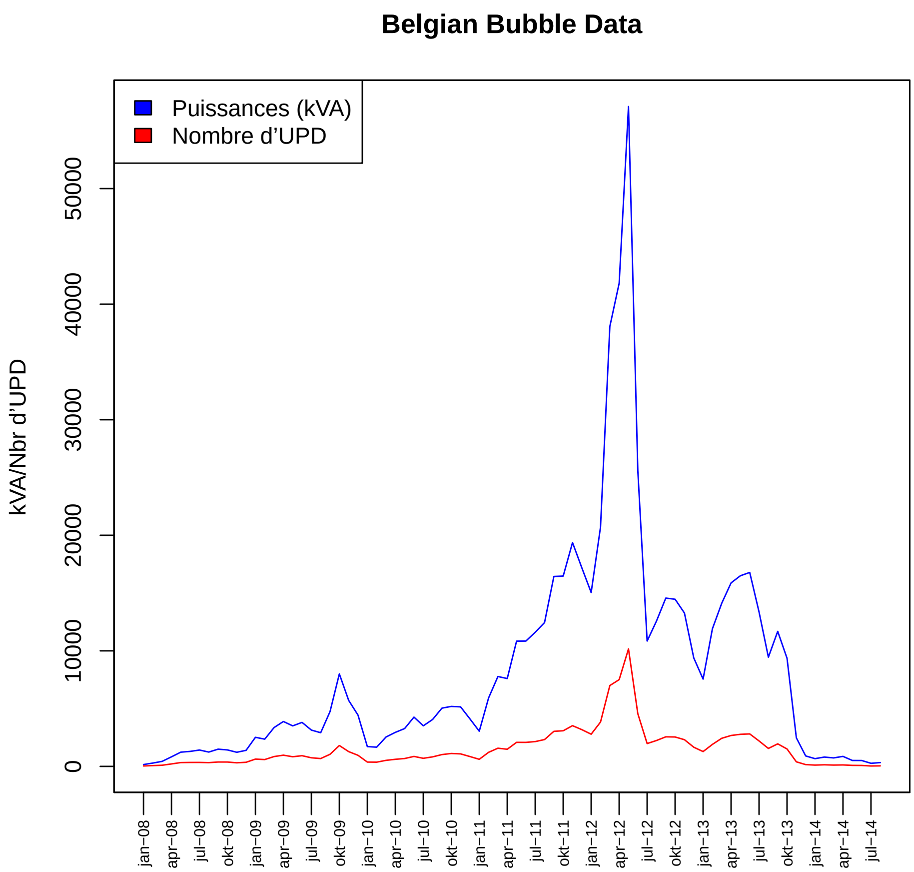


Figure 6: Number of solar panel units and power

been communicated far in advance. Consequently, this announcement has created a boom and then the burst phenomenon characteristic of such a bubble. It is described in the literature that regulatory changes give rise to a sharp decrease contrarily to expectations and sentiment reversals that produce a more gradual decrease (dot.com bubble in the US for instance).

5.2 Implementation of MAR(r, s) Models

In order to test for causal and noncausal models in those series we will follow the following strategy:

1. We determine an autoregressive process using information criteria, AIC, BIC or HQ for instance. Even in the presence of a noncausal component we obtain an estimate of the pseudo-causal order $p = r + s$. Simulation results (available upon request) would favor the use of BIC.

2. We test for the null of normality on the residuals of the AR(p) using e.g., the Jarque-Bera[8] and one-sample Kolmogorov-Smirnov (KS) test. It should be noted, that in case the null hypothesis of normality is not rejected, there is no need to consider noncausal or mixed causal-noncausal models, as the causal and noncausal polynomial cannot be distinguished.

3. We test for *iid*-ness. If the true process (i.e., the DGP) is *iid* causal, then the errors from the AR(p) must be *iid*. One should reject the *iid* null hypothesis if we estimate a pseudo-causal model using data generated by an *iid* noncausal DGP. Several tests can be used (e.g., BDS test) but we propose for simplicity to look at the residuals from the pseudo-causal AR(p), denoted $\hat{\varepsilon}_t$, and to run a multivariate regression from $\hat{\varepsilon}_t$ on $\hat{\varepsilon}^2_{t-1}$ up to $\hat{\varepsilon}^2_{t-m}$:
$$\hat{\varepsilon}_t = \mu + \delta_1 \hat{\varepsilon}^2_{t-1} + ...\delta_m \hat{\varepsilon}^2_{t-m} + u_t,$$
and to test for *iid*-ness using the null hypothesis $H_0 : \delta_1 = ... = \delta_m = 0$. This test is asymptotically $\chi^2_{(m)}$ distributed under normality. Because this is not the case here, its distribution can be tabulated. This is the natural extension to the procedure proposed by Gouriéroux and Zakoïan (2013) for $m = 1$ lag.

[8]Strictly speaking, the Jarque-Bera test does not test for normality, but for the skewness and kurtosis of the data to match the normal distribution. We acknowledge this fact, but because of the close correspondence, we treat the Jarque-Bera test as a test for normality.

4. After we established the order p that captures the autocorrelation in the series and we find that non-normality is detected and the *iid*-ness of the pseudo-causal residuals can be rejected, the next step is to select a model among all MAR(r, s) specifications with $r + s = p$. To do so, we apply the Student's t MLE as well as the LAD estimators that we described in Section 3 for each potential model (within $p = r + s$); we take the one that gives the highest value for the likelihood. Note that every model has got the same number of regressors, hence the use of information criteria is superfluous here. Note that, although an intercept is included in our specifications (6) and (9), demeaning the series first (but leaving the intercept in the regression) seems to improve the results in terms of speed and accuracy of the convergence process.

5.3 MLE and LAD Results

Table 2 provides the pseudo-AR(p) models (estimated by OLS) for both series using usual BIC, the value of the Bera-Jarque normality test, the LM tests for the null hypothesis of no-autocorrelation and *iid*-ness. Those tests are obtained on residuals after having estimated AR(1) processes. The white noise null hypothesis is not rejected but both the normality and *iid* hypotheses are. The Dickey-Fuller test does not reject the null marginally (p-values of 0.06) but it is well known that the test suffers from both size and power distortions in this situation (see e.g., Gouriéroux and Zakoïan, 2013 or Saikkonen and Sandberg, 2013).

We observe that there is a strong rejection of the null of normality. This is a first indication that it is justified to investigate the presence of noncausality in the series of interest. Using BIC, an AR(1) is selected for both series. We will thus consider either a purely causal MAR(1,0) or a purely noncausal MAR(0,1). Table 2 also reports the value of the log-likelihood (LL) and we see that for both series a noncausal MAR(0,1) is preferred for $p = 1$.

The two MAR(0,1) equations are estimated both by maximum likelihood and LAD. The 95% confidence intervals are in brackets below estimated parameters. There are no large differences

	$Units_t$	$Power_t$
BIC	AR(1)	AR(1)
$BJ : H_0 = normality$	727.76	748.21
$LM[1-1] : H_0 = white\ noise$	0.00	0.03
$LM[1-4] : H_0 = white\ noise$	1.17	1.11
$LM[1-1] : H_0 = iid$	19.18	19.62
$LM[1-4] : H_0 = iid$	9.78	9.38
$ADF - test$	-2.78	-2.77
LAD		
LL MAR(1,0)	-680.071	-817.068
LL MAR(0,1)	-678.725*	-814.834*
MLE		
LL MAR(1,0)	-593.189	-724.129
LL MAR(0,1)	-588.327*	-718.403*

Table 2: Summary Statistics

whether one takes the number of solar panels or the power they generate, but we observe that the parameters estimated by LAD are more persistent such that the confidence intervals marginally include the unit root case. We find the MLE results more plausible although the small number of observations does not allow to draw sharp conclusions; the estimated degrees of freedom tell us that we are close to the Cauchy distribution for the error process.

$$\text{Student's } t \text{ MLE} \begin{cases} \widehat{Units}_t = \underset{[0.893,0.964]}{0.928} Units_{t+1}, & \hat{\nu} = 1.09 \\ \widehat{Power}_t = \underset{[0.878,0.967]}{0.907} Power_{t+1}, & \hat{\nu} = 0.98 \end{cases}$$

$$\text{LAD} \begin{cases} \widehat{Units}_t = \underset{[0.908,1.007]}{0.957} Units_{t+1}, \\ \widehat{Power}_t = \underset{[0.928,1.010]}{0.968} Power_{t+1}, \end{cases}$$

An additional outcome that is out of the scope of this paper is whether the econometrician could have anticipated the crash of the bubble. Gouriéroux and Jasiak (2015) propose such a risk measure and proceed in the following way. One cuts of the data from the crash onwards, estimates a MAR(r,s) on the remaining observations, say $(y_1,...,y_{T^*})$, and uses these results to derive a joint predictive density for future values $(y_{T^*+1},...,y_{T^*+H})$. From these predictive densities probabilities can be computed for different types of scenarios like e.g., $y_{T^*} < y_{T^*+1} < y_{T^*+2}$, which signifies the probability that the bubble keeps on increasing in the upcoming two time periods (given the last data point y_{T^*}). Our sample is too small to realize this exercise.

6 Application 2: Realized Volatilities

The previous application was based on a very small sample size, around 75 data points. This can affect the outcomes in two ways. First, we have seen in Table 1 that results are obviously more accurate for larger samples. A second opposite effect is that the bubble phenomena that we can clearly observe in small samples (both in simulated graphs of Section 2 and on Belgian solar panels data) can be considered as a zoom of a larger process. When one constructs a series with thousands of observations, we stack to some extent those processes next to each other and the concatenated series generated/observed on a larger sample looks more random. In this second application we identify mixed causal-noncausal models in daily realized variances for almost 4,000 observations. To the best of our knowledge, this is the first time that these type of models are used on this kind of data. There are no doubts that Gaussianity as well as *iid*-ness is rejected for all series. Hence, we do not repeat steps 2 and 3 of the estimation strategy of subsection 5.2.

The realized variances are computed using

$$RV_t \equiv \sum_{i=1}^{M} r_{t,i}^2,$$

where $r_{t,i}$ are the high frequency intra-day returns, observed for M intra-day periods each day. For instance $M = 288$ for 5-min returns on 24 hours or around 80 when the market is open between 9am to 4pm. In order to attenuate the impact of jumps, we work in this paper with bipower variation (BV henceforth) series instead of realized variances (Barndorff-Nielsen and Shephard, 2004). They are computed as (Bauwens, Hafner and Laurent, 2012)

$$BV_t \equiv \frac{\pi}{2}\left(\frac{M}{M-1}\right)\sum_{i=2}^{M}|r_{t,i}||r_{t,i-1}|.$$

The bipower variations also provide a consistent measure of the integrated volatility associated with standard theoretical continuous time diffusion models. However, BVs are designed to be more robust to jumps than RVs because they are computed on products between two consecutive returns instead of the squared returns. We also consider in the application the median realized volatility measure (see Andersen, Dobrev, Schaumburg, 2010)

$$Med - RV_t \equiv \frac{\pi}{6-4\sqrt{3}+\pi}\left(\frac{M}{M-2}\right)\sum_{i=2}^{M-1}\text{med}\left(|r_{t,i}||r_{t,i-1}||r_{t,i+1}|\right)^2.$$

Our series are obtained from 5 minute data and are from Oxford-Man Institute of Quantitative Finance, Library version 0.2 (series code bv5 & medrv).[9] We use the period 03/01/2000 to 08/10/2014 (i.e., $T = 3858$ observations[10]) for twenty-one equity indexes whose names are given in the following Table 3

In this analysis we have replaced for simplicity the missing days, that are due to closing market, by the average of the series over the whole period. Alternative approaches can obviously be used. We have not divided the sample into different subperiods and hence the financial crisis is also included in our samples. The last two columns of Table 3 report the final models chosen by BIC using the Student's t MLE approach. We report results for both BV and Med-RV series. As an example, Figure 7 shows the computed Med-RV and BV series for the equity indexes of S&P500, FTSE, Nikkei and DAX.

On similar series (Corsi, Audrino and Renò, 2012) the heterogenous autoregressive model

[9] See Heber, Gerd, Asger Lunde, Neil Shephard and Kevin K. Sheppard (2009). "Oxford-Man Institute's realized library", Oxford-Man Institute, University of Oxford.

[10] We do not consider the first 600 observations that are not available for the Canadian index.

Ox.Man name	Common name	Country	MAR(r,s)-BV	MAR(r,s)-Med RV
AEX	AEX Index	Netherlands	MAR(6,2)	MAR(7,1)
AORD2	All Ordinaries	Australia	MAR(8,2)	MAR(2,7)
BVSP	Bovespa Index	Brazil	MAR(10,4)	MAR(12,6)
DJI2	DJIA	USA	MAR(7,5)	MAR(8,3)
FCHI2	CAC 40	France	MAR(9,0)	MAR(5,3)
FTSE2	FTSE 100	UK	MAR(7,2)	MAR(5,3)
FTSEMIB	FTSE MIB	Italy	MAR(5,1)	MAR(6,3)
FTSTI	FT straits time index	Singapore	MAR(6,15)	MAR(0,16)
GDAXI2	DAX	Germany	MAR(12,2)	MAR(12,2)
GSPTSE	S&P/TSX Composite Index	Canada	MAR(14,13)	MAR(15,13)
HSENG	Hang Seng	Asia	MAR(12,1)	MAR(17,1)
IBEX2	IBEX 35	Spain	MAR(5,1)	MAR(5,1)
IXIC2	Nasdaq 100	USA	MAR(4,7)	MAR(8,3)
KS11	KOSPI Composite Index	South Korea	MAR(5,2)	MAR(6,1)
MXX	IPC Mexico	Mexico	MAR(11,0)	MAR(6,2)
N2252	Nikkei 225	Japan	MAR(17,1)	MAR(12,6)
NSEI	S&P CNX Nifty	India	MAR(5,1)	MAR(1,1)
RUT2	Russell 2000	USA	MAR(9,5)	MAR(4,7)
SPX2	S&P 500	USA	MAR(6,7)	MAR(10,1)
SSMI	Swiss Market Index	Switzerland	MAR(7,2)	MAR(7,6)
STOXX50E	Euro STOXX 50	Eurozone	MAR(12,4)	MAR(10,2)

Table 3: Equity Indexes and MAR(r, s) identification

(HAR hereafter) proposed by Corsi (2009) to approximate the long memory features of realized volatilities is commonly used. The HAR for the RV_t variable is such that

$$RV_t^{(d)} = \beta_0 + \beta^{(d)} RV_{t-1d}^{(d)} + \beta^{(w)} RV_{t-1d}^{(w)} + \beta^{(m)} RV_{t-1d}^{(m)} + v_t, \quad t = 1, 2, ..., T,$$

where (d), (w), and (m) denote, respectively, the influence from yesterday and moving average impacts of last week average (assuming 5 days a week) and last month average (assuming 20 days a month) with $RV_t^{(w)} = \frac{1}{5}\sum_{j=0}^{4} RV_{t-jd}^{(d)}$ and $RV_t^{(m)} = \frac{1}{20}\sum_{j=0}^{19} RV_{t-jd}^{(d)}$. Hence, this HAR is also a restricted form of a causal (i.e, autoregressive) AR(20).

Developing a similar framework for mixed causal-noncausal models is out of the scope of this paper, in particular the multiplicative nature of the MAR(r, s) necessitates to adapt the

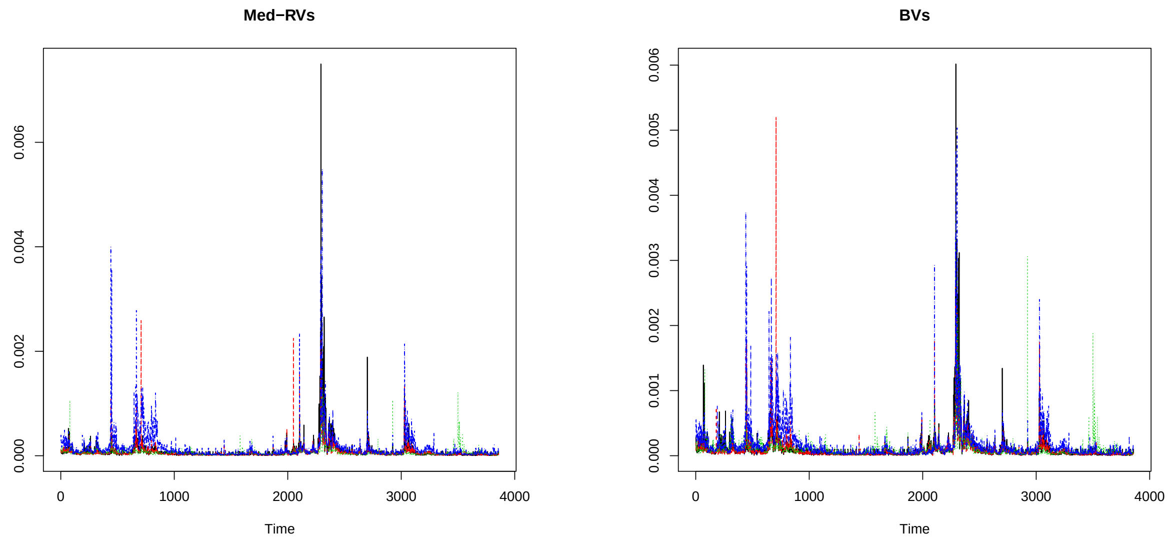


Figure 7: Med-RVs and BVs for S&P500, FTSE, Nikkei and DAX

HAR structure. However we want to allow for a dependence in the past and in the future at rather long displacements, namely at least one month in both time directions. We start by determining the pseudo-causal model for each series allowing a $p_{\max}$ of 50 days. Then, BIC is used to determine p in the pseudo-causal models, even though it is known that information criteria probably overestimate the number of lags for those volatile series (see e.g., Hecq, 1996). Depending on the indexes, p ranges from 2 to 28. Next, we search using the Student's t MLE approach, within $p = r + s$, the MAR(r, s) model that maximizes the log-likelihood. Table 3 reports those MAR(r, s) models for both volatility indicators. There are only two entries (resp. one entry) out of 42 series for which we get a pure causal (resp. purely noncausal) model. In most situations mixed causal-noncausal models are selected. We find degrees of freedom for the $t-$distribution between 1.1 and 2 for almost every series in our dataset. We leave the value added of considering the additional noncausal part for forecasting volatilities for further research, but this is clearly an important issue. Also the interaction between MAR(r, s) models and time-varying volatilities is not considered here.

7 Conclusion

We do not claim that noncausality is everywhere. However we must recognize that the Gaussianity assumption is rejected in many macroeconomic and financial time series. This gives the opportunity to extend the usual Box-Jenkins approach in three different ways: (*i*) y_t is assumed to have a two-sided moving average representation, including both a causal component (current and lagged values of the disturbance term) and a noncausal component (future values of the error term) (*ii*) The error term is strong white noise, instead of weak white noise; (*iii*) The distribution of the error term can be leptokurtic, with the possibility of having infinite variance or even infinite expectation.

Our paper compares the small sample performances of the MLE, assuming a t-distribution, with the LAD estimator. We show in a simulation study that both methods capture the causal and/or noncausal dependency particularly well when the distribution has very fat tails. These results are in line with findings in Davis et al. (1992) who find that both the MLE and LAD estimator converge faster for causal autoregressive processes with infinite variance than their counterparts with finite variance. The simulation study in this paper shows that these results hold for mixed causal-noncausal processes as well.

In the empirical section we investigate two types of data: monthly solar panel data and daily realized volatility series. In both applications, the presence of a noncausal component is detected (in most cases) and the estimated degrees of freedom for the t-distribution are found to be very small. Hence, the data confirms the interest in mixed causal-noncausal processes with potentially infinite variance or even expectation. The results of the MLE and LAD estimator are found to be very similar.

References

[1] Alessi, L., Barigozzi, M. and M. Capasso (2011), Non-Fundamentalness in Structural Econometric Models: A Review, *International Statistical Review,* 79, 1.

[2] Andersen, T.G., Dobrev, D. and E. Schaumburg (2010), Jump-Robust Volatility Estimation Using Nearest Neighbor Truncation, Staff Report no. 465, Federal Reserve Bank of New York.

[3] Andrews, B., Breidt, F. and R. Davis (2006), Maximum Likelihood Estimation For All-Pass Time Series Models. *Journal of Multivariate Analysis, 97, 1638-1659.*

[4] Barndorff-Nielsen, O. E. and N. Shephard (2004), Power and Bipower Variation with Stochastic Volatility and Jumps (with Discussion), *Journal of Financial Econometrics*, 2, 1-37.

[5] Bauwens, L., Hafner, Ch. and S. Laurent (2012), Volatility Models, *Handbook of Volatility Models and Their Applications* (eds. Bauwens, L., Hafner, Ch. and S. Laurent), Wiley, Ch. 1.

[6] Beaudry, P., Fève, P., Guay, A. and F. Portier (2015), When is Nonfundamentalness in VARs a Real Problem? An Application to News Shocks, CEPR discussion paper 10763.

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

[8] Brockwell, P. and R. Davis (1991), *Time Series: Theory and Methods*, Springer-Verlag New York, Second Edition.

[9] Brockwell, P. and R. Davis (2002), *Introduction to Time Series and Forecasting*, Springer-Verlag New York, Second Edition.

[10] Corsi, F. (2009), A Simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics*, Vol. 7, No.2, 174-196.

[11] Corsi, F., Audrino, F. and R. Renò (2012), *HAR Modeling for Realized Volatility Forecasting*, Ch. 15 in Handbook of Volatility Models and Their Applications, Wiley, Ed. by Bauwens, Hafner and Laurent.

[12] Davis, R., Knight, K. and J. Liu (1992), M-Estimation for Autoregressions with Infinite Variance, *Stochastic Processes and their Applications*, 40, 145-180

[13] Fonseca, T., Ferreira, M. and H. Mignon (2008), Objective Bayesian Analysis for the Student-t Regression Model, *Biometrika*, 95, 325-333.

[14] Gouriéroux, C. and J. Jasiak (2014), Misspecification of Causal and Noncausal Orders in Autoregressive Processes, *CREST Working Paper*.

[15] Gouriéroux, C. and J. Jasiak (2015), Filtering and Prediction in Noncausal Processes. *CREST Working Paper*.

[16] Gouriéroux, C., and J.M. Zakoïan (2013), Explosive Bubble Modeling by Noncausal Process, *CREST Discussion Papers 2013-04*.

[17] Gouriéroux, C., and J.M. Zakoïan (2014), On Uniqueness of Moving Average Representations of Heavy-Tailed Stationary Processes, *CREST Discussion Papers 2014-01*.

[18] Hansen, L. and T. Sargent (1991), Two Difficulties in Interpreting Vector Autoregressions, in *Rational Expectation Econometrics*, ed. Hansen and Sargent, pp. 77-119, Westview Press, Boulder.

[19] Hecq, A. (1996), IGARCH Effects on Autoregressive Lag Length Selection and Causality Tests, *Applied Economic Letters*, 3, 317-323.

[20] Hecq, A., Lieb, L. and S. Telg (2015), Forecasting Inflation in Europe with Mixed Causal-Noncausal Models, *Working paper*.

[21] Hencic, A. and C. Gouriéroux (2014), Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rate, *Crest Working paper*.

[22] Homm, U. and J. Breitung (2012), Testing for Speculative Bubbles in Stock Markets: a Comparison of Alternative Methods, *Journal of Financial Econometrics*, 10(1), 198-231.

[23] Huang J. and Y. Pawitan (2000), Quasi-likelihood Estimation of Noninvertible Moving Average Processes, *Scandinavian Journal of Statistics*, 27, 689-710.

[24] Knight K. (1998), Limiting Distributions for L_1 Regression Estimators under General Conditions, *The Annals of Statistics*, 26(2), 755-770.

[25] Lanne, M., Luoto J. and P. Saikkonen (2012), Optimal Forecasting of Noncausal Autoregressive Time Series, *International Journal of Forecasting*, 28, 623-631.

[26] Lanne, M., Nyberg, H. and E. Saarinen (2012). Does Noncausality Help in Forecasting Economic Time Series?, *Economics Bulletin*, 32(4), 2849-2859.

[27] Lanne, M. and P. Saikkonen (2011), Noncausal Autoregressions for Economic Time Series, *J ournal of Time Series Econometrics*, vol.3, iss.3.

[28] Lii, K. and Rosenblatt, M. (1993), Non-Gaussian Autoregressive Moving Average Processes. *Proceedings of the National Academy of Science of the United States of America, 90(19), 9168-9170.*

[29] Lippi, M. and L. Reichlin (1993a), The dynamic Effects of Aggregate Demand and Supply Disturbances, *American Economic Review*, 83(3), 644-52.

[30] Lippi, M. and L. Reichlin (1993b), VAR Analysis, Nonfundamental Representations, Blaschke Matrices, *Journal of Econometrics*, 63, 307-325.

[31] Lof, M. and H. Nyberg (2015), Noncausality and the Commodity Currency Hypothesis. *Working Paper*, Available at SSRN: http://ssrn.com/abstract=2597815.

[32] Phillips, P.C., Wu, Y. and J. Yu (2011), Explosive behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values?, *International Economic Reviews*, 52(1), 201-226.

[33] Saikkonen, P. and R. Sandberg (2013), Testing for a Unit Root in Noncausal Autoregressive Models, *Discussion Papers of Bank of Finland Research.*

[34] Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London – New York, 175

[35] Scherbina, A. (2013), Asset Price Bubbles: A Selective Survey, *IMF working paper WP/13/45.*

[36] Stiglitz, J. (1990), Symposium on Bubbles, *Journal of Economic Perspectives, Vol. 4, No. 2, 13-18.*

[37] Wu R. and R. Davis (2010), Least Absolute Deviation Estimation for General Autoregressive Moving Average Time-Series Models. *Journal of Time Series Analysis*, 31, 98-112.

8 Appendix

8.1 Appendix A: Pseudo-Causal Representation

To get to the pseudo-causal model from a mixed framework, let us consider a MAR(r, s)

$$\phi(L)(1 - \varphi_1 L^{-1} - ... - \varphi_s L^{-s})y_t = \varepsilon_t.$$

Then one takes the polynomial $\tilde{\varphi}(L)$ with roots equal to the inverse of the roots of $\tilde{\varphi}(L)$ such that

$$\phi(L)\left(-\varphi_s L^{-s}\right)\left(-\frac{1}{\varphi_s}L^s + \frac{\varphi_1}{\varphi_s}L^{s-1} + ... + 1\right)y_t = \varepsilon_t$$
$$\Leftrightarrow \phi(L)\tilde{\varphi}(L)y_t = -\frac{\varepsilon_{t-s}}{\varphi_s}.$$

As an example (see Gouriéroux and Jasiak, 2014), a MAR(1,1) model

$$\begin{aligned}
(1-\phi_1 L)(1-\varphi_1 L^{-1})y_t &= \varepsilon_t, \\
(1-\phi_1 L)\left(1 - \frac{1}{\varphi_1}L\right)y_t &= -\frac{\varepsilon_{t-1}}{\varphi_1}, \\
\left(1 - \left(\phi_1 + \frac{1}{\varphi_1}\right)L + \frac{\phi_1}{\varphi_1}L^2\right)y_t &= -\frac{\varepsilon_{t-1}}{\varphi_1},
\end{aligned}$$

is the pseudo-AR(2) causal model. Note that the weak white noise disturbance term $-\frac{\varepsilon_{t-1}}{\varphi_1}$ is not a MA(1) process but only another white noise sequence, shifted by one period in this case.

8.2 Appendix B: Weak White Noise

Brockwell and Davis (1991) usually introduce a noncausal specification as an autoregressive model, for instance an AR(1), with roots inside the unit circle

$$y_t = \phi_1 y_{t-1} + \varepsilon_t,$$

where ε_t is a non-Gaussian $iid(0, \sigma^2)$ noise and $|\phi_1| > 1$. Then y_t admits the following representation (by rearranging terms and increasing all subscripts one time unit):

$$y_t = \phi_1^{-1} y_{t+1} - \phi_1^{-1} \varepsilon_{t+1},$$

or after iterating[11]

$$y_t = -\sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j} = -\phi_1^{-1} \varepsilon_{t+1} - \phi_1^{-2} \varepsilon_{t+2} - \phi_1^{-3} \varepsilon_{t+3} - ....$$

with

$$\begin{aligned}
\gamma_y(h) &= Cov(y_{t+h}, y_t) \\
&= Cov\left(-\sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+h+j}, -\sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j}\right) = \sigma^2 \sum_{j=1}^{\infty} \phi_1^{-j} \phi_1^{-(h+j)} \\
&= \sigma^2 \phi_1^{-h} \sum_{j=1}^{\infty} \phi_1^{-2j} = \sigma^2 \phi_1^{-h} \frac{\phi_1^{-2}}{(1 - \phi_1^{-2})} = \sigma^2 \frac{1}{\phi_1^{|h|} (\phi_1^2 - 1)} \\
&.
\end{aligned}$$

Now let us take $\varepsilon_t^* = y_t - \frac{1}{\phi_1} y_{t-1}$. It is straightforward (solving exercise 3.8 in Brockwell and Davis, 2002) that

$$\begin{aligned}
\gamma_{\varepsilon^*}(h) &= Cov(\varepsilon_{t+h}^*, \varepsilon_t^*) \\
&= Cov\left(y_{t+h} - \phi_1^{-1} y_{t+h-1}, y_t - \phi_1^{-1} y_{t-1}\right) \\
&= \gamma_y(h) - \phi_1^{-1} \gamma_y(h+1) - \phi_1^{-1} \gamma_y(h-1) + \phi_1^{-2} \gamma_y(h).
\end{aligned}$$

[11] Complete iteration would yield $y_t = \phi_1^{-T} y_{t+T} - \sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j}$. However, since $|\phi_1| > 1$, it follows that $|\phi_1^{-1}| < 1$ and thus y_t is stationary. Then $||y_t||^2 = E(y_t^2)$ is constant so that $||y_t - (-\sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j})||^2 = \phi_1^{-2T} ||y_{t+T}|| \to 0$ as $T \to \infty$. Hence, it is reasonable to omit the value $\phi_1^{-T} y_{t+T}$ in our future computations.

That is, $\gamma_{\varepsilon^*}(h) = 0$ for $h \neq 0$ and $\gamma_{\varepsilon^*}(h) = \frac{\sigma^2}{\phi_1^2}$ for $h = 0$, and ε_t^* is a white noise process. Consequently, y_t admits the pseudo-causal representation

$$y_t = \frac{1}{\phi_1} y_{t-1} + \varepsilon_t^*.$$

However, ε_t^* is not a strong white noise in general. Using the following relations,

$$\begin{aligned} y_t &= -\sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j} = -\phi_1^{-1}\varepsilon_{t+1} - \phi_1^{-2}\varepsilon_{t+2} - \phi_1^{-3}\varepsilon_{t+3}..., \\ y_{t-1} &= -\phi_1^{-1}\varepsilon_t - \phi_1^{-2}\varepsilon_{t+1} - \phi_1^{-3}\varepsilon_{t+2}..., \end{aligned}$$

and

$$\begin{aligned} y_t^2 &= (-\sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j})^2 \\ &= (-\phi_1^{-1}\varepsilon_{t+1} - \phi_1^{-2}\varepsilon_{t+2} - \phi_1^{-3}\varepsilon_{t+3}....) \times (-\phi_1^{-1}\varepsilon_{t+1} - \phi_1^{-2}\varepsilon_{t+2} - \phi_1^{-3}\varepsilon_{t+3}....), \\ y_{t-1}^2 &= (-\phi_1^{-1}\varepsilon_t - \phi_1^{-2}\varepsilon_{t+1} - \phi_1^{-3}\varepsilon_{t+2}....) \times (-\phi_1^{-1}\varepsilon_t - \phi_1^{-2}\varepsilon_{t+1} - \phi_1^{-3}\varepsilon_{t+2}....), \end{aligned}$$

we find that

$$\begin{aligned} Cov(\varepsilon_t^*, y_{t-1}^2) &= Cov(y_t - \frac{1}{\phi_1} y_{t-1}, y_{t-1}^2) \\ &= E\left(-\sum_{j=1}^{\infty} \phi_1^{-j} \phi_1^{-2(j+1)} \varepsilon_{t+j}^3 + \frac{1}{\phi_1} \sum_{j=1}^{\infty} \phi_1^{-j} \phi_1^{-2j} \varepsilon_{t+j}^3 \right) \\ &= \mu_3 \left(-\frac{1}{\phi_1^2} \sum_{j=1}^{\infty} \phi_1^{-j} \phi_1^{-2j} + \frac{1}{\phi_1} \sum_{j=1}^{\infty} \phi_1^{-j} \phi_1^{-2j} \right) \\ &= \mu_3 \left(-\frac{1}{\phi_1^2} \frac{\phi_1^{-3}}{(1-\phi_1^{-3})} + \frac{1}{\phi_1} \frac{\phi_1^{-3}}{(1-\phi_1^{-3})} \right) \end{aligned}$$

$$
\begin{aligned}
&= \mu_3 \left(-\frac{1}{\phi_1^2} \frac{1}{(\phi_1^3 - 1)} + \frac{1}{\phi_1} \frac{1}{(\phi_1^3 - 1)} \right) \\
&= \mu_3 \left(-\frac{1}{\phi_1^2} \frac{1}{(\phi_1^3 - 1)} + \frac{1}{\phi_1^2} \frac{\phi_1}{(\phi_1^3 - 1)} \right) \\
&= \mu_3 \left(\frac{\phi_1 - 1}{\phi_1^2(\phi_1^3 - 1)} \right) \\
&= \mu_3 \left(\frac{(\phi_1 - 1)}{\phi_1^2(\phi_1 - 1)(\phi_1^2 + \phi_1 + 1)} \right) \\
&= \frac{\mu_3}{\phi_1^2(\phi_1^2 + \phi_1 + 1)},
\end{aligned}
$$

which is unequal to zero if $E(\varepsilon_t^3) = \mu_3 \neq 0$.

8.3 Appendix C: Matlab Toolbox Main Features

Note to the referees: please tell us whether we should add a description/illustration of our toolbox or whether "available upon request" is enough.