

Alain Hecq and Li Sun*

Selecting between causal and noncausal models with quantile autoregressions

<https://doi.org/10.1515/snde-2019-0044>

Received April 25, 2019; accepted August 4, 2020; published online September 19, 2020

Abstract: We propose a model selection criterion to detect purely causal from purely noncausal models in the framework of quantile autoregressions (QAR). We also present asymptotics for the i.i.d. case with regularly varying distributed innovations in QAR. This new modelling perspective is appealing for investigating the presence of bubbles in economic and financial time series, and is an alternative to approximate maximum likelihood methods. We illustrate our analysis using hyperinflation episodes of Latin American countries.

Keywords: causal and noncausal time series; financial bubbles; model selection criterion; quantile autoregressions; regularly varying variables.

1 Motivation

Mixed causal and noncausal time series models have been recently used in order (i) to obtain a stationary solution to explosive autoregressive processes, (ii) to improve forecast accuracy, (iii) to model expectation mechanisms implied by economic theory, (iv) to interpret non-fundamental shocks resulting from the asymmetric information between economic agents and econometricians, (v) to generate non-linear features from simple linear models with non-Gaussian disturbances, (vi) to test for time reversibility. When the distribution of innovations is known, a non-Gaussian likelihood approach can be used to discriminate between lag and lead polynomials of the dependent variable. For instance, the R package MARX developed by (Hecq, Lieb, and Telg 2017a) estimates univariate mixed models under the assumption of a Student's t -distribution with ν degrees of freedom (see also Lanne and Saikkonen 2011, 2013) as well as the Cauchy distribution as a special case of the Student's t when $\nu = 1$. Gouriéroux and Zakoian (2016) privilege the latter distribution to derive analytical results. Gouriéroux and Zakoian (2015); Fries and Zakoian (2017) provide an additional flexibility to involve some skewness by using the family of alpha-stable distributions. However, all those aforementioned results require the estimation of a parametric distributional form. In this article we take another route.

The objective of this paper is to select between causal and noncausal models without using parametric distributional assumptions. To achieve that, we adopt a quantile regression (QR) framework and apply quantile autoregressions (QCAR hereafter) (Koenker and Xiao 2006) on candidate models. Although we obviously also require non-Gaussian innovations in time series, we do not make any parametric distributional assumption about the innovations. By using quantile regressions, we consider a statistic called the sum of rescaled absolute residuals (SRAR hereafter) to measure model performances and reveal properties of time series. Remarkably we find that SRAR cannot always favour a model uniformly along quantiles. This issue is common for time series of asymmetric distributed innovations, which causes confusion in model selection and calls for a robust statistic to meet the goal. Considering that, we propose to aggregate the SRAR information over quantiles. It is worth mentioning that when coefficients are constant in the underlying model with a

***Corresponding author: Li Sun, Ph.D.,** Maastricht University, School of Business and Economics, Department of Quantitative Economics, P.O.Box 616, 6200 MD Maastricht, The Netherlands, Phone: +31619341416, E-mail: l.sun@maastrichtuniversity.nl. <https://orcid.org/0000-0001-7662-7174>

Alain Hecq, Maastricht University, School of Business and Economics, Department of Quantitative Economics, P.O.Box 616, 6200 MD Maastricht, The Netherlands, E-mail: a.hecq@maastrichtuniversity.nl

symmetrically i.i.d. error term, the aggregate SRAR criterion is equivalent to select between forward and backward conditional mean models (termed by Gourieroux and Zakoian (2017)). However, the aggregate SRAR is a measure based on the whole dynamics of the underlying process, which is not dominated by the conditional mean information any more. This characteristic of the aggregate SRAR criterion indeed makes it robust in model selection even for some general situations such as with asymmetric distributed innovations. Another remark on this paper is that our method is restricted to the model framework of purely causal or noncausal autoregressions without other explanatory variables, thereby this method can be used to questions like asset pricing of exchange rate where current exchange rate is associated with future exchange rates. However, it cannot be used to questions like the Taylor (1993) rule which associates the dynamics of the nominal interest rate with dynamics of some endogenous variables (e.g. inflation).

The rest of this paper is constructed as follows. Section 2 introduces mixed causal and noncausal models and our research background. In Section 3, we propose quantile autoregressions in the time reverse version called quantile noncausal autoregressions (QNCAR) along with a generalized asymptotic theorem in a stable law for both QCAR and QNCAR. Section 4 brings out a common issue in the model selection through SRAR comparison. The use of the aggregate SRAR over all quantiles as a new model selection criterion is then proposed with the shape of SRAR curves being analysed. Furthermore, we illustrate our analysis using hyperinflation episodes of four Latin American countries in Section 6. Section 7 concludes this paper.

2 Causal and noncausal time series models

Brockwell and Davis introduce in their textbooks (Brockwell and Davis 2016; Brockwell, Davis, and Fienberg 1991) a univariate noncausal specification as a way to rewrite an autoregressive process with explosive roots into a process in reverse time with roots outside the unit circle. This noncausal process possesses a stable forward looking solution whereas the explosive autoregressive process in direct time does not. This approach can be generalized to allow for both lead and lag polynomials. This is the so called mixed causal-noncausal univariate autoregressive process for y_t that we denote $MAR(r, s)$

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

where $\pi(L) = 1 - \pi_1 L - \dots - \pi_r L^r$, $\phi(L^{-1}) = 1 - \phi_1 L^{-1} - \dots - \phi_s L^{-s}$. L is the usual backshift operator that creates lags when raised to positive powers and leads when raised to negative powers, i.e. $L^j y_t = y_{t-j}$ and $L^{-j} y_t = y_{t+j}$. The roots of both polynomials are assumed to lie strictly outside the unit circle, that is $\pi(z) = 0$ and $\phi(z) = 0$ for $|z| > 1$ and therefore

$$y_t = \pi(L)^{-1} \phi(L^{-1})^{-1} \varepsilon_t = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{t-i} \quad (2)$$

has an infinite two sided moving average representation. We also have that $E(|\varepsilon_t|^\delta) < \infty$ for $\delta > 0^1$ and the Laurent expansion parameters are such that $\sum_{i=-\infty}^{\infty} |a_i|^\delta < \infty$. The representation (2) is sometimes clearer than (1) to motivate the terminology “causal/noncausal”. Indeed those terms refer to as the fact that y_t depends on a causal (resp. noncausal) component $\sum_{i=0}^{\infty} a_i \varepsilon_{t-i}$ (resp. noncausal $\sum_{i=-\infty}^{-1} a_i \varepsilon_{t-i}$). With this in mind, it is obvious that an autoregressive process with explosive roots will be defined as noncausal.

Note that in (1), the process y_t is a purely causal $MAR(r, 0)$, also known as the conventional causal $AR(r)$ process, when $\phi_1 = \dots = \phi_s = 0$,

$$\pi(L)y_t = \varepsilon_t, \quad (3)$$

while the process is a purely noncausal $MAR(0, s)$

¹ The errors do not necessarily have finite second order moments. For $\delta \geq 2$ the second order moment exists, for $\delta \in [1, 2)$ the errors have infinite variance but finite first order moment, for $\delta \in (0, 1)$ the errors do not have finite order moments.

$$\phi(L^{-1})y_t = \varepsilon_t, \quad (4)$$

when $\pi_1 = \dots = \pi_r = 0$.

A crucial point of this literature is that innovation terms ε_t must be i.i.d. non-Gaussian to ensure the identifiability of a causal from a noncausal specification (Breidt et al. 1991). The departure from Gaussianity is not as such an ineptitude as a large part of macroeconomic and financial time series display nonlinear and non-normal features.

We have already talked in Section 1 about the reasons for looking at models with a lead component. Our main motivation in this paper lies in the fact that MAR(r, s) models with non-Gaussian disturbances are able to replicate non-linear features (e.g. bubbles, asymmetric cycles) that previously were usually obtained by highly nonlinear models. As an example, we simulate in Figure 1 an MAR(1,1) of $(1-0.8L)(1-0.6L^{-1})y_t = \varepsilon_t$ with $\varepsilon_t \stackrel{d}{\sim} t(3)$ for 200 observations.² One can observe asymmetric cycles and multiple bubbles.³

Once a distribution or a group of distributions is chosen, the parameters in $\pi(L)\phi(L^{-1})$ can be estimated. Assuming for instance a non-standardized t -distribution for the innovation process, the parameters of mixed causal-noncausal autoregressive models of the form (1) can be consistently estimated by the approximate maximum likelihood (AML) method (Hecq, Lieb, and Telg 2016). Let $(\varepsilon_1, \dots, \varepsilon_T)$ be a sequence of i.i.d. zero mean t -distributed random variables, then its joint probability density function can be characterized as

$$f_{\varepsilon}(\varepsilon_1, \dots, \varepsilon_T | \sigma, \nu) = \prod_{t=1}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma} \left(1 + \frac{1}{\nu} \left(\frac{\varepsilon_t}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}},$$

where $\Gamma(\cdot)$ denotes the gamma function. The corresponding (approximate) log-likelihood function conditional on the observed data $y = (y_1, \dots, y_T)$ can be formulated as

$$l_y(\phi, \boldsymbol{\varphi}, \boldsymbol{\lambda}, \alpha | y) = (T-p)[\ln(\Gamma((\nu+1)/2)) - \ln(\sqrt{\nu\pi}) - \ln(\Gamma(\nu/2)) - \ln(\sigma)] \\ - (\nu+1)/2 \sum_{t=r+1}^{T-s} \ln\left(1 + ((\pi(L)\phi(L^{-1})y_t - \alpha)/\sigma)^2/\nu\right), \quad (5)$$

where $p = r + s$ and $\varepsilon_t = \pi(L)\phi(L^{-1})y_t - \alpha$ is replaced by a nonlinear function of the parameters when expanding the product of polynomials. The distributional parameters are collected in $\boldsymbol{\lambda} = [\sigma, \nu]'$, with σ representing the scale parameter and ν the degrees of freedom. α denotes an intercept that can be introduced in model (1). Thus, the AML estimator corresponds to the solution $\hat{\boldsymbol{\theta}}_{ML} = \arg \max_{\boldsymbol{\theta} \in \Theta} l_y(\boldsymbol{\theta} | y)$, with $\boldsymbol{\theta} = [\phi', \boldsymbol{\varphi}', \boldsymbol{\lambda}']'$ and Θ is a permissible parameter space containing the true value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$, as an interior point. Since an analytical solution of the score function is not directly available, gradient based numerical procedures can be used to find $\hat{\boldsymbol{\theta}}_{ML}$. If $\nu > 2$, and hence $E(|\varepsilon_t|^2) < \infty$, the AML estimator is \sqrt{T} -consistent and asymptotically normal. Lanne and Saikonen (2011)

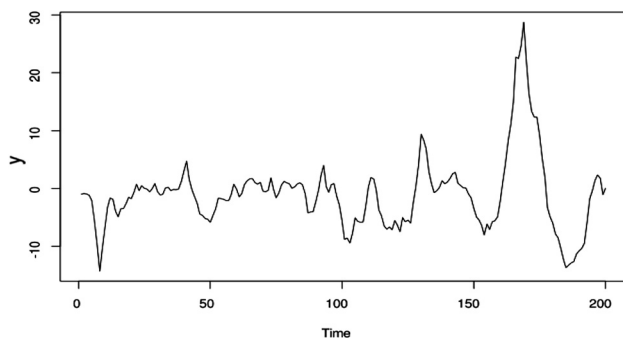


Figure 1: Simulation of a MAR(1,1) model, $T = 200$.

² We use the package MARX developed in R by Hecq, Lieb, and Telg (2017a, 2017b).

³ MAR(r, s) models can be generated in two steps (see Gouriéroux and Jasiak 2016; Hecq, Lieb, and Telg 2016). We propose in the Appendix an alternative method based on matrix representation that is very compact in code writing and intuitive in understanding.

also show that a consistent estimator of the limiting covariance matrix is obtained from the standardized Hessian of the log-likelihood. For the estimation of the parameters and the standard innovations as well as for the selection of mixed causal-noncausal models we can also follow the procedure proposed by Hecq, Lieb, and Telg (2016).

However, the AML estimation is based on a parametric form of the innovation term in (1), which makes this method not flexible enough to adapt uncommon distributions as complex in reality. To be more practical and get rid of strong distribution assumptions on innovations, in next section we adopt quantile regression methods with some properties discussed there. This paper only focuses on purely causal and noncausal models.

3 QCAR & QNCAR

Koenker and Xiao (2006) have introduced a quantile autoregressive model of order p denoted as QAR(p) which is formulated as the following form:

$$y_t = \theta_0(u_t) + \theta_1(u_t)y_{t-1} + \dots + \theta_p(u_t)y_{t-p}, \quad t = p+1, \dots, T, \quad (6)$$

where u_t is a sequence of i.i.d. standard uniform random variables. In order to emphasize the causal characteristic of this kind of autoregressive models, we refer to (6) as QCAR(p) hereafter. Provided that the right-hand side of (6) is monotone increasing in u_t , the τ -th conditional quantile function of y_t can be written as

$$Q_{y_t}(\tau|y_{t-1}, \dots, y_{t-p}) = \theta_0(\tau) + \theta_1(\tau)y_{t-1} + \dots + \theta_p(\tau)y_{t-p}. \quad (7)$$

If an observed time series $\{y_t\}_{t=1}^T$ can be written into a QCAR(p) process, its parameters as in (7) can be obtained from the following minimization problem.

$$\hat{\theta}(\tau) = \arg \min_{\theta \in \mathbb{R}^{p+1}} \sum_{t=1}^T \rho_{\tau}(y_t - \mathbf{x}_t' \theta), \quad (8)$$

where $\rho_{\tau}(u) := u(\tau - I(u < 0))$ is called the check function, $\mathbf{x}_t' := [1, y_{t-1}, \dots, y_{t-p}]$, and $\theta' := [\theta_0, \theta_1, \dots, \theta_{t-p}]$. We define the sum of rescaled absolute residuals (SRAR) for each pair of (τ, θ) as

$$\text{SRAR}(\tau, \theta) := \sum_{t=1}^T \rho_{\tau}(y_t - \mathbf{x}_t' \theta). \quad (9)$$

Substitute (9) into (8) and write the minimization problem (8) as

$$\hat{\theta}(\tau) = \arg \min_{\theta \in \mathbb{R}^{p+1}} \text{SRAR}(\tau, \theta). \quad (10)$$

The estimation consistency and asymptotic normality in the minimization problem (8) have been provided by Koenker and Xiao (2006). A modified simplex algorithm proposed by Barrodale and Roberts (1973) can be used to solve the minimization, and in practise parameters for each τ -th quantile can be obtained, for instance, through the `rq()` function from the **quantreg** package in R or in EViews.

3.1 QNCAR

A QNCAR(p) specification is introduced here as the noncausal counterpart of the QCAR(p) model by reversing time, explicitly as follows:

$$Q_{y_t}(\tau|y_{t+1}, \dots, y_{t+p}) = \phi_0(\tau) + \phi_1(\tau)y_{t+1} + \dots + \phi_p(\tau)y_{t+p}. \quad (11)$$

Analogously to the QCAR(p), the estimation of the QNCAR(p) goes through solving

$$\hat{\boldsymbol{\theta}}(\tau) = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \text{SRAR}(\tau, \boldsymbol{\theta})$$

with

$$\mathbf{x}_t' = [1, y_{t+1}, \dots, y_{t+p}],$$

where for the simplicity of notations, we use $\hat{\boldsymbol{\theta}}(\tau)$ to denote the estimate in quantile noncausal autoregression. Drawing on the asymptotic derived by Koenker and Xiao (2006), we present the following theorem for QNCAR(p) based on three assumptions which are made to ensure covariance stationarity of the time series (by (A1) and (A2)) and the existence of quantile estimates (by (A3)).

Remark. There is an issue in the estimation consistency of QCAR(p) as reported by Fan and Fan (2010). This is due to the violation on the monotonicity requirement of the right side of (6) in u_t but not exclusively the monotonicity of $\theta_i(u_t)$ in u_t . So to recover an AR(p) process of coefficients $\theta_i(u_t)$ ($i = 0, \dots, p$) monotonic in u_t , quantile autoregression is not a 100% match tool unless the monotonicity requirement is met beforehand. This issue is also illustrated in Section 4.1.

Theorem 1. A QNCAR(p) model can be written in the following vectorized companion form:

$$\tilde{\mathbf{x}}_t = \mathbf{A}_t \tilde{\mathbf{x}}_{t+1} + \mathbf{v}_t, \quad (12)$$

where $\tilde{\mathbf{x}}_t' := [y_t, y_{t+1}, \dots, y_{t+p-1}]$, $\mathbf{x}_t' := [1, \tilde{\mathbf{x}}_t']$, $\mathbf{A}_t := \begin{bmatrix} \phi_{1,t} & \phi_{2,t} & \dots & \phi_{p,t} \\ \mathbf{I}_{p-1} & & & \mathbf{0}_{(p-1) \times 1} \end{bmatrix}$ and $\mathbf{v}_t := \begin{bmatrix} \varepsilon_t \\ \mathbf{0}_{(p-1) \times 1} \end{bmatrix}$, satisfying the following assumptions:

- (A1): $\{\varepsilon_t\}_{t=1}^n$ are i.i.d. innovations with mean 0 and variance $\sigma^2 < \infty$. The distribution function of ε_t , denoted as $F(\cdot)$, has a continuous density $f(\cdot)$ with $f(\varepsilon) > 0$ on $\mathcal{H} := \{\varepsilon : 0 < F(\varepsilon) < 1\}$.
 (A2): The eigenvalues of $E[\mathbf{A}_t \otimes \mathbf{A}_t]$ have moduli less than one.
 (A3): $F_{y_t|\tilde{\mathbf{x}}_{t+1}}(\cdot) := \mathbf{P}[y_t < \cdot | y_{t+1}, y_{t+2}, \dots, y_{t+p}]$ has derivative $f_{y_t|\tilde{\mathbf{x}}_{t+1}}(\cdot)$ which is uniformly integrable on \mathcal{H} and non-zero with probability one.

Then,

$$\Sigma^{-\frac{1}{2}} \sqrt{T} (\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\phi}(\tau)) \stackrel{d}{\rightarrow} \mathbb{B}_{p+1}(\tau), \quad (13)$$

where $\boldsymbol{\Sigma} := \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_1^{-1}$, $\boldsymbol{\Sigma}_0 := E[\mathbf{x}_t \mathbf{x}_t']$, $\boldsymbol{\Sigma}_1 := \lim T^{-1} \sum_{t=1}^T f_{y_t|\tilde{\mathbf{x}}_{t+1}}(F_{y_t|\tilde{\mathbf{x}}_{t+1}}^{-1}(\tau)) \mathbf{x}_t \mathbf{x}_t'$, $\boldsymbol{\phi}(\tau)' := [F^{-1}(\tau), \phi_1(\tau), \dots, \phi_p(\tau)]$, $\mathbb{B}_{p+1}(\tau) := \mathcal{N}(0, \tau(1-\tau)\mathbf{I}_{p+1})$ with sample size T .

The above result can be further simplified into Corollary 2 by adding the following assumption:

- (A4): The coefficient matrix \mathbf{A}^t in (12) is constant over time. (We denote $\mathbf{A} := \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ \mathbf{I}_{p-1} & & & \mathbf{0}_{(p-1) \times 1} \end{bmatrix}$ for \mathbf{A}^t under this assumption.)

Corollary 2. Under assumptions (A1), (A2), (A3) and (A4),

$$\sqrt{T} f(F^{-1}(\tau)) \Sigma_0^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\phi}_\tau) \stackrel{d}{\rightarrow} \mathbb{B}_{p+1}(\tau), \quad (14)$$

where $\boldsymbol{\phi}_\tau := [F^{-1}(\tau), \phi_1, \dots, \phi_p]$.

As can be seen, QCAR(p) and QNCAR(p) generalize the classical purely causal and purely noncausal models respectively by allowing random coefficients on lag or lead regressors over time. Corollary 2 provides additional results when the same coefficients except the intercept are used to generate each quantile. However,

the moment requirement in (A1) is very strict for heavy tailed time series. In order to study noncausality by QAR in heavy tailed distributions, we have to show its applicability when weakening the assumption (A1). This goal is achieved by Theorem 3 which presents the asymptotic behaviour of the QAR estimator for a classical purely noncausal model. Similarly, the asymptotic in a classical purely causal model follows right after reversing time.

Theorem 3. (Asymptotics in regularly varying distributed innovations).

Under Assumption (A4), a purely noncausal AR(p) of the following form

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

where $\phi(L^{-1}) = 1 - \phi_1 L^{-1} - \dots - \phi_p L^{-p}$, also satisfies the following assumptions:

(A5): $\{\varepsilon_t\}_{t=1}^n$ are i.i.d. innovation variables with regularly vary tails defined as

$$P(|\varepsilon_t| > x) = x^{-\alpha} L(x), \quad (15)$$

where $L(x)$ is slowly varying at ∞ and $0 < \alpha < 2$. There is a sequence $\{a_T\}$ satisfying

$$T \cdot P\{\|\varepsilon_t\| > a_T x\} \rightarrow x^{-\alpha} \quad \text{for all } x > 0. \quad (16)$$

with $b_T = \mathbb{E}[\varepsilon_t I(|\varepsilon_t| \leq a_T)] = 0$.⁴ The distribution function of ε_t , denoted as $F(\cdot)$, has continuous density $f(\cdot)$ with $f(\varepsilon) > 0$ on $\{\varepsilon: 0 < F(\varepsilon) < 1\}$ in probability one;

(A6): The roots of the polynomial $\phi(z)$ are greater than one, such that y_t can be written into

$$y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t+j}, \quad (17)$$

where $\sum_{j=0}^{\infty} j |c_j|^\delta < \infty$ for some $\delta < \alpha$, $\delta \leq 1$.

Then

$$\frac{f(F^{-1}(\tau)) \cdot a_T \sqrt{T}}{\sqrt{\tau(1-\tau)}} (\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\phi}_\tau) \stackrel{d}{\rightarrow} \begin{bmatrix} 1 & 0 \\ 0 & \left(\int_0^1 S_\alpha^2(s) ds \right)^{-1} \end{bmatrix} \begin{bmatrix} W(1), \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) dW(s), \dots, \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) dW(s) \end{bmatrix}_{(p+1) \times 1}. \quad (18)$$

where $\boldsymbol{\phi}_\tau := \left[\frac{F^{-1}(\tau)}{a_T}, \phi_1, \dots, \phi_p \right]$, $\boldsymbol{\Omega} = [\omega_{ik}]$ being a $p \times p$ matrix with entry $\omega_{ik} := \sum_{j=0}^{\infty} c_j c_{j+|k-i|}$ at the i th row and the k th column, $\{S_\alpha(s)\}$ being a process of stable distributions with index α which are independent of Brownian motions $\{W(s)\}$. In this theorem the intercept regressor in QNCAR(p) is changed to a_T such that $\mathbf{x}'t = [a_T, y_t, y_{t+1}, \dots, y_{t+p-1}]$.

Proof. See the appendix

Heuristically, next we restrict our focus on the classical models and explore consequences of causality misspecification in quantile regressions.

3.2 Causal and noncausal models with Gaussian i.i.d. disturbances

Suppose a causal AR(1) process $\{y_t\}_{t=1}^T$, $y_t = \alpha + \beta y_{t-1} + \varepsilon_t$, with for instance $[\alpha, \beta] = [1, 0.5]$, i.i.d. standard normal $\{\varepsilon_t\}$ and $T = 200$. Figure 2 displays a simulated sample following this data generating process (DGP).

⁴ Without loss of generality, we assume b_T to be zero in the derivation for the simplicity.

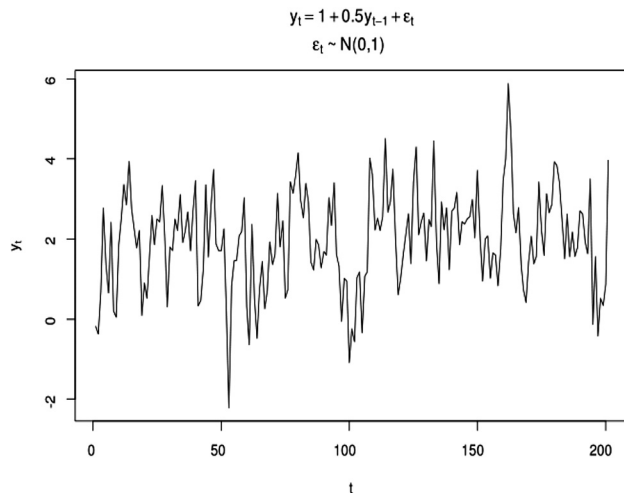


Figure 2: Simulation of a one-regime process with $N(0, 1)$ innovations, $T = 200$.

The information displayed in Figure 3 is the SRAR(τ) of each candidate model along quantiles, indicating their goodness of fit. The two SRAR curves almost overlap at every quantile, which implies no discrimination between QCAR and QNCAR in Gaussian innovations, in line with results in the OLS case. The Gaussian distribution is indeed time reversible, weak and strict stationary. Its first two moments characterize the whole distribution and consequently every quantile. Note that we obtain similar results for a stationary noncausal AR(p) process with i.i.d. Gaussian $\{\varepsilon_t\}$. The results are not reported to save space.

3.3 Causal and noncausal models with Student's t distributed innovations

Things become different if we depart from Gaussianity. Suppose now a causal AR(1) process $y_t = \alpha + \beta y_{t-1} + \varepsilon_t$ with again $[\alpha, \beta] = [1, 0.5]$ but where $\{\varepsilon_t\}$ are i.i.d. Student's t -distributed with 2 degrees of freedom (hereafter using shorthand notation: $t(2)$). Figure 4 depicts a simulation in this AR(1) with $T = 200$. Applying QCAR and QNCAR respectively on this series results in the SRAR curves displayed in Figure 5. The distance between the two curves is obvious compared to the Gaussian case, favouring the causal specification at almost all quantiles. Figure 6 is the SRAR plot of a purely noncausal process with i.i.d. Cauchy innovations. The noncausal specification is preferred in the SRAR comparison.

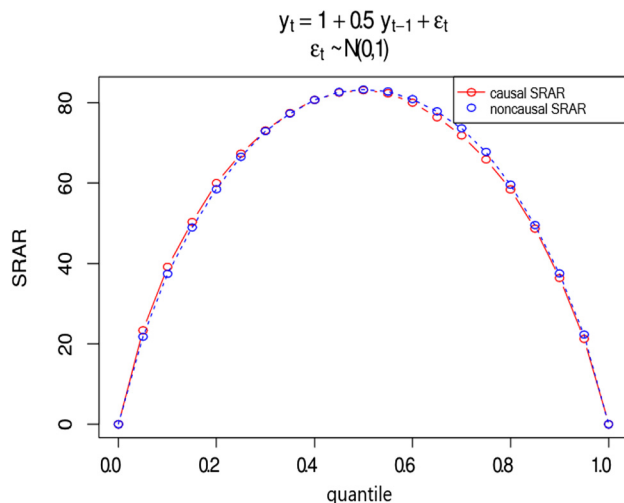


Figure 3: SRAR plot under an AR(1) with $N(0, 1)$ innovations, $T = 200$.

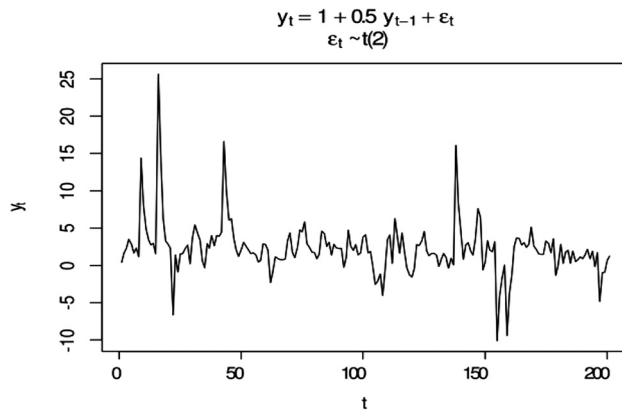


Figure 4: Simulation of an AR(1) with $t(2)$ innovations, $T = 200$.

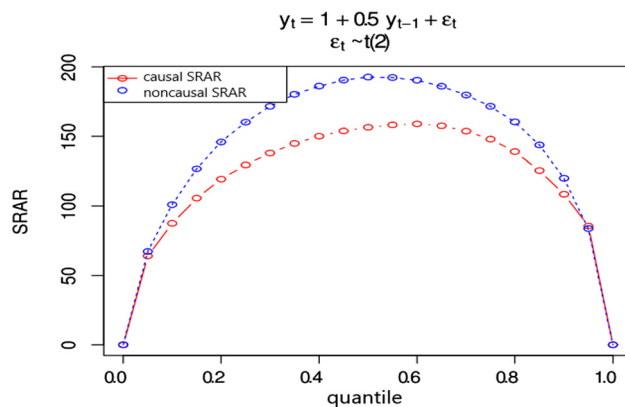


Figure 5: SRAR plot under an AR(1) with $t(2)$, $T = 200$.

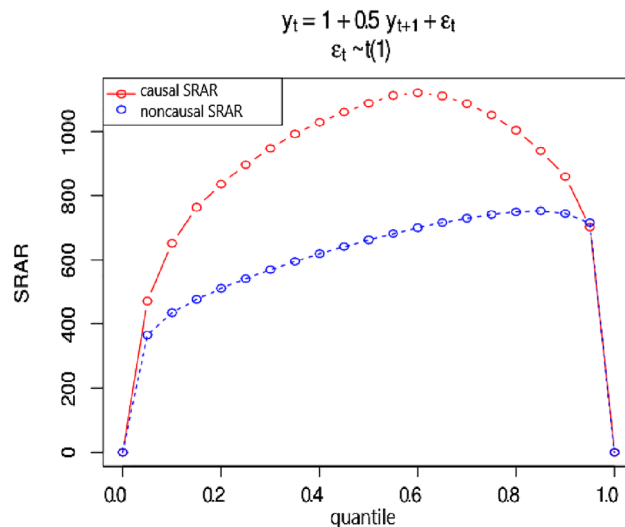


Figure 6: SRAR plot under a noncausal model with Cauchy innovations, $T = 200$.

It seems now that applying the SRAR comparison at one quantile, such as the median, is sufficient for model identification, but it is not true in general. In Section 4, we will spot an identification issue in the SRAR plots, the true model even having higher SRAR values at certain quantiles than the misspecified model.

So far we have applied QCAR and its counterpart QNCAR on the classical purely causal or noncausal models with symmetrically i.i.d. innovations. Within this restricted scope, the conditional mean models of those data generating processes only differ from their conditional quantile models in the intercept term. And we see that model selection by the SRAR comparison gives uniform decisions along quantiles. However, such a model selection is not always that clear in practise. For example, in the empirical study later, we will encounter an identification issue which can be seen by checking the SRAR plots. In the next section, we will present this

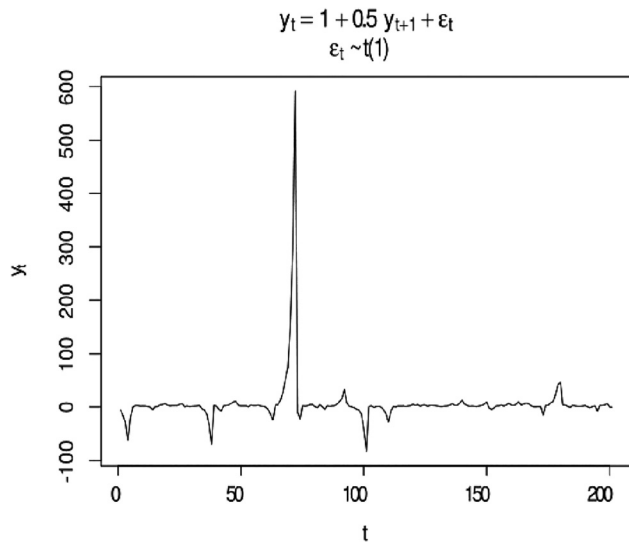


Figure 7: Simulation of a noncausal model with Cauchy innovations, $T = 200$.

identification issue with some possible reasons, and propose a robust model selection criterion called the aggregate SRAR to cope with this issue.

4 SRAR as a model selection criterion

It is natural to think about SRAR as a model selection criterion since a lower SRAR means a better goodness of fit in quantile regressions. However, SRAR is a function of quantile, which raises a question on which quantile to be considered for model selection. It is empirically common to see an identification problem by checking the SRAR plots, which gives different model selections at certain quantiles and makes a selection unreliable if only one quantile is considered. In this section, we discuss this issue and propose a more robust model selection criterion based on aggregating SRARs.

4.1 Identification issue spotted from the SRAR plots

First let us see some possible model settings causing the identification problem in SRAR plots. The first case is linked to the existence of multi-regimes in coefficients.

Suppose a regime-shift model is specified as follows:

$$y_t = \beta_t y_{t+1} + \varepsilon_t, \quad (19)$$

where $\{\varepsilon_t\}$ is an i.i.d. innovation process with cumulative probability function $F(\cdot)$, and β_t is defined as follows:

$$\beta_t = \begin{cases} \beta_1, & \text{if } 0 < F(\varepsilon_t) \leq \tau^*; \\ \beta_2, & \text{if } \tau^* < F(\varepsilon_t) \leq 1, \end{cases} \quad (20)$$

with $\tau^* \in (0, 1)$ and $\beta_1 < \beta_2$. In essence, the regime shift of β_t depends on the quantile occurrence of ε_t which is indexed by $\tau_t = F(\varepsilon_t)$ with $\{\tau_t\}$ being i.i.d. in the standard uniform distribution.

If $\{y_t\}$ can be negative, then there is a problem in using QNCAR to recover the coefficients in the underlying model (19) because the τ -th regime is not necessary to produce the τ -th conditional quantile of y_t . So the comonotonicity condition of the linear quantile regressions (19) is not satisfied. However, by restricting to the non-negative region of the covariate y_{t+1} (also see Fan and Fan 2010) we force the regression model to satisfy the comonotonicity requirement without losing its association with $\{\tau_t\}$. The obtained estimator is also

consistent to the true coefficients in (19).⁵ QCAR (or QNCAR) with such a restriction, hereafter denoted as RQCAR (or RQNCAR) shorthand for restricted quantile causal autoregression (or restricted quantile noncausal autoregression), is formulated as follows:

$$\hat{\boldsymbol{\theta}}(\tau) = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \sum_{t=1}^T I[t \in \mathfrak{S}] \rho_{\tau}(y_t - \mathbf{x}'_t \boldsymbol{\theta}) \quad (21)$$

where \mathfrak{S} is the set restricting the quantile regression on a particular information set. The restriction is usually imposed in order for quantile regressions to meet the comonotonicity condition. To study the regime-shift model (19), we restrict the QNCAR on non-negative covariates, i.e. $\mathfrak{S} = \{t : \mathbf{x}_t \geq 0\}$.

Figure 8 shows four SRAR curves estimated from QCAR, QNCAR, RQCAR and RQNCAR. We consider a time series $\{y_t\}_{t=1}^{600}$ simulated from the model (19) with $\tau^* = 0.7$, $\beta_1 = 0.2$, $\beta_2 = 0.8$ and i.i.d. innovation process in $t(3)$, i.e. $F^{-1}(\cdot) = F_{t(3)}^{-1}(\cdot)$.

Figure 8 illustrates such an identification issue in which the SRAR curve from a true model is not always lower than one from misspecification. Applying restriction helps to enlarge the SRAR difference between a true model and a misspecified time direction.

The second case we investigate is the presence of skewed distributed innovations.

Let us consider a time series $\{y_t\}$ following a purely noncausal AR(1): $y_t = 0.8y_{t+1} + \varepsilon_t$ with $\{\varepsilon_t\}$ i.i.d. in a demeaned skewed t distribution with skewing parameter $\gamma = 2$ and $\nu = 3$ degrees of freedom (hereafter $t(\nu, \gamma)$ is the shorthand notation for a skewed t -distribution). The probability density function of $t(\nu, \gamma)$ (see Francq and Zakoïan 2007) is defined as

$$f(x) = \begin{cases} \frac{2}{\gamma + \frac{1}{\gamma}} f_t(\gamma x) & \text{for } x < 0 \\ \frac{2}{\gamma + \frac{1}{\gamma}} f_t\left(\frac{x}{\gamma}\right) & \text{for } x \geq 0 \end{cases} \quad (22)$$

where $f_t(\cdot)$ is the probability density function of the symmetric $t(\nu)$ distribution. Figure 9 shows four SRAR curves obtained from the estimation of the QCAR, the QNCAR, the RQCAR and the RQNCAR respectively. The

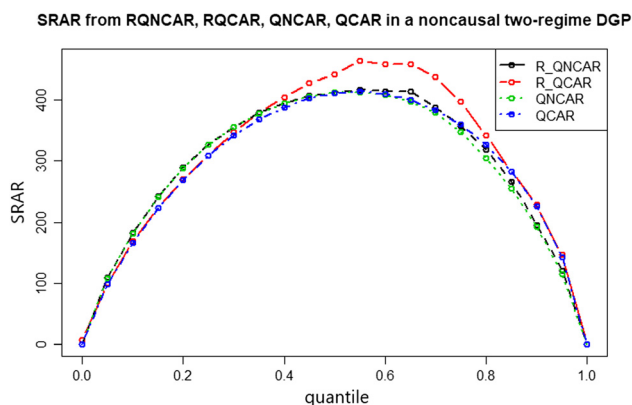


Figure 8: Identification problem spotted in the SRAR plot for restricted quantile autoregressions.

⁵ An alternative is to use simultaneous linear quantile regressions (see Tokdar and Kadane 2012; Liu 2019) which assumes a base quantile function for all coefficients in a (reparametrized) regression model. The dependence structure of these coefficients and priors on parameters in the base quantile function are further assumed for simulation, model fitting and posterior distribution summary. We do not include this method here as this paper aims to avoid assumptions on parametrizing underlying conditional distributions in the model setting. Additionally, regarding the data in our empirical study, very few negative points are observed. So the concern on losing data by such a restriction can be addressed.

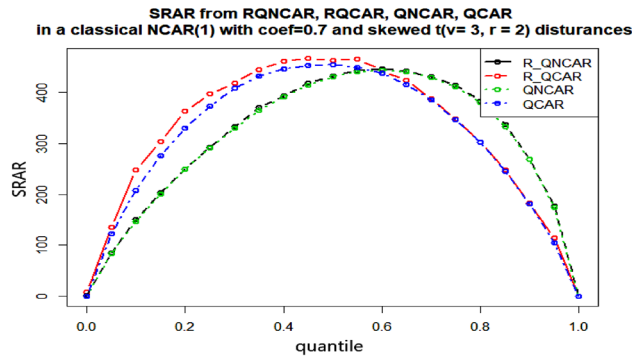


Figure 9: Identification issue spotted from the SRAR plot for a skewed distribution.

curves from the QNCR and the RQNCR almost overlap each other, which confirms our understanding that the monotonicity requirement is met in the true model. The estimations and the corresponding SRAR curves should be the same unless many observations are omitted by the restriction. On the other hand, the SRAR curve gets enlarged from the QCAR to the RQCAR, which is very reasonable as the feasible set in the QCAR is larger and the misspecification is not ensured to satisfy the monotonicity requirement. Again we see this identification problem from the SRAR plot. Remarkably, the SRAR curve from a true model can be higher at certain quantiles than the one from a misspecified model. Consequently the SRAR comparison relying only on particular quantiles, such as the least absolute deviation (LAD) method for the median only is not robust in general. Therefore, we propose a new model selection criterion in next subsection by including the information over all quantiles.

Table 1: Frequencies of selecting the correct model using the SRAR criteria.

Quantiles	Gaussian (Figure 2)	$t(2)$ (Figure 4)	$t(1)$ (Figure 7)	Two-regime (Figure 8)	$t(v=3, \gamma=2)$ (Figure 9)
0	0.698	0.678	0.601	0.787	0.476
0.05	0.516	0.416	0.653	0.044	1.000
0.10	0.51	0.677	0.763	0.059	1.000
0.15	0.519	0.858	0.841	0.095	1.000
0.20	0.512	0.948	0.907	0.167	1.000
0.25	0.513	0.981	0.947	0.305	1.000
0.30	0.488	0.992	0.978	0.487	1.000
0.35	0.487	0.998	0.996	0.654	1.000
0.40	0.486	0.999	0.996	0.798	1.000
0.45	0.487	1.000	0.996	0.892	1.000
0.50	0.5	1.000	0.995	0.950	1.000
0.55	0.499	0.999	0.995	0.974	0.994
0.60	0.492	0.999	0.995	0.988	0.533
0.65	0.478	0.997	0.995	0.991	0.018
0.70	0.467	0.994	0.979	0.996	0.001
0.75	0.49	0.984	0.951	0.998	0.000
0.80	0.493	0.954	0.903	0.999	0.000
0.85	0.481	0.862	0.858	1.000	0.000
0.90	0.469	0.72	0.791	1.000	0.000
0.95	0.484	0.454	0.668	0.997	0.000
1	0.653	0.58	0.595	0.780	0.420
Aggregate SRAR	0.483	0.998	0.995	0.995	0.999

4.2 The aggregate SRAR criterion

Based on the same number of explanatory variables in QCAR and QNCAR with a fixed sample size in quantile regressions, the best model is supposed to exhibit the highest goodness of fit among candidate models. Similarly to the R-squared criterion in the OLS, when turning to quantile regressions, we are led to use the SRAR criterion for model selection. The aggregate SRAR is regarded as an overall performance of a candidate model over all quantiles such as:

$$\text{aggregate SRAR} := \int_0^1 \text{SRAR}(\tau) d\tau.$$

There are many ways to calculate this integral. One way is to approximate the integral by the trapezoidal rule. Another way is to sum up SRARs over a fine enough quantile grid with equal weights. In other words, this aggregation is regarded as an average of performances ($\text{SRAR}(\tau)$, $\tau \in (0, 1)$) of a candidate model. In practise, there is almost no difference in model selection between the two aggregation methods.

As equal weights are used on all quantiles in the aggregate SRAR above, people may argue to use a different weighting scheme. The weighting scheme indeed can be different as when weight being one for one quantile and zero for others is to select a conditional quantile model. We agree that the weighting scheme can be customized in justice of users' purpose. The equal-weight scheme proposed here is inspired to calculate the area under the SRAR curve over quantiles when we check the SRAR plot. Intuitively, such areas are directly linked to model performance when we concerns the whole dynamics of the underlying process. And we compare models by viewing the gap between their SRAR curves, which is the difference between the areas under their SRAR curves. This leads us to the aggregate SRAR measure.

Performances of the SRAR model selection criteria in Monte Carlo simulations are reported in Table 1. It shows the average frequencies with which we find the correct model based on the SRAR criterion per quantile and the aggregate SRAR criterion. The sample size T is 200 and each reported number is based on 2000 Monte Carlo simulations. Columns of Table 1 refer to as a particular distribution previously illustrated in this paper. As observed, the aggregate SRAR criterion performs very well even in situations with the identification issue. The Gaussian distribution being weakly and strictly stationary we cannot obviously discriminate between causal and noncausal specifications leading to an average frequency of around 50% to detect the correct model.

4.3 Shape of SRAR curves

By observing SRAR plots, we see that SRAR curves vary when the underling distribution varies. It is interesting to investigate the reasons. In this subsection, we will provide some insights on the slope and concavity of $\text{SRAR}_{y_t}(\tau, \hat{\theta}(\tau))$ curves under assumptions (A1), (A2), (A3) and (A4). Since $\rho_\tau(y_t - \mathbf{x}_t' \theta)$ is a continuous function in $\theta \in \mathbb{R}^{(p+1)}$, by the continuous mapping theorem and $\tau, \hat{\theta}(\tau)$, we know that

$$\rho_\tau(y_t - \mathbf{x}_t' \hat{\theta}) \xrightarrow{p} \rho_\tau(y_t - \mathbf{x}_t' \phi_\tau).$$

We also know that

$$\rho_\tau(y_t - \mathbf{x}_t' \phi_\tau) = \rho_\tau(\varepsilon_t - F^{-1}(\tau)).$$

Therefore instead of directly deriving the shape of a $\text{SRAR}_{y_t}(\tau, \hat{\theta}(\tau))$ curve, we look at the properties of its intrinsic curve $\text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))$. We derive the first and second order derivatives of $\text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))$ with respect to τ in order to determine the shape of $\text{SRAR}_{y_t}(\tau, \hat{\theta}(\tau))$.

4.3.1 The slope property

One major difference between SRAR curves in a plot is their slopes. We can compute the first-order derivative of SRAR with respect to τ if the derivative exists. Under the following assumption:

(A7) : The inverse distribution function $F^{-1}(\cdot)$ of innovation ε_t is continuous and differentiable on $(0, 1)$ to the second order;

we can then take the first-order derivative of $\text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))$ with respect to τ .

Suppose $0 < \tau < \tau + \Delta\tau < 1$, $\Delta\tau > 0$ and denote $\Delta F^{-1}(\tau) = F^{-1}(\tau + \Delta\tau) - F^{-1}(\tau)$.

$$\begin{aligned} \text{SRAR}_{\varepsilon_t}(\tau + \Delta\tau, F^{-1}(\tau + \Delta\tau)) - \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau)) &= \sum_{t=1}^T (\rho_{\tau+\Delta\tau}(\varepsilon_t - F^{-1}(\tau + \Delta\tau)) - \rho_{\tau}(\varepsilon_t - F^{-1}(\tau))) \\ &= \sum_{t=1}^T ((\varepsilon_t - F^{-1}(\tau + \Delta\tau))(\tau + \Delta\tau - \mathbb{1}_{\{\varepsilon_t - F^{-1}(\tau + \Delta\tau) \leq 0\}}) \\ &\quad - (\varepsilon_t - F^{-1}(\tau))(\tau - \mathbb{1}_{\{\varepsilon_t - F^{-1}(\tau) \leq 0\}})) \\ &= \sum_{t=1}^T \left(\varepsilon_t (\Delta\tau - \mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}}) + \tau (F^{-1}(\tau) - F^{-1}(\tau + \Delta\tau)) - \Delta\tau F^{-1}(\tau + \Delta\tau) \right. \\ &\quad \left. + F^{-1}(\tau + \Delta\tau) \mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}} - F^{-1}(\tau) \mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau)\}} \right) \\ &= \sum_{t=1}^T (\Delta\tau (\varepsilon_t - F^{-1}(\tau + \Delta\tau)) + (F^{-1}(\tau + \Delta\tau) \\ &\quad - F^{-1}(\tau)) (\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}} - \tau) + \mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}} \\ &\quad \times (F^{-1}(\tau) - \varepsilon_t)) \end{aligned} \quad (23)$$

Divide the above difference by $\Delta\tau$, and take the limit $\Delta\tau \downarrow 0$. It gives us

$$\lim_{\Delta\tau \downarrow 0} \frac{\text{SRAR}_{\varepsilon_t}(\tau + \Delta\tau, F^{-1}(\tau + \Delta\tau)) - \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{\Delta\tau} = \sum_{t=1}^T \left(\varepsilon_t - F^{-1}(\tau) + \frac{dF^{-1}(\tau)}{d\tau} (\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau)\}} - \tau) \right), \quad (24)$$

because

$$\begin{aligned} \lim_{\Delta\tau \downarrow 0} \frac{\Delta\tau (\varepsilon_t - F^{-1}(\tau + \Delta\tau))}{\Delta\tau} &= \varepsilon_t - F^{-1}(\tau), \\ \lim_{\Delta\tau \downarrow 0} \frac{(F^{-1}(\tau + \Delta\tau) - F^{-1}(\tau)) (\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}} - \tau)}{\Delta\tau} &= \frac{dF^{-1}(\tau)}{d\tau} (\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau)\}} - \tau), \\ \lim_{\Delta\tau \downarrow 0} \frac{\mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}} (F^{-1}(\tau) - \varepsilon_t)}{\Delta\tau} &= 0. \end{aligned} \quad (25)$$

The last line is from

$$\begin{cases} \mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}} (F^{-1}(\tau) - \varepsilon_t) = 0, & \text{when } \varepsilon_t \notin (F^{-1}(\tau), F^{-1}(\tau + \Delta\tau)]; \\ (F^{-1}(\tau) - F^{-1}(\tau + \Delta\tau)) \leq (F^{-1}(\tau) - \varepsilon_t) < 0, & \text{when } \varepsilon_t \in (F^{-1}(\tau), F^{-1}(\tau + \Delta\tau)]; \end{cases} \quad (26)$$

and

$$0 = \mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau)\}} \frac{dF^{-1}(\tau)}{d\tau} \leq \lim_{\Delta\tau \downarrow 0} \frac{\mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}} (F^{-1}(\tau) - \varepsilon_t)}{\Delta\tau} \leq 0. \quad (27)$$

In analogue, the left-hand limit $\lim_{\Delta\tau \downarrow 0} \frac{\text{SRAR}_{\varepsilon_t}(\tau + \Delta\tau, F^{-1}(\tau + \Delta\tau)) - \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{\Delta\tau}$ gives the same result. Therefore, we have the first-order derivative as below.

$$\frac{d \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{d\tau} = \sum_{t=1}^T \left(\varepsilon_t - F^{-1}(\tau) + \frac{dF^{-1}(\tau)}{d\tau} (\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau)\}} - \tau) \right). \quad (28)$$

To manifest this result with a fixed T , we take expectation and obtain

$$\mathbb{E} \left[\frac{d \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{d\tau} \right] = T (\mathbb{E}[\varepsilon_t] - F^{-1}(\tau)), \quad (29)$$

when $\mathbb{E}[\varepsilon_t]$ exists. In practise, we are not strict with $\mathbb{E}[\varepsilon_t] < \infty$ since the mean of an i.i.d. $\{\varepsilon_t\}_{t=1}^T$ can be estimated empirically to replace $\mathbb{E}[\varepsilon_t]$ in (29) without affecting other terms.

Now we have the expectation of $\frac{d \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{d\tau}$ which can be regarded as the underlying guideline for the slope of a SRAR curve. Before interpreting this result, let us derive the second-order derivative of $\text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))$ with respect to τ and make an interpretation together.

4.3.2 The concave property

One empirically observed property of SRAR curves is their concavity which can be explained through the second-order derivative of $\text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))$ with respect to τ under assumptions (A1), (A2), (A3), (A4) and (A7). Suppose $0 < \tau < \tau + \Delta\tau < 1, \Delta\tau > 0$.

$$\begin{aligned} \Delta^2 \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau)) &:= \text{SRAR}_{\varepsilon_t}(\tau + \Delta\tau, F^{-1}(\tau + \Delta\tau)) - 2 \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau)) + \text{SRAR}_{\varepsilon_t}(\tau - \Delta\tau, F^{-1}(\tau - \Delta\tau)) \\ &= \sum_{t=1}^T (\varepsilon_t - F^{-1}(\tau)) (\mathbb{1}_{\{F^{-1}(\tau - \Delta\tau) < \varepsilon_t \leq F^{-1}(\tau)\}} - \mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}}) + \tau (2F^{-1}(\tau) - F^{-1}(\tau + \Delta\tau) - F^{-1}(\tau - \Delta\tau)) \\ &\quad + \Delta\tau (F^{-1}(\tau - \Delta\tau) - F^{-1}(\tau + \Delta\tau)) + (F^{-1}(\tau + \Delta\tau) + F^{-1}(\tau - \Delta\tau) - 2F^{-1}(\tau)) \mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau - \Delta\tau)\}} + (F^{-1}(\tau + \Delta\tau) \\ &\quad - F^{-1}(\tau)) \mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}} \end{aligned} \quad (30)$$

Divide the above second order central difference by $\Delta\tau^2$, and take the limit $\Delta\tau \downarrow 0$. It gives us

$$\begin{aligned} \frac{d^2 \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{d\tau^2} &= \lim_{\Delta\tau \downarrow 0} \frac{\Delta^2 \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{\Delta\tau^2} \\ &= \sum_{t=1}^T \left(\frac{d^2 F^{-1}(\tau)}{d\tau^2} (\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau)\}} - \tau) - 2 \frac{dF^{-1}(\tau)}{d\tau} \right), \end{aligned} \quad (31)$$

the last line of which is obtained similarly to (25). To interpret this result, we take expectation and get the following:

$$\mathbb{E} \left[\frac{d^2 \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{d\tau^2} \right] = -2 \frac{dF^{-1}(\tau)}{d\tau} \quad T < 0. \quad (32)$$

where the inequality holds with probability one since $f(\varepsilon) > 0$ with probability one in the assumption (A1). Now we have the expectation of $\frac{d^2 \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{d\tau^2}$ which can be regarded as the underlying guideline for the concavity of a SRAR curve. Together with the slope information, it implies that SRAR curves are always in arch shapes, going upward and then downward, with a peak point at $\mathbb{E}[\varepsilon_t] = F^{-1}(\tau)$. We can also know the skewness of ε_t from the location of the peak point: ε_t is left-skewed when the SRAR curve reaches its peak in the region $\tau < 0.5$, or right-skewed when the peak in $\tau > 0.5$. If ε_t is symmetrically distributed, its SRAR curve is symmetric, and vice versa.

5 Binding functions

Plotting SRAR is a way to present the goodness of fit in quantile regressions for each candidate model. Quantile regressions are the path to get residuals for SRAR calculation. As we know and provide unbiased consistent estimation for true models. To study the estimation in misspecification we adopt the concept of binding function (Dhaene, Gouriéroux, and Scaillet 1998). Binding function is defined as a mapping from coefficients in the true model to pseudo-true coefficients in a misspecified model.

The estimator of a pseudo-true coefficient in quantile regression for a misspecified QCAR(p) or QNCAR(p) converges to a limiting value which is characterized into the binding function. It is difficult to derive the binding functions explicitly in a general case so that they are studied by means of simulations (see Gouriéroux and Jasiak 2017). Suppose a noncausal AR(1): $y_t = \pi_1 y_{t+1} + \varepsilon_t$, with $\{\varepsilon_t\}$ i.i.d. $t(v)$ for $v = 1, 3, 5$ and 10 . It is observed that the binding function in the misspecified QCAR(1) varies with two factors: (i) the distribution of ε_t and (ii) the distance function in regression which is the check function $\rho_\tau(\cdot)$ in quantile regression. Figure 10, Figure 11 and Figure 12 illustrate the effect of those factors. Each point is an average value of estimates based on 1000

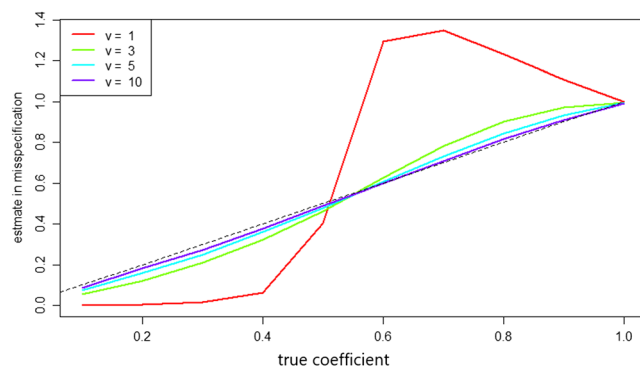


Figure 10: Quarterly inflation rate series plot for 4 Latin American countries.

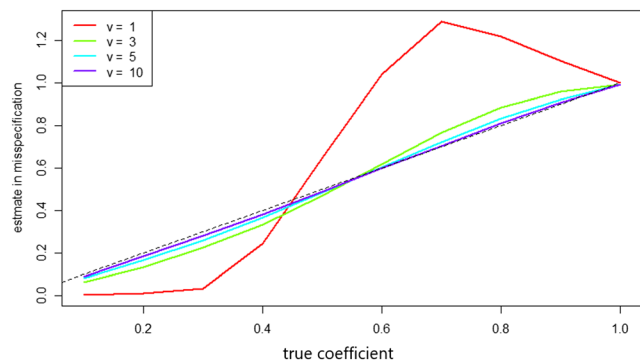


Figure 11: Binding function for a misspecified QCAR(1) in 0.3rd-quantile regression.

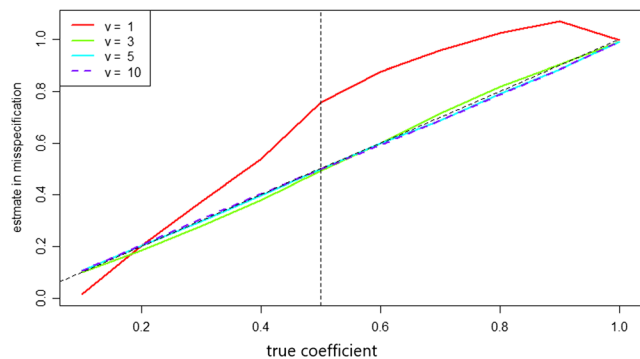


Figure 12: Binding function for a misspecified QCAR(1) in 0.1st-quantile regression.

simulations and 600 observations. Since $t(v)$ is symmetric, the estimation results are in the same pattern for negative true coefficient region and $(1-\tau)$ th-quantile regression as in these three figures. Sometimes the binding function is not injective, which is evidenced in Figure 10 and Figure 11 for small absolute true coefficients. The non-injectivity of the binding function for Cauchy distributed innovations is also illustrated in Gouriéroux and Jasiak (2017) result that disables encompassing tests. On the other hand, we see that on Figure 12 the injectivity of binding functions seems recovered at $\tau = 10\%$. In the case of Cauchy distributed innovations, there are no binding functions from extreme quantile regressions like 0.1 th- or 0.9 th-quantile regression because the estimate is not convergent. Although a value for $\pi_1 \in (0,1)$ is plotted in Figure 12, it is just the average of binding function estimates for π_1 for illustration.

6 Modelling hyperinflation in Latin America

6.1 The model specification

The motivation of our empirical analysis comes from the rational expectation (RE) hyperinflation model originally proposed by Cagan (1956) and investigated by several authors (see e.g. Adam and Szafarz 1992; Broze and Szafarz 1985). We follow Broze and Szafarz (1985) notations with

$$m_t^d = \alpha p_t + \beta E(p_{t+1}|I_t) + x_t. \quad (33)$$

In (33), m_t^d and p_t respectively denote the logarithms of money demand and price, x_t is the disturbance term summarizing the impact of exogenous factors. $E(p_{t+1}|I_t)$ is the rational expectation, when it is equal to conditional expectation, of p_{t+1} at time t based on the information set I_t . Assuming that the money supply $m_t^s = z_t$ is exogenous, the equilibrium $m_t^d = m_t^s$ provides the following equation for prices

$$\begin{aligned} p_t &= -\frac{\beta}{\alpha} [E(p_{t+1}|I_t)] + \frac{z_t - x_t}{\alpha}, \\ &= \phi [E(p_{t+1}|I_t)] + u_t. \end{aligned}$$

Broze and Szafarz (1985) show that a forward-looking recursive solution of this model exists when x_t is stationary and $|\phi| < 1$. The deviation from that solution is called the bubble B_t with $p_t = \sum_{i=0}^{\infty} \phi^i E(u_{t+i}|I_t) + B_t$. Finding conditions under which this process has rational expectation equilibria (forward and or backward looking) is out of the scope of our paper. We only use this framework to illustrate the interest of economists for models with leads components. Under a perfect foresight scheme $E(p_{t+1}|I_t) = p_{t+1}$ we obtain the purely noncausal model

$$p_t = \phi p_{t+1} + \tilde{\varepsilon}_t, \quad (34)$$

with $\tilde{\varepsilon}_t = u_t$. In the more general setting, for instance when $E(p_{t+1}|I_t) = p_{t+1} + v_t$ with v_t a martingale difference, the new disturbance term is $\tilde{\varepsilon}_t = v_t + u_t$. Empirically, a specification with one lead only might be too restrictive to capture the underlying dynamics of the observed variables. We consequently depart from the theoretical model proposed above and we consider empirical specifications with more leads or lags. Lanne and Luoto (2013, 2017) and Hecq, Lieb, and Telg (2017a, 2017b) in the context of the new hybrid Keynesian Phillips curve assume for instance that $\tilde{\varepsilon}_t$ is a $MAR(r-1, s-1)$ process such as

$$\rho(L)\pi(L^{-1})\tilde{\varepsilon}_t = c + \varepsilon_t, \quad (35)$$

where ε_t is *iid* and c an intercept term. Inserting (35) in (34) we observe that if $\tilde{\varepsilon}_t$ is a purely noncausal model (i.e. a $MAR(0, s-1)$ with $\rho(L) = 1$) we obtain a noncausal $MAR(0, s)$ motion for prices

$$\begin{aligned} (1 - \phi L^{-1})p_t &= \pi(L^{-1})^{-1}(c + \varepsilon_t), \\ (1 - \phi L^{-1})(1 - \pi_1 L^{-1} - \dots - \pi_{s-1} L^{-(s-1)})p_t &= c + \varepsilon_t, \end{aligned}$$

We would obtain a mixed causal and noncausal model if $\rho(L) \neq 1$. Our guess that the same specification might in some circumstances empirically (although not mathematically as the lag polynomial does not annihilate the lead polynomial) gives rise to a purely causal model in small samples when the autoregressive part dominates the lead component.

The above illustration presents a context of pure causal and noncausal models so that we can apply our approach to give an empirical analysis. It would be interesting to extend our modelling to investigate theoretical models with both forward and backward behaviours such as backward- and forward-looking Taylor rule for instance. To do that however we have to introduce additional regressors and extend the approach of Hecq, Issler, and Telg (2020) to quantile regressions, which can be further investigated by future research and is out of the scope of this paper.

6.2 The data and unit root testing

We consider seasonally unadjusted quarterly Consumer Price Index (CPI) series for four Latin American countries: Brazil, Mexico, Costa Rica and Chile. Monthly raw price series are downloaded at the OECD database for the largest span available (in September 2018). Despite the fact that quarterly data are directly available at OECD, we do not consider those series as they are computed from the unweighted average over three months of the corresponding quarters. Hence, these data are constructed using a linear filter, leading to undesirable properties for the detection of mixed causal and noncausal models (see Hecq, Telg, and Lieb 2017a, b on this specific issue). As a consequence, we use quarterly data computed by point-in-time sampling from monthly variables. The first observation is 1969Q1 for Mexico, 1970Q1 for Chile, 1976Q1 for Costa Rica and 1979Q4 for Brazil. Our last observation is 2018Q2 for every series. We do not use monthly data in this paper as monthly inflation series required a very large number of lags to capture their dynamic feature. Moreover, the detection of seasonal unit roots in the level of monthly price series was quite difficult.

Applying seasonal unit root tests (here HEGY tests, see Hylleberg et al. 1990) with a constant, a linear trend and deterministic seasonal dummies, we reject (see Table 2 in which a * denotes a rejection of the null unit root hypothesis at a specific frequency corresponding to 5% significance level) the null of seasonal unit roots in each series whereas we do not reject the null of a unit root at the zero frequency. The implementation of the unit root tests here is concerned with conditional mean models of the raw data to ensure that we process the data and use its weakly stationary time series in quantile regressions for analysis. The unit root testing can also been done per quantile (Koenker and Xiao 2004) to relate short-term explosiveness of time series to unit-root quantile models, which is an interesting perspective to treat explosive time series and alternative to causal and noncausal modelling. We do not go deeper in the unit-root direction for this paper but with its outlook on future research.

The number of lags of the dependent variable used to whiten for the presence of autocorrelation is chosen by AIC. From these results we compute quarterly inflation rates for the four countries in annualized rate, i.e. $\Delta \ln P_t^i \times 400$. Next we carry out a regression of $\Delta \ln P_t^i \times 400$ on seasonal dummies to capture the potential presence of deterministic seasonality. The null of no deterministic seasonality is not rejected for the four series. Figure 13 displays quarterly inflation rates and it illustrates the huge inflation episodes that the countries had faced. Among the four inflation rates, Brazil and Mexico show the typical pattern closer to the intuitive notion

Table 2: Seasonal HEGY unit root tests in the log levels of prices.

Country	$H_0: \pi_1=0$	$H_0: \pi_2=0$	$H_0: \pi_3=\pi_4=0$	Sample
$\ln P_t^{Bra}$	-1.39	-5.75*	48.28*	1979Q4–2018Q2
$\ln P_t^{Chi}$	-2.98	-6.32*	20.13*	1970Q1–2018Q2
$\ln P_t^{Costa}$	-1.80	-4.23*	7.81*	1976Q1–2018Q2
$\ln P_t^{Mex}$	-0.88	-11.92*	60.10*	1969Q1–2018Q2

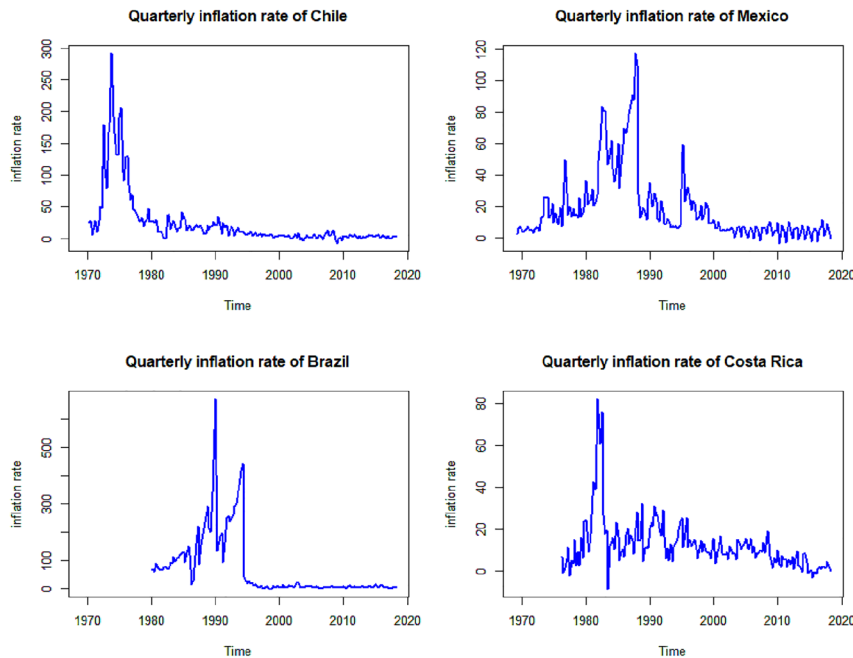


Figure 13: Quarterly inflation rate series plot for 4 Latin American countries.

Table 3: Descriptive statistics for quarterly inflation rates.

Country	HQ	BJ	skew.	kurt.	LM[1–2]	ARCH[1–2]
$\Delta \ln P_t^{Bra}$	1	<0.001	−2.54	56.96	0.19	<0.001
$\Delta \ln P_t^{Chi}$	7	<0.001	2.84	22.45	0.09	0.09
$\Delta \ln P_t^{Costa}$	4	<0.001	1.01	8.73	0.47	0.30
$\Delta \ln P_t^{Mex}$	3	<0.001	−0.40	13.81	0.20	<0.001

Table 4: SRAR identification results.

Country	SRAR _{$\tau=0.1$}	SRAR _{$\tau=0.3$}	SRAR _{$\tau=0.5$}	SRAR _{$\tau=0.7$}	SRAR _{$\tau=0.9$}	SRAR _{total}
$\Delta \ln P_t^{Bra}$	MAR(0, 1)	MAR(0, 1)	MAR(0, 1)	MAR(1, 0)	MAR(1, 0)	MAR(0, 1)
$\Delta \ln P_t^{Chi}$	MAR(7, 0)	MAR(7, 0)	MAR(7, 0)	MAR(0, 7)	MAR(0, 7)	MAR(7, 0)
$\Delta \ln P_t^{Costa}$	MAR(0, 4)	MAR(0, 4)	MAR(0, 4)	MAR(4, 0)	MAR(4, 0)	MAR(0, 4)
$\Delta \ln P_t^{Mex}$	MAR(0, 3)	MAR(0, 3)	MAR(0, 3)	MAR(3, 0)	MAR(3, 0)	MAR(0, 3)

of what a speculative bubble is, namely a rapid increase of the series until a maximum value is reached before the bubble bursts.

6.3 Empirical findings and identification of noncausal models

Table 3 reports for each quarterly inflation rates the autoregressive model obtained using the Hannan–Quinn information criterion. Given our results on the binding function (see also Gouriéroux and Jasiak 2017) it is safer to determine the pseudo-true autoregressive lag length using such an OLS approach than using quantile regressions or using maximum likelihood method. Indeed there is the risk that a regression in direct time from a noncausal DGP provides an underestimation of the lag order for some distributions (e.g. the Cauchy) and some values of the parameters.

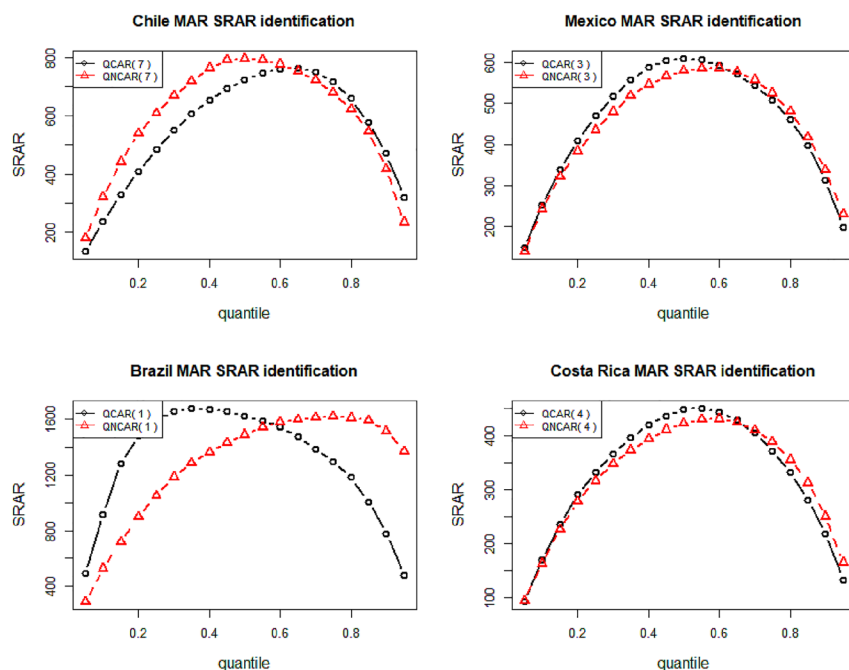


Figure 14: SRAR plots of the inflation rates of four Latin American countries respectively.

Estimating autoregressive univariate models gives the lag length range from $p = 1$ for Brazil to $p = 7$ for the Chilean inflation rate. The p -values of the Breusch-Pagan LM test (see column labelled $LM[1-2]$) for the null of no-autocorrelation after having included those lags show that we do not reject the null in every four cases. On the other hand, we reject the null of normality (Jarque-Bera test) in the disturbances of each series. We should consequently be able to identify causal from noncausal models. From columns *skew.* and *kurt.* it emerges that the residuals are skewed to the left for Brazil and Mexico and skewed to the right for Chile and Costa Rica. Heavy tails are present in each series. At a 5% significance level we reject the null of no ARCH (see column $ARCH[1-2]$) for Brazil and Mexico. Gouriéroux and Zakoian (2017) have derived the closed form conditional moments of a misspecified causal model obtained from a purely noncausal process with alpha stable disturbances. They show that the conditional mean (in direct time) is a random walk with a time varying conditional variance in the Cauchy case. This result would maybe favour the presence of a purely noncausal specification for Brazil and Mexico as the null of no ARCH is rejected. But this assertion must be carefully evaluated and tested, for instance using our comparison of quantile autoregressions in direct and reverse time. The results by the Q(N)CAR are reported in Table 4, and the RQ(N)CAR produces the same results. Each cell of Table 4 provides the selection frequency of $MAR(0, p)$ or $MAR(p, 0)$ identified by the SRAR at quantiles 0.1, 0.3, 0.5, 0.7, 0.9 as well as the aggregated SRAR. Figure 14 displays the SRAR curves from 0.05th-quantile to 0.95th-quantile by the Q(N)CAR for the four economies respectively, similarly to the ones by the RQ(N)CAR with restriction on non-negative regressors. As observed, the identification problem is raised in the SRAR plots. Especially in the SRAR plot for Brazil, it is hard to trust a model from evidence at single quantiles. However, the aggregate SRAR criterion comes to help in this situation from an overall perspective. We conclude that Brazil, Mexico and Costa Rica are better characterized as being purely noncausal while Chile being purely causal according to the aggregate SRAR criterion.

7 Conclusions

This paper introduces a new way to select between causal and noncausal models by comparing residuals from quantile autoregressions developed by Koenker and Xiao (2006) and from the time-reverse specifications. To adapt to heavy tailed distributions, we generalize the quantile autoregression theory for regularly varying

distributions. This also confirms the validity of quantile autoregressions in analysing heavy tailed time series, such as explosive or bubble-type dynamics. It is natural to consider SRAR as a model selection criterion in the quantile regression framework. However due to the identification problem spotted in the SRAR plots as presented in this paper, we propose to use the aggregate SRAR criterion for model selection. The robustness in its performance has been seen from all the results in this paper. It is worth mentioning that when coefficients are constant in the underlying model with a symmetrically i.i.d. error term, the aggregate SRAR criterion is equivalently to select between forward and backward conditional mean models (termed by Gouriéroux and Zakoian (2017)). However, the aggregate SRAR is a measure based on the whole dynamics of the underlying process, which is not dominated by the conditional mean information any more. This characteristic of the aggregate SRAR criterion indeed makes it robust in model selection even for some general situations such as with asymmetric distributed innovations. In the empirical study on the inflation rates of four Latin American countries, we found that the purely noncausal specification is favoured in three cases.

Finally some possible extensions of our approach can be to the identification of mixed models in addition to purely causal and noncausal specifications, to enhancing QCAR and QNCAR with some explanatory variables in order to investigate the Taylor (1993) rule, and to investigating the unit-root testing per quantile for QCAR as well as QNCAR. Also, a formal testing on SRAR differences would require the application of a bootstrap approach which is beyond the scope of this paper but in our outlook for the future research.

Acknowledgements: We would like to thank Eric Beutner, Sean Telg, Sebastien Fries as well as the participants in the following conferences and seminar places in which we have presented our paper: The 26th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, Tokyo, March 2018, The Eighth International Conference on Mathematical and Statistical Methods for Actuarial Sciences and Finance, Madrid, April 2018, The 12th NESG 2018 Conference, Amsterdam, May 2018, The 2nd International Conference on Econometrics and Statistics, Hong Kong, June 2018, The 4th Dongbei Econometrics Workshop, Dalian, June 2018, The 2018 IAAE Annual Conference International Association for Applied Econometrics, Montreal, June 2018, The 12th International Conference on Computational and Financial Econometrics, Pisa, December 2018, The University of Milano-Bicocca, The BI Norwegian Business School and The University of Nottingham. All errors are ours.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Appendix

Alternative way to simulate MAR models

Suppose that the DGP is a MAR(r, s) as in (1). First, we rewrite (1) into a matrix representation as follows:

$$\begin{aligned} \mathbf{M}\mathbf{y} &= \boldsymbol{\varepsilon}, \\ \mathbf{M} &:= \begin{bmatrix} \pi(L)\phi(L^{-1}) & 0 & \dots & 0 \\ 0 & \pi(L)\phi(L^{-1}) & \dots & 0 \\ & \dots & \dots & \dots \\ 0 & 0 & \dots & \pi(L)\phi(L^{-1}) \end{bmatrix}, \\ \mathbf{y} &:= [y_1 \ y_2 \ \dots \ y_T]', \\ \boldsymbol{\varepsilon} &:= [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_T]', \end{aligned} \quad (36)$$

where \mathbf{M} is $T \times T$ matrix and T is the sample size. The equivalence to (1) holds by assuming $y_{1-r}, y_{2-r}, \dots, y_0$ and $y_{T+1}, y_{T+2}, \dots, y_{T+s}$ are all zeros. This assumption effect can be neglected by deleting enough observations from the beginning and the end of a simulated sample, for instance, $\{y_t\}_{t=201}^{T-200}$ kept for analysis from a first simulated $\{y_t\}_{t=1}^T$. Next, \mathbf{M} can be decomposed into a product of two diagonal matrices, denoted as \mathbf{L} and \mathbf{U} , of main diagonal entries being $\pi(L)$ and $\phi(L^{-1})$ respectively as follows.

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -\pi_1 & 1 & 0 & 0 & \dots & 0 \\ -\pi_2 & -\pi_1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & -\pi_r & \dots & -\pi_1 & 1 \end{bmatrix},$$

$$\mathbf{U} = \begin{bmatrix} 1 & -\psi_1 & \dots & -\psi_s & 0 & \dots & 0 \\ 0 & 1 & -\psi_1 & \dots & -\psi_s & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & \dots & 1 \end{bmatrix} \quad (37)$$

Substitute (37) into (36). We get

$$\mathbf{L}\mathbf{U}\mathbf{y} = \boldsymbol{\varepsilon},$$

such as

$$\mathbf{y} = \mathbf{U}^{-1}\mathbf{L}^{-1}\boldsymbol{\varepsilon}. \quad (38)$$

Given $\boldsymbol{\varepsilon}$, \mathbf{y} can be obtained directly since \mathbf{L} and \mathbf{U} are positive definite triangular matrices. This $MAR(r, s)$ simulating method can easily be generalized, for instance, for an $MAR(r, s)$ involving some exogenous independent variables presented by Hecq, Issler, and Telg (2020). In practise this vector-wise simulation method is slower than the element-wise method because of the matrix creation and storage in simulation.

Proof of Theorem 3

Proof.

First, we rewrite $SRAR(\tau, \hat{\boldsymbol{\theta}}(\tau))$ as follows:

$$\begin{aligned} SRAR(\tau, \hat{\boldsymbol{\theta}}(\tau)) &= \sum_{t=1}^T \rho_{\tau}(y_t - \mathbf{x}_t' \hat{\boldsymbol{\theta}}(\tau)) \\ &= \sum_{t=1}^T \rho_{\tau}(y_t - \mathbf{x}_t' \boldsymbol{\phi}_{\tau} + \mathbf{x}_t' \boldsymbol{\phi}_{\tau} - \mathbf{x}_t' \hat{\boldsymbol{\theta}}(\tau)) \\ &= \sum_{t=1}^T \rho_{\tau}\left(u_{t\tau} - \frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t\right), \end{aligned} \quad (39)$$

where $\mathbf{x}_t' := [a_T, y_{t+1}, \dots, y_{t+p}]$, $u_{t\tau} := y_t - \mathbf{x}_t' \boldsymbol{\phi}_{\tau} = \varepsilon_t - F^{-1}(\tau)$, $\mathbf{v} := a_T \sqrt{T}(\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\phi}_{\tau})$. We know from Davis and Resnick (1985) and Knight (1989, 1991) that

$$\begin{aligned} \frac{1}{a_T} \left(\sum_{t=1}^{\lfloor T \cdot s \rfloor} (\varepsilon_t - b_T) \right) &\stackrel{d}{\sim} S_{\alpha}(s), \\ \frac{1}{a_T \sqrt{T}} \sum_{t=1}^T \left(y_t - \lfloor T \cdot s \rfloor \sum_{j=0}^{\infty} c_j b_T \right) &\stackrel{d}{\sim} \sum_{j=0}^{\infty} c_j \int_0^1 S_{\alpha}(s) ds, \\ \frac{1}{a_T^2 T} \sum_{t=1}^T \left(y_t \cdot y_{t+h} - \lfloor T \cdot s \rfloor \sum_{j=0}^{\infty} c_j c_{j+h} b_T^2 \right) &\stackrel{d}{\sim} \sum_{j=0}^{\infty} c_j c_{j+h} \int_0^1 S_{\alpha}^2(s) ds, \end{aligned} \quad (40)$$

where $t = \lfloor T \cdot s \rfloor$, and $\{S_{\alpha}(s)\}$ is a process of stable distributions with index α . Without loss of generality, we assume $b_T = 0$ for the proof below. In use of the limiting behaviour information presented in (40), we get that

$$\frac{1}{a_T^2 T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \stackrel{d}{\sim} \begin{bmatrix} 1 & 0 \\ 0 & \int_0^1 S_{\alpha}^2(s) ds \Omega \end{bmatrix}_{(p+1) \times (p+1)} \quad (41)$$

where

$$\begin{aligned} \Omega &:= [\omega_{ik}]_{p \times p}, \\ \omega_{ik} &:= \sum_{j=0}^{\infty} c_j c_{j+|k-i|}, \end{aligned}$$

with ω_{ik} being the entry at Ω 's i th row and k th column. Ω is positive definite symmetric. Note that $\hat{\theta}(\tau) = \arg \min_{\theta \in \mathbb{R}^{p+1}} \text{SRAR}(\tau, \theta)$ which also minimizes

$$Z_T(\mathbf{v}) := \sum_{t=1}^T \left[\rho_\tau \left(u_{tr} - \frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t \right) - \rho_\tau(u_{tr}) \right]. \quad (42)$$

$Z_T(\mathbf{v})$ is a convex random function. Knight (1989) showed that if $Z_T(\mathbf{v})$ converges in distribution to $Z(\mathbf{v})$ and $Z(\mathbf{v})$ has unique minimum, then the convexity of $Z_T(\mathbf{v})$ ensures $\hat{\mathbf{v}} = \arg \min_{\mathbf{v} \in \mathbb{R}^{p+1}} Z_T(\mathbf{v})$ converging in distribution to $\arg \min Z(\mathbf{v})$.

By using the following check function identity:

$$\begin{aligned} \rho_\tau(v_1 - v_2) - \rho_\tau(v_1) &= -v_2 \xi_\tau(v_1) + (v_1 - v_2) (I(0 > v_1 > v_2) - I(0 < v_1 < v_2)) \\ &= -v_2 \xi_\tau(v_1) + \int_0^{v_2} (I(v_1 \leq s) - I(v_1 < 0)) ds, \end{aligned} \quad (43)$$

where $\xi_\tau(v) = \tau - I(v < 0)$, we can rewrite $Z_T(\mathbf{v})$ into

$$\begin{aligned} Z_T(\mathbf{v}) &= - \sum_{t=1}^T \frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t \xi_\tau(u_{tr}) + \sum_{t=1}^T \int_0^{\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t} (I(u_{tr} \leq s) - I(u_{tr} < 0)) ds \\ &= Z_T^{(1)}(\mathbf{v}) + Z_T^{(2)}(\mathbf{v}), \end{aligned} \quad (44)$$

where $Z_T^{(2)}(\mathbf{v}) := \sum_{t=1}^T \int_0^{\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t} (I(u_{tr} \leq s) - I(u_{tr} < 0)) ds$ and $Z_T^{(1)}(\mathbf{v}) := - \sum_{t=1}^T \frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t \xi_\tau(u_{tr})$. Further denote $\eta_t(\mathbf{v}) := \int_0^{\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t} (I(u_{tr} \leq s) - I(u_{tr} < 0)) ds$, $\tilde{\eta}_t(\mathbf{v}) := E[\eta_t(\mathbf{v}) | \mathbf{x}_t]$ and $\tilde{Z}_T^{(2)}(\mathbf{v}) := \sum_{t=1}^T \tilde{\eta}_t(\mathbf{v})$. By Assumption (A5) and small enough $\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t$, we further rewrite $\tilde{Z}_T^{(2)}(\mathbf{v})$ as follows:

$$\begin{aligned} \tilde{Z}_T^{(2)}(\mathbf{v}) &= \sum_{t=1}^T E \left[\int_0^{\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t} (I(u_{tr} \leq s) - I(u_{tr} < 0)) ds \middle| \mathbf{x}_t \right] \\ &= \sum_{t=1}^T \int_0^{\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t} \left[\int_{F^{-1}(\tau)}^{s+F^{-1}(\tau)} f(r) dr \right] ds \\ &= \sum_{t=1}^T \int_0^{\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t} \frac{F(s + F^{-1}(\tau)) - F(F^{-1}(\tau))}{s} s ds \\ &= \sum_{t=1}^T \int_0^{\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t} f(F^{-1}(\tau)) s ds \\ &= \frac{1}{2a_T^2 T} \sum_{t=1}^T f(F^{-1}(\tau)) \mathbf{v}' \mathbf{x}_t \mathbf{x}_t' \mathbf{v} + o_p(1) \\ &= \frac{1}{2a_T^2 T} f(F^{-1}(\tau)) \mathbf{v}' \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right) \mathbf{v} + o_p(1) \end{aligned} \quad (45)$$

Using the limiting behaviour information presented in (40), we get the limiting distribution for $\tilde{Z}_T^{(2)}(\mathbf{v})$ so as for $Z_T^{(2)}(\mathbf{v})$ as follows:

$$Z_T^{(2)}(\mathbf{v}) \stackrel{d}{\sim} \frac{1}{2} f(F^{-1}(\tau)) \mathbf{v}' \begin{bmatrix} 1 & 0 \\ 0 & \int_0^1 S_\alpha^2(s) ds \Omega \end{bmatrix} \mathbf{v}, \quad (46)$$

by the fact that $Z_T^{(2)}(\mathbf{v}) - \tilde{Z}_T^{(2)}(\mathbf{v}) \xrightarrow{p} 0$ which can be proved by following the arguments of Knight (1989).

The limiting distribution of $Z_T^{(1)}(\mathbf{v})$ can also be deduced in using (40) as follows.

$$-\sum_{t=1}^T \frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t \xi_\tau(u_{t\tau}) \stackrel{d}{\sim} \mathbf{v}' \left[\sigma_\xi W(1), \sum_{j=0}^{\infty} c_j \sigma_\xi \int_0^1 S_\alpha(s) dW(s), \dots, \sum_{j=0}^{\infty} c_j \sigma_\xi \int_0^1 S_\alpha(s) dW(s) \right]_{(p+1) \times 1}, \quad (47)$$

where $[\dots]_{(p+1) \times 1}$ is a column vector of $(p+1)$ elements, $\int dW(s)$ is a stochastic integral with Brownian motion $\{W(s)\}$ independent of $\{S_\alpha(s)\}$ (see Knight (1991)), and σ_ξ is the standard deviation of $\xi_\tau(u_{t\tau})$ which equals $\sqrt{\tau(1-\tau)}$. Therefore by Davis and Resnick (1985) and Knight (1989, 1991),

$$Z_T^{(1)}(\mathbf{v}) \stackrel{d}{\sim} \mathbf{v}' \sqrt{\tau(1-\tau)} \left[W(1), \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) dW(s), \dots, \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) dW(s) \right]_{(p+1) \times 1}. \quad (48)$$

Thus,

$$Z_T(\mathbf{v}) \stackrel{d}{\sim} Z(\mathbf{v}) := \mathbf{v}' \sqrt{\tau(1-\tau)} \left[W(1), \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) dW(s), \dots, \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) dW(s) \right]_{(p+1) \times 1} \\ + \frac{1}{2} f(F^{-1}(\tau)) \mathbf{v}' \begin{bmatrix} 1 & 0 \\ 0 & \int_0^1 S_\alpha^2(s) ds \Omega \end{bmatrix} \mathbf{v}. \quad (49)$$

and so

$$\frac{f(F^{-1}(\tau)) \cdot a_T \sqrt{T}}{\sqrt{\tau(1-\tau)}} (\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\phi}_\tau) \stackrel{d}{\sim} \begin{bmatrix} 1 & 0 \\ 0 & \left(\int_0^1 S_\alpha^2(s) ds \Omega \right)^{-1} \end{bmatrix} \left[W(1), \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) dW(s), \dots, \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) dW(s) \right]_{(p+1) \times 1}.$$

follows by setting the derivative of $Z(\mathbf{v})$ to 0 and solving for \mathbf{v} . \square

References

- Adam, M. C., and A. Szafarz. 1992. "Speculative Bubbles and Financial Markets." *Oxford Economic Papers* 44: 626–40.
- Barrodale, I., and F. D. Roberts. 1973. "An Improved Algorithm for Discrete l_1 Linear Approximation." *SIAM Journal on Numerical Analysis* 10: 839–48.
- Breid, F. J., R. A. Davis, K.-S. Lh, and M. Rosenblatt. 1991. "Maximum Likelihood Estimation for Noncausal Autoregressive Processes." *Journal of Multivariate Analysis* 36: 175–98.
- Brockwell, P. J., and R. A. Davis. 2016. *Introduction to Time Series and Forecasting*. New York: Springer.
- Brockwell, P. J., R. A. Davis, and S. E. Fienberg. 1991. *Time Series: Theory and Methods*. New York: Springer Science & Business Media.
- Broze, L., and A. Szafarz. 1985. "Solutions des modèles linéaires à anticipations rationnelles." *Annales de l'INSEE*: 99–118, <https://doi.org/10.2307/20076541>.
- Cagan, P. 1956. "The Monetary Dynamics of Hyperinflation." In *Studies in the Quantity Theory of Money*. Chicago: The University of Chicago Press.
- Davis, R., and S. Resnick. 1985. "Limit Theory for Moving Averages of Random Variables with Regularly Varying Tail Probabilities." *Annals of Probability*: 179–195, <https://doi.org/10.1214/aop/1176993074>.
- Dhaene, G., C. Gouriéroux, and O. Scaillet. 1998. "Instrumental Models and Indirect Encompassing." *Econometrica*: 673–688, <https://doi.org/10.2307/2998579>.
- Fan, J., and Y. Fan. 2010. *Issues on Quantile Autoregression*. Also available at <https://orfe.princeton.edu/~jqfan/papers/06/quantilereg.pdf>.
- Francq, C., and J.-M. Zakoïan. 2007. "Hac Estimation and Strong Linearity Testing in Weak Arma Models." *Journal of Multivariate Analysis* 98: 114–144.

- Fries, S., and J.-M. Zakoian. 2017. *Mixed Causal-Noncausal Ar Processes and the Modelling of Explosive Bubbles*. Munich Personal RePEc Archive. Also available at <https://mpra.ub.uni-muenchen.de/86926/>.
- Gourieroux, C., and J. Jasiak. 2016. "Filtering, Prediction and Simulation Methods for Noncausal Processes." *Journal of Time Series Analysis* 37: 405–430.
- Gourieroux, C., and J. Jasiak. 2015. *Semi-Parametric Estimation of Noncausal Vector Autoregression*. CREST. Also available at https://www.researchgate.net/profile/Joann_Jasiak/publication/308047972_REVAR/links/57d812ce08ae6399a399137e.pdf.
- Gourieroux, C., and J.-M. Zakoian. 2017. "Local Explosion Modelling by Non-causal Process." *Journal of the Royal Statistical Society: Series B Statistical Methodology* 79: 737–56.
- Hecq, A., J. V. Issler, and S. Telg. 2020. "Mixed Causal–Noncausal Autoregressions with Exogenous Regressors." *Journal of Applied Econometrics* 35: 328–343.
- Hecq, A., L. Lieb, and S. Telg. 2016. "Identification of Mixed Causal-Noncausal Models in Finite Samples." *Annals of Economics and Statistics/Annales d'Économie et de Statistique*: 307–331, <https://doi.org/10.15609/annaeconstat2009.123-124.0307>.
- Hecq, A., L. Lieb, and S. Telg. 2017a. *Simulation, Estimation and Selection of Mixed Causal-Noncausal Autoregressive Models: The Marx Package*. 3015797, SSRN. Also available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3015797.
- Hecq, A., S. Telg, and L. Lieb. 2017b. "Do seasonal Adjustments Induce Noncausal Dynamics in Inflation Rates?." *Econometrics* 5: 48.
- Hylleberg, S., R. F. Engle, C. W. Granger, and B. S. Yoo. 1990. "Seasonal Integration and Cointegration." *Journal of Econometrics* 44: 215–38.
- Knight, K. 1989. "Limit Theory for Autoregressive-Parameter Estimates in an Infinite-Variance Random Walk." *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*: 261–78, <https://doi.org/10.2307/3315522>.
- Knight, K. 1991. "Limit Theory for M-Estimates in an Integrated Infinite Variance Process." *Econometric Theory*: 200–12, <https://doi.org/10.1017/s0266466600004400>.
- Koenker, R., and Z. Xiao. 2004. "Unit Root Quantile Autoregression Inference." *Journal of the American Statistical Association* 99: 775–787.
- Koenker, R., and Z. Xiao. 2006. "Quantile Autoregression." *Journal of the American Statistical Association* 101: 980–90.
- Lanne, M., and J. Luoto. 2013. "Autoregression-based Estimation of the New Keynesian Phillips Curve." *Journal of Economic Dynamics and Control* 37: 561–70.
- Lanne, M., and J. Luoto. 2017. "A New Time-Varying Parameter Autoregressive Model for US Inflation Expectations." *Journal of Money, Credit, and Banking* 49: 969–95.
- Lanne, M., and P. Saikkonen. 2011a. "Gmm Estimation with Non-causal Instruments." *Oxford Bulletin of Economics & Statistics* 73: 581–592.
- Lanne, M., and P. Saikkonen. 2011b. "Noncausal Autoregressions for Economic Time Series." *Journal of Time Series Econometrics* 3, <https://doi.org/10.2202/1941-1928.1080>.
- Lanne, M., and P. Saikkonen. 2013. "Noncausal Vector Autoregression." *Econometric Theory*: 447–481, <https://doi.org/10.1017/s0266466612000448>.
- Liu, X. 2019. "Quantile-based Asymmetric Dynamics of Real Gdp Growth." *Macroeconomic Dynamics*: 1–29, <https://doi.org/10.1017/S1365100519000063>.
- Taylor, J. B. 1993. "Discretion versus Policy Rules in Practice," In *Carnegie-Rochester Conference Series on Public Policy*, 195–214. Amsterdam, The Netherlands: Elsevier.
- Tokdar, S. T., J. B. Kadane 2012. "Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach." *Bayesian Analysis* 7: 51–72.