City Research Online

City, University of London Institutional Repository

This is the draft version of the paper.

This version of the publication may differ from the final published version.

Department of Economics

Threshold Quantile Autoregressive Models

Antonio F. Galvao, Jr.[1]
University of Illinois at Urbana-Champain

Gabriel Montes-Rojas[2]
City University London

Jose Olmo[3]
City University London

[1] Department of Economics, University of Illinois at Urbana-Champaign, 419 David Kinley Hall, 1407 W. Gregory Dr., Urbana, IL 61801, USA.
Email: *galvao@illinois.edu*
[2] Department of Economics, City University London, D308 Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK.
Email: *Gabriel.Montes-Rojas.1@city.ac.uk*. T: +44 (0)20 7040 8919 F: +44 (0)20 7040 8580
[3] Department of Economics, City University London, D308 Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK.
Email: *J.Olmo@city.ac.uk*. T: +44 (0)20 7040 4129 F: +44 (0)20 7040 8580

Threshold Quantile Autoregressive Models

ANTONIO F. GALVAO, *University of Illinois at Urbana-Champaign*
GABRIEL MONTES-ROJAS , *City University, London*
JOSE OLMO*, *City University, London*

November, 2008

Abstract

We study in this article threshold quantile autoregressive processes. In particular we propose estimation and inference of the parameters in nonlinear quantile processes when the threshold parameter defining nonlinearities is known for each quantile, and also when the parameter vector is estimated consistently. We derive the asymptotic properties of the nonlinear threshold quantile autoregressive estimator. In addition, we develop hypothesis tests for detecting threshold nonlinearities in the quantile process when the threshold parameter vector is not identified under the null hypothesis. In this case we propose to approximate the asymptotic distribution of the composite test using a p-value transformation. This test contributes to the literature on nonlinearity tests by extending Hansen's (Econometrica 64, 1996, pp.413-430) methodology for the conditional mean process to the entire quantile process. We apply the proposed methodology to model the dynamics of US unemployment growth after the Second World War. The results show evidence of important heterogeneity associated with unemployment, and strong asymmetric persistence on unemployment growth.

Keywords: Nonlinear models; quantile regression; threshold models

JEL Classification: C14, C22, C32, C50

---

*Corresponding Author: Department of Economics, City University London, Northampton Square, London EC1V 0HB, UK, email: `J.Olmo@city.ac.uk`, tel: +44 (0)20-7040-4129.

1 Introduction

During the last twenty years the choice of linear time series models has proven to be rather limited and inadequate for describing the dynamics of macroeconomic and financial time series. These data are plagued with nonlinear phenomena such as asymmetries in the occurrence and persistence of negative and positive shocks, time-irreversibility, different tail behavior of the distribution of the data, asymmetric effects of the heteroskedasticity in the mean process. These phenomena have sparked a growing interest in applied and theoretical econometrics for developing time series processes modeling these nonlinearities in various ways.

Among the models that have enjoyed greater popularity are the linear threshold autoregressive models (TAR) of Tong and Lim (1980), Tong (1983, 1990) or Tsay (1989). Extensions of these models to accommodate stylized facts in macroeconomics and finance are abundant in the literature. Thus, Nelson (1991), Glosten, Jagannatthan, and Runkle (1993) or Zakoian (1994) propose threshold models in the conditional volatility process capable of describing the asymmetries observed in the conditional volatility process due to feedback and leverage effects. Another example is Sarno, Valente, and Leon (2006) or Clarida, Sarno, Taylor, and Valente (2006) that make use of these processes to model the dynamics and asymmetries of exchange rates and the term structure of interest rates, respectively. An alternative widely explored in the literature for modeling nonlinearities are the Smooth Transition Models (STAR). These models are characterized by an infinite number of regimes and a state variable that changes smoothly from one state to the other, see Chan and Tong (1986), Terasvirta (1994) and the references therein for general models.

All these models are aimed at describing the dynamics of the conditional mean process. These methods, however, are inadequate to describe nonlinearities in the data produced by general forms of heterogeneity, when different regimes coincide in the mean process. The idea that time series may have different behavior across quantiles has attracted considerable attention in the theoretical literature (see for instance Koenker and Xiao, 2006, and the references therein) and a natural extension is the existence of different regimes depending on the quantile of the series to be modeled. Thus, the first contribution of this article is to introduce threshold quantile autoregressive processes (T-QAR), and to derive the asymptotic properties of the nonlinear threshold quantile autoregressive estimator. We build on the studies of Koenker and Xiao (2006) for quantile autoregressive processes, and Cai and Stander (2008) for the study of self-exciting processes in quantiles. In particular, we concentrate on studying estimation and inference for the parameters of nonlinear threshold quantile autoregressive processes. More specifically, our study will be divided in two different cases. One,

when the threshold parameter defining the nonlinearity in the quantile is known, and a second scenario when the parameter needs to be estimated.

The second contribution of this article consists on developing hypothesis tests to detect threshold nonlinearities in the quantile process. Following Hansen's (1996) approach we propose heteroskedasticity-robust Wald tests for the case when the threshold parameter of each quantile is known and also for the case when the parameter is not identified under the null hypothesis for some quantile. Whereas for the former case the asymptotic distribution is chi-squared for each quantile, for the latter case the asymptotic distribution is nonstandard and is approximated by simulation methods. We choose the p-value transformation method for the supremum and average of the relevant Wald statistic. Finally, we extend these tests to detect nonlinearities in the entire quantile process. In order to do this we will use Kolmogorov-Smirnov type tests over the quantile process. These tests contribute to the literature on nonlinearity tests by extending Hansen (1996) methodology for the conditional mean process to the entire quantile process.

To motivate the usefulness of this methodology, consider the following example that illustrates the lack of power of a standard threshold nonlinearity test to detect nonlinearities in the quantile process.

Consider the following two location-scale models:

$$y_t = \theta_0 + \theta_1 y_{t-1} + \delta y_{t-1} u_t, \tag{1}$$

and

$$y_t = \theta_0 + \theta_1 y_{t-1} + \delta |y_{t-1}| u_t, \tag{2}$$

where u_t is an independent and identically distributed (iid) process with mean 0, variance $\sigma_u^2 < \infty$, and symmetric and differentiable distribution function F_u.

Koenker (2005, p.57) shows that the quantile process of (1) is linear if F_u is symmetric. This is also true for model (2) if $y_t > 0$ for all t, but interesting nonlinearities appear if y_t can take positive and negative values. In this case, $|y_{t-1}|$ generates differences across quantiles depending on the sign of y_t. In particular, the model becomes a two-regime model of the form

$$y_t = \begin{cases} \theta_0 + \theta_1 y_{t-1} - \delta y_{t-1} u_t, & y_{t-1} \leq 0, \\ \theta_0 + \theta_1 y_{t-1} + \delta y_{t-1} u_t, & y_{t-1} > 0. \end{cases} \tag{3}$$

Although the model is nonlinear, its first conditional moments are identical across regimes,

$$E[y_t | y_{t-1} \leq 0] = E[y_t | y_{t-1} > 0] = \theta_0 + \theta_1 y_{t-1}, \tag{4}$$

3

and

$$Var[y_t|y_{t-1} \leq 0] = Var[y_t|y_{t-1} > 0] = \delta^2 y_{t-1}^2 \sigma_u^2. \tag{5}$$

This model generates heterogeneity and nonlinearity in the quantile process that the standard $F-$test and the heteroskedasticity-robust Wald test are not able to detect. The heteroskedastic residual sequence $a_t = \delta y_{t-1} u_t$ from (1) and $a_t'(0) = (\delta - 2\delta I(y_{t-1} > 0))y_{t-1}u_t$ from model (2) are different but satisfy that $a_t^2 = a_t'(0)^2$ for all $t \in \mathbb{N}$, implying the insensitiveness of the Hansen test to this type of nonlinearities. In fact, any procedure based on the squared residual sequence will fail because of the equality of squared residuals. However, the conditional τ-quantile process of model (3), assuming $\delta > 0$ is

$$Q_{y_t}(\tau|y_{t-1} \leq 0) = \theta_0 + (\theta_1 - \delta F_u^{-1}(\tau))y_{t-1}, \tag{6}$$

and

$$Q_{y_t}(\tau|y_{t-1} > 0) = \theta_0 + (\theta_1 + \delta F_u^{-1}(\tau))y_{t-1}. \tag{7}$$

Note that unless $F_u^{-1}(\tau) = 0$ both regimes are clearly identified. If the distribution of u is symmetric, then $F_u^{-1}(1/2) = 0$, and therefore, the conditional quantile processes will coincide at the median, but they will differ as we go to the upper or lower extreme quantiles.

Finally, we illustrate the two contributions of this paper with an application to unemployment growth series. In particular we postulate that there is heterogeneity on US unemployment growth, *i.e.* growth in the number of people unemployed, after the second world war. In our study, we claim that this heterogeneity can be due to asymmetric dynamics in the quantile process that for some quantiles of the process can potentially depend in different ways on previous values of the variable and lead us, therefore, to propose different autoregressive regimes for different quantiles. The results show evidence of important heterogeneity associated with unemployment, and strong asymmetric persistence on unemployment growth.

The article is structured as follows. Section 2 introduces the family of threshold quantile regime switching models and the statistical properties, in particular, estimation procedures, consistency and asymptotic normality of the estimators. Section 3 discusses different hypothesis tests, alternative to Hansen (1996), for linearity of the process at a given quantile and for the entire quantile process. In Section 4 we carry out a Monte Carlo experiment where we study the estimators' properties and different features of nonlinearity tests. Section 5 discusses an application to study nonlinearities in the dynamics of the quantile process of unemployment growth. Section 6 concludes. The proofs are gathered in a mathematical appendix.

2 Statistical Model, Estimation and Inference

2.1 Statistical Model

Consider a nonlinear process $\{y_t\}$ with two possible regimes defined by the known function $q_t = q(x_t)$ with $x_t = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p_y})$, and threshold parameter γ, $-\infty < \gamma < \infty$. To make allowance for different values of γ across quantiles we will index these parameters by τ.

We consider an extension of Koenker and Xiao (2006) quantile autoregressive model to incorporate the existence of different regimes. Let $\{u_t\}$ be a sequence of *iid* standard uniform random variables. The response variable y_t is modeled by

$$y_t = \mu(u_t, \gamma(u_t)), \tag{8}$$

where $\mu(\cdot, \cdot)$ is a piece-wise continuous process. For our purposes, assume that

$$\mu(u_t, \gamma(u_t)) = x_t(\gamma(u_t))'\theta(u_t) \tag{9}$$

where $\theta(\cdot)$ is a $2 \times (1 + p_y)$ dimensional vector of parameters, and $x(\gamma(\cdot)) = x \otimes (1\{q \leq \gamma(\cdot)\}, 1\{q > \gamma(\cdot)\})$, with $1\{\cdot\}$ an indicator function and $\otimes$ the Kronecker product.

Assuming that the right hand side of equation (8) is monotone increasing in u_t, it follows that the conditional quantile function of y_t can be written as[1]

$$Q_{y_t}(\tau|\Im_{t-1}) = \mu(\tau, \gamma(\tau)), \tag{10}$$

with $\Im_{t-1}$ denoting the σ-field generated by $\{y_s, s \leq t - 1\}$.

The model allows for different values of the threshold parameter $\gamma(.)$ across different quantiles, but such that the threshold parameter is given for each fixed τ and $\gamma(.)$ is a continuous function in τ. This model captures systematic influences on the location, scale and shape of the conditional distribution function, and thus, it allows for a significant extension of linear regression and time series models in two dimensions; by allowing for threshold effects and by modeling the whole quantile process.

---

[1]The transition from (9) to (10) is an immediate consequence of the fact that for any monotone increasing function g and standard uniform random variable, u, we have

$$Q_{g(u)}(\tau) = g(Q_u(\tau)) = g(\tau),$$

where $Q_u(\tau)$ is the quantile function of u.

2.2 Estimation and Inference

For models with a known threshold parameter $\gamma_0 \equiv \gamma_0(\tau)$ standard quantile regression (QR) estimation procedures can be applied. These methods consist on finding

$$\widehat{\theta}_{\gamma_0}(\tau) = \arg\min_\theta \sum_{t=1}^{T} \rho_\tau \left(y_t - x_t(\gamma_0)'\theta\right), \tag{11}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$, as in Koenker and Bassett (1978).

The consistency of this vector of estimators for each fixed τ is achieved by the argmax theorem in van der Vaart and Wellner (1996), and it is shown for pure quantile autoregressive processes in Koenker and Xiao (2006). These authors also show the asymptotic normality of this type of QR estimators under proper standardization. In order to generalize Koenker and Xiao (2006) to the regime switching model we need the following set of assumptions.

Assumptions:

A1: $\{y_t\}$ is a strictly stationary and ergodic sequence;

A2: $\tau \in \mathcal{T} = [c, 1 - c]$ with $c \in (0, 1/2)$. $\gamma(\tau)$ lies in a compact set $\mathcal{G} \subset \mathbb{R}$ for every $\tau \in \mathcal{T}$;

A3: $E(\|x_t\|^2) < \infty$ with $\|\cdot\|$ the usual Euclidean norm, and $\max_t \|x_t\| = O(\sqrt{T})$;

A4: Let $F(.|\Im_t) = F_t(.)$ denote the conditional distribution function of y_t given $\Im_t$. $F_t(.)$ has a continuous Lebesgue density, f_t, with $0 < f_t(u) < \infty$ on $\mathcal{U} = \{u : 0 < F_t(u) < 1\}$ and f_t is uniformly integrable on $\mathcal{U}$;

A5: Let $\theta_\gamma(\tau) = \arg\min_\theta E\left[\rho_\tau\left(y_t - x_t(\gamma)'\theta\right)\right]$. For all $\tau \in \mathcal{T}$ and $\gamma \in \mathcal{G}$, $\theta_\gamma(\tau)$ exists, is unique, and $\theta_\gamma(\tau) \in int \, \Theta$, with Θ compact and convex;

A6: $\widehat{\Omega}_0(\gamma, \gamma^*) = \frac{1}{T}\sum_{t=1}^{T} x_t(\gamma)x_t'(\gamma^*)$ and $\widehat{\Omega}_1(\tau, \gamma) = \frac{1}{T}\sum_{t=1}^{T} \widehat{f}_{t-1}(F_{t-1}^{-1}(\tau))x_t(\gamma)x_t(\gamma)'$ converge uniformly to $\Omega_0(\gamma, \gamma^*) = E[x_t(\gamma)x_t'(\gamma^*)]$ and $\Omega_1(\tau, \gamma) = E[f_{t-1}(F_{t-1}^{-1}(\tau))x_t(\gamma)x_t(\gamma)']$, respectively, for $\gamma(\tau) \in \mathcal{G}$ and all $\tau \in \mathcal{T}$.

A7: $det(\Omega_0(\gamma, \gamma^*)) > 0$, for all $\gamma, \gamma^* \in \mathcal{G}$.

These assumptions are common in the quantile regression (QR) and regime switching literature. A1 guarantees that the process is stationary. Depending on the specific type of model, this

assumption can be replaced by milder conditions on the values that the parameters in θ can take. A2 imposes a constraint on the value of $\gamma(.)$. A3 and A4 are common in the QR literature. A5 guarantees that for each threshold parameter value and quantile the QR problem has a unique solution. This will be needed for the case where the threshold parameter is not identified. A6 and A7 guarantee that the variance-covariance matrices in the Wald test statistics derived below exist and can be estimated for the case when the threshold parameter $\gamma(.)$ is not identified.

With these assumptions in place we are ready to derive the asymptotic distribution of the parameter estimators.

Lemma 1. *Given assumptions A1-A7, and for γ_0 known with $\tau \in \mathcal{T}$ fixed,*

$$\sqrt{T}\left(\widehat{\theta}_{\gamma_0}(\tau) - \theta_{\gamma_0}(\tau)\right) \Rightarrow N(0, \tau(1-\tau)\Sigma(\tau, \gamma_0)),$$

where $\Sigma(\tau, \gamma_0) = \Omega_1(\tau, \gamma_0)^{-1}\Omega_0(\gamma_0, \gamma_0)\Omega_1(\tau, \gamma_0)^{-1}$.

Proof. The proof of this result uses standard QR asymptotic theory when γ_0 is known. The proof is a simple extension of that in Koenker and Xiao (2006) for the asymptotic normality in quantile autoregressive models. ∎

The most interesting case is, however, when the threshold value is not known and must be estimated. For this case we need to impose the following identification condition:

A8: For all $\tau \in \mathcal{T}$, $(\theta_0(\tau), \gamma_0(\tau)) = \underset{(\theta, \gamma)}{\arg\min} E\left[\rho_\tau\left(y_t - x_t(\gamma)'\theta\right)\right]$ exists and is unique.

In this case, $\theta_0(\tau) \equiv \theta_{\gamma_0}(\tau)$. In this framework this involves estimating the following nonlinear QR model,

$$(\widehat{\theta}(\tau), \widehat{\gamma}(\tau)) = \underset{(\theta, \gamma)}{\arg\min} \sum_{t=1}^{T} \rho_\tau\left(y_t - x_t(\gamma)'\theta\right). \tag{12}$$

Regarding estimation, there are two different possible scenarios determined by the continuity or discontinuity of the threshold model. Each method implies a different methodology and properties of estimators. Thus, under OLS estimation if the threshold model is continuous on the threshold

7

variable q_t, the corresponding estimator $\widehat{\gamma}$ is a $\sqrt{T}-$consistent estimator of γ. Furthermore, Chan and Tsay (1998), still in the OLS context, show that this estimator converges in distribution to a normal distribution with mean zero and variance that depends on the correlation of $\widehat{\gamma}$ with the vector of estimators $\widehat{\theta}$. On the other hand, if the threshold model is discontinuous, the estimator $\widehat{\gamma}$ converges at a faster rate ($T-$consistent), and is therefore independent of the vector of estimators of θ. In this case two-stage estimation procedures can be used to estimate consistently the parameters of the model, see Hansen (1997, 2000) for details on the method. Moreover, Caner (2002) extends the results in Chan (1993) and Hansen (2000) to the least absolute deviation estimation of a threshold model.

Similarly, we can study the consistency of the parameter estimators of $(\theta_0(\tau), \gamma_0(\tau))$ in the QR framework. We will concentrate on the discontinuous case. Here, the two-stage estimation procedure is as follows. For fixed τ, consider a grid of γ values in the real line, and for each value estimate model (11) and save $\widehat{\theta}_\gamma(\tau)$. Then minimize,

$$\widehat{\gamma}(\tau) = \arg\min_\gamma \sum_{t=1}^T \rho_\tau\left(y_t - x_t(\gamma)'\widehat{\theta}_\gamma(\tau)\right). \tag{13}$$

Interestingly, this two-stage procedure is similar to that developed in Hansen (1996, 1997), where now we are minimizing the sum of the convex check functions $\rho_\tau(\cdot)$ instead of the sum of squared residuals. Therefore, our procedure is parallel to Hansen's procedure but in the QR framework instead of the OLS one.

The consistency of $(\widehat{\theta}(\tau), \widehat{\gamma}(\tau))$ is provided in the following lemma:

Lemma 2. *Given assumptions A1-A8, for fixed $\tau \in \mathcal{T}$,*

$$(\widehat{\theta}(\tau), \widehat{\gamma}(\tau)) = (\theta_0(\tau), \gamma_0(\tau)) + o_p(1).$$

Proof. The proof is given in the Appendix. ∎

Now, in order to derive the asymptotic distribution of the standardized parameter estimators we show that the estimation of $\gamma_0(\tau)$ has no first order effect on the asymptotic distribution of $\widehat{\theta}(\tau)$ properly standardized. The reason, is that $\widehat{\gamma}(\tau)$ is consistent for $\gamma_0(\tau)$ at a faster rate than $\widehat{\theta}(\tau)$. The following theorem and lemma formalize this fact.

Theorem 1. *Given assumptions A1-A8, and $\tau \in \mathcal{T}$ fixed,*

$$\sqrt{T}\|\widehat{\theta}(\tau) - \theta_0(\tau)\| = O_p(1) \quad \text{and} \quad T|\widehat{\gamma}(\tau) - \gamma_0(\tau)| = O_p(1).$$

Proof. The proof is given in the Appendix. ∎

From these two results and Lemma 1 it immediately follows the asymptotic normality of $\sqrt{T}(\widehat{\theta}(\tau) - \theta_0(\tau))$.

Lemma 3 *Given assumptions A1-A8, and $\tau \in \mathcal{T}$ fixed,*

$$\sqrt{T}\left(\widehat{\theta}(\tau) - \widehat{\theta}_{\gamma_0}(\tau)\right) = o_p(1).$$

Proof. The result follows from Theorem 1 as an application of the first part of Theorem 2 in Chan (1993). ∎

These results allow us to make inference about the parameters of T-QAR processes introduced here, when the threshold parameter is known, and also, and more importantly when this value is not known and needs to be estimated from the data.

In order to make feasible inference about $\theta(\tau)$ the conditional density function $f_t(\cdot)$ needs to be consistently estimated. Koenker and Xiao (2006) show that this conditional distribution can be estimated by the difference quotients,

$$\widehat{f}_{t-1}(F_{t-1}^{-1}(\tau)) = \frac{\tau_i - \tau_{i-1}}{\widehat{Q}_{y_t}(\tau_i|\Im_{t-1}) - \widehat{Q}_{y_t}(\tau_{i-1}|\Im_{t-1})}, \tag{14}$$

with the quantile process consistently estimated by $\widehat{Q}_{y_t}(\tau|\Im_{t-1}) = x_t(\gamma(\tau))'\widehat{\theta}(\tau)$, for some appropriately chosen sequence of different τ values indexed by i.

The following section makes use of these asymptotic results and introduces an innovative hypothesis test for the nonlinearity of the different threshold processes defined for each $\tau-$quantile, and of the entire quantile process indexed by $\tau \in \mathcal{T}$.

3 Nonlinearity tests

General hypotheses on the vector $\theta_\gamma(\tau)$ can be accommodated by Wald-type tests. These test statistics and associated limiting theory provide a natural foundation for the hypothesis $H_0 :$ $R\theta_{\gamma_0}(\tau) = r$, when r is known. Here R is a $m \times (2 \times (1 + p_Y))$ full rank matrix with rank m and r is an m-dimensional vector. These tests in the QR framework are introduced in Koenker and Xiao (2006) and for γ_0 known we obtain

$$\mathcal{W}_T(\tau, \gamma_0) = \frac{T(R\widehat{\theta}_{\gamma_0} - r)'[R\widehat{\Sigma}(\tau, \gamma_0)R']^{-1}(R\widehat{\theta}_{\gamma_0} - r)}{\tau(1 - \tau)}, \tag{15}$$

with $\widehat{\Sigma}(\tau, \gamma_0) = \widehat{\Omega}_1(\tau, \gamma_0)^{-1}\widehat{\Omega}_0(\gamma_0, \gamma_0)\widehat{\Omega}_1(\tau, \gamma_0)^{-1}$, and show that under H_0, and for fixed τ, $\mathcal{W}_T(\tau, \gamma_0)$ is asymptotically χ_m^2 with m-degrees of freedom.

Important tests within this class are nonlinearity tests for detecting regime switching. For example, in model (3) the selector matrix is given by $R = [1\ 0\ -1\ 0;\ 0\ 1\ 0\ -1]$ and $r = 0_m$, with 0_m a vector of zeros of dimension $m = 2$. The null hypothesis of interest is that of linearity of the model, $H_0 : R\theta_\gamma(\tau) = 0$. For the most interesting cases, γ is a nuisance parameter that is not identified under the null hypothesis. This is particularly important for the composite test involving the whole range of quantiles since it is sufficient with γ not identified for a value of τ for the asymptotic distribution of the test to be nonstandard.

Hansen (1996), following Davies (1977, 1987) and Andrews and Ploberger (1994), proposes a supremum Wald test and an average Wald test. We also choose these statistics considered now on the space $\gamma \in \mathcal{G}$ for fixed τ, and later, in order to test for nonlinearity on the entire quantile process, on the space defined by $(\tau, \gamma) \in \mathcal{T} \times \mathcal{G}$. Hansen (1996), in a OLS context and for hypothesis tests only for the mean, base his test statistics on the score function. In parallel, we use the Bahadur representation of the QR model (see for instance Koenker, 2005, p.122):

$$\sqrt{T}\left(\hat{\theta}_\gamma(\tau) - \theta_\gamma(\tau)\right) = \Omega_1^{-1}(\tau, \gamma)S_T(\tau, \gamma) + o_p(1), \tag{16}$$

where $S_T(\tau, \gamma) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t(\gamma)\psi_\tau(y_t - F_{t-1}^{-1}(\tau))$ is the score function and $\psi_\tau(u) = \tau - I(u < 0)$ is the influence function in the quantile regression models.

The next theorem derives the asymptotic distribution of the bivariate process (16). This theorem is instrumental for the derivation of the asymptotic distribution of the nonlinearity tests later introduced.

Theorem 2. *Given assumptions A1-A7,*

$$\sqrt{T}\left(\hat{\theta}_\gamma(\tau) - \theta_\gamma(\tau)\right) \Rightarrow B(\tau, \gamma), \tag{17}$$

with $B(\tau, \gamma)$ a bivariate Gaussian process with mean zero and covariance kernel defined by

$$K((\tau_i, \gamma_i), (\tau_j, \gamma_j)) = E\left(B(\tau_i, \gamma_i)B(\tau_j, \gamma_j)\right) = (\tau_i \wedge \tau_j - \tau_i\tau_j)\,\Omega_1(\tau_i, \gamma_i)^{-1}\Omega_0(\gamma_i, \gamma_j)\Omega_1(\tau_j, \gamma_j)^{-1},$$

with $\tau_i, \tau_j \in \mathcal{T}$ and $\gamma_i, \gamma_j \in \mathcal{G}$.

Proof. The proof is given in the Appendix. ∎

Corollary 1. *Given assumptions A1-A7,*

- *For $\tau \in \mathcal{T}$ fixed and γ varying in $\mathcal{G}$,*

$$\sqrt{T}\left(\hat{\theta}_\gamma(\tau) - \theta_\gamma(\tau)\right) \Rightarrow B_\tau(\gamma), \tag{18}$$

  with $B_\tau(\gamma)$ a univariate Gaussian process with mean zero and covariance kernel defined by,

$$K_\tau(\gamma_i, \gamma_j) = \tau(1-\tau)\Omega_1(\tau, \gamma_i)^{-1}\Omega_0(\gamma_i, \gamma_j)\Omega_1(\tau, \gamma_j)^{-1}, \tag{19}$$

  with $\gamma_i, \gamma_j \in \mathcal{G}$.

- *For $\gamma \in \mathcal{G}$ fixed and τ varying in $\mathcal{T}$,*

$$\sqrt{T}\left(\hat{\theta}_\gamma(\tau) - \theta_\gamma(\tau)\right) \Rightarrow B_\gamma(\tau), \tag{20}$$

  with $B_\gamma(\tau)$ a univariate Gaussian process with mean zero and covariance kernel defined by,

$$K_\gamma(\tau_i, \tau_j) = (\tau_i \wedge \tau_j - \tau_i\tau_j)\,\Omega_1(\tau_i, \gamma)^{-1}\Omega_0(\gamma, \gamma)\Omega_1(\tau_j, \gamma)^{-1},$$

  where $\tau_i, \tau_j \in \mathcal{T}$.

Proof. The proof follows trivially from fixing one or other parameter in process (16), and from applying Theorem 2. ∎

The composite hypothesis of linearity of the model is now $H_0 : R\theta_\gamma(\tau) = 0$ for all $\gamma \in \mathcal{G}$, and $\tau \in \mathcal{T}$. However, if one is only interested in testing for linearity for one fixed quantile τ_0 the relevant test statistics are

$$\sup_{\gamma \in \mathcal{G}} \mathcal{W}_T^{(1)}(\tau_0, \gamma), \tag{21}$$

11

and

$$\underset{\gamma \in \mathcal{G}}{\text{ave}} \, \mathcal{W}_T^{(1)}(\tau_0, \gamma), \tag{22}$$

with

$$\mathcal{W}_T^{(1)}(\tau_0, \gamma) = T(R\widehat{\theta}_\gamma(\tau_0) - r)'[R\widehat{K}_{\tau_0}(\gamma, \gamma)R']^{-1}(R\widehat{\theta}_\gamma(\tau_0) - r), \tag{23}$$

and where $\widehat{K}_{\tau_0}(\gamma, \gamma) = \tau_0(1 - \tau_0)\widehat{\Omega}_1(\tau_0, \gamma)^{-1}\widehat{\Omega}_0(\gamma, \gamma)\widehat{\Omega}_1(\tau_0, \gamma)^{-1}$.

Theorem 3. *Given assumptions A1-A7, $\tau_0 \in \mathcal{T}$ fixed, and under the null hypothesis of linearity $H_0 : R\theta_\gamma(\tau_0) = 0$,*

$$\underset{\gamma \in \mathcal{G}}{\sup} \, \mathcal{W}_T^{(1)}(\tau_0, \gamma) \Rightarrow \underset{\gamma \in \mathcal{G}}{\sup} \mathcal{W}_0^{(1)}(\tau_0, \gamma), \tag{24}$$

$$\underset{\gamma \in \mathcal{G}}{\text{ave}} \mathcal{W}_T^{(1)}(\tau_0, \gamma) \Rightarrow \underset{\gamma \in \mathcal{G}}{\text{ave}} \mathcal{W}_0^{(1)}(\tau_0, \gamma), \tag{25}$$

with $\mathcal{W}_0^{(1)}$ a process defined by

$$\mathcal{W}_0^{(1)} = E[S_T(\tau_0, \gamma)]'\Omega_1^{-1}(\tau_0, \gamma)'R'[RK_{\tau_0}(\gamma, \gamma)R']^{-1}R\Omega_1^{-1}(\tau_0, \gamma)E[S_T(\tau_0, \gamma)].$$

Proof. The proof of this theorem follows from Corollary 1 and the continuous mapping theorem. ∎

For testing for linearity for the entire $\tau-$quantile process the test statistic that we propose is of Kolmogorov-Smirnov type. Let the relevant test statistics be defined as

$$KS \, sup\mathcal{W} = \underset{\tau \in \mathcal{T}}{\sup} \underset{\gamma \in \mathcal{G}}{\sup} \, \mathcal{W}_T^{(2)}(\tau, \gamma) \tag{26}$$

and

$$KS \, ave\mathcal{W} = \underset{\tau \in \mathcal{T}}{\sup} \underset{\gamma \in \mathcal{G}}{\text{ave}} \mathcal{W}_T^{(2)}(\tau, \gamma), \tag{27}$$

with

$$\mathcal{W}_T^{(2)}(\tau, \gamma) = T(R\widehat{\theta}_\gamma(\tau) - r)'[R\widehat{K}((\tau, \gamma), (\tau, \gamma))R']^{-1}(R\widehat{\theta}_\gamma(\tau) - r), \tag{28}$$

and where $\widehat{K}((\tau, \gamma), (\tau, \gamma)) = \tau(1 - \tau)\widehat{\Omega}_1(\tau, \gamma)^{-1}\widehat{\Omega}_0(\gamma, \gamma)\widehat{\Omega}_1(\tau, \gamma)^{-1}$.

Theorem 4. *Given assumptions A1-A7, and under the null hypothesis of linearity $H_0 : R\theta_\gamma(\tau) = 0$, for all $\gamma \in \mathcal{G}$, and $\tau \in \mathcal{T}$,*

$$KS \, sup\mathcal{W} \Rightarrow \underset{\tau \in \mathcal{T}}{\sup} \underset{\gamma \in \mathcal{G}}{\sup} \, \mathcal{W}_0^{(2)}(\tau, \gamma) \tag{29}$$

12

and

$$KS\ aveW \Rightarrow \sup_{\tau \in \mathcal{T}} ave_{\gamma \in \mathcal{G}} \mathcal{W}_0^{(2)}(\tau, \gamma), \qquad (30)$$

with $\mathcal{W}_0^{(2)}$ a process defined by

$$\mathcal{W}_0^{(2)} = E[S_T(\tau, \gamma)]' \Omega_1^{-1}(\tau, \gamma)' R' [RK((\tau, \gamma), (\tau, \gamma))R']^{-1} R\Omega_1^{-1}(\tau, \gamma) E[S_T(\tau, \gamma)].$$

Proof. The proof of this theorem follows from Theorem 2 and the continuous mapping theorem.
∎

Since the asymptotic null distributions of these Wald type tests are highly nonstandard, depend upon the covariance functions $K_{\tau_0}(\tau, \gamma)$ and $K((\tau, \gamma), (\tau, \gamma))$, critical values are data-dependent and in turn cannot be tabulated. We follow the solution proposed in Hansen (1996) and approximate the p-value of the asymptotic null distributions by an alternative p-value transformation adapted to QR. In particular, we are interested in approximating the p-value for the supremum and average tests for τ_0 fixed. The procedure is detailed as follows.

Given a sample of observations $\{y_t, x_t\}$, $t = 1, \ldots, T$, denote

$$\widehat{S}_T(\tau_0, \gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t(\gamma) \psi_{\tau_0}(y_t - \widehat{F}_{t-1}^{-1}(\tau_0)),$$

where $\widehat{F}_{t-1}(\cdot)$ is an estimator of the corresponding conditional distribution function $F_{t-1}(\cdot)$. The asymptotic p-value of the supremum and average of the $\mathcal{W}_0^{(1)}(\tau_0, \gamma)$ process can be approximated by generating independent replicas of the test statistic $\mathcal{W}_T^{(1)}(\tau_0, \gamma)$, for fixed τ_0 and taking the supremum over γ. Now, using the multiplier central limit theorem in van der Vaart and Wellner (1996, p.176) we have that

$$\widehat{S}_T^*(\tau_0, \gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t(\gamma) \psi_{\tau_0}(y_t - \widehat{F}_{t-1}^{-1}(\tau_0)) v_t,$$

with $\{v_t\}_{t=1}^{T}$ an *iid* sequence of $N(0, 1)$ random variables, also satisfies (16). By the continuous mapping theorem and the Bahadur representation discussed above, the corresponding replica

$$\mathcal{W}_T^{(1)*}(\tau_0, \gamma) = \left(R\Omega_1^{-1}(\tau_0, \gamma) S_T^*(\tau_0, \gamma) \right)' \left(RK_{\tau_0}(\gamma, \gamma)R' \right)^{-1} \left(R\Omega_1^{-1}(\tau_0, \gamma) S_T^*(\tau_0, \gamma) \right), \qquad (31)$$

converges to independent and identical replicas of the asymptotic process in Theorem 2.

13

4 Monte Carlo experiments

4.1 Estimating γ

Consider a baseline linear location-scale two-regime switching SETAR process, as a particular T-QAR model,

$$y_t = \begin{cases} \theta_{01} + \theta_{11}y_{t-1} + \delta y_{t-1}u_t, & y_{t-1} \leq \gamma_0, \\ \theta_{02} + \theta_{12}y_{t-1} - \delta y_{t-1}u_t, & y_{t-1} > \gamma_0. \end{cases} \tag{32}$$

We construct a Monte Carlo experiment with $\theta_{01} = \theta_{02} = 0.5, \theta_{11} = \theta_{12} = 0.5, \delta = 1, \gamma_0 = 1, u \sim iid\ N(0,1)$, $T = 500$, where T is the sample size, and $R = 500$ where R is the number of repeated experiments. Figure 1 plots the root mean square error (RMSE) of estimating γ_0. For the QR estimates we consider $\tau \in [0.1, 0.9]$.

Let $\hat{\theta}_\gamma(\tau)$ denote the QR estimate for fixed τ and γ, and $\bar{\theta}_\gamma$ be the OLS estimate for fixed γ. The following estimators of γ_0 are considered:

$$Solid\ line: \widehat{\gamma}(\tau) = \arg\min_\gamma \sum_{t=1}^{T} \rho_\tau \left(y_t - x_t(\gamma)'\widehat{\theta}_\gamma(\tau) \right)$$

$$Dashed\ line: \bar{\gamma}(\tau) = \arg\min_\gamma \sum_{t=1}^{T} \left(y_t - x_t(\gamma)'\widehat{\theta}_\gamma(\tau) \right)^2$$

$$Dotted\ line: \bar{\gamma} = \arg\min_\gamma \sum_{t=1}^{T} \left(y_t - x_t(\gamma)'\bar{\theta}_\gamma \right)^2$$

Here the regressor variable $x_t(\gamma)$ is defined as in Section 2. $\widehat{\gamma}(\tau)$ is our preferred estimator, while $\bar{\gamma}$ is the least squares SETAR estimator for the threshold parameter, and $\bar{\gamma}(\tau)$ is a hybrid estimator that runs QR for each pair (τ, γ) but minimizes the sum of square errors as in the least squares framework.

The figure shows that $\widehat{\gamma}(\tau)$ has the smallest RMSE, which decreases as $|\tau - 0.5|$ increases[2]. This gives evidence that least squares SETAR models cannot identify the presence of two regimes that coincide in the mean process. Moreover, this generalizes to the case where a symmetric loss function such as the square loss function is used to estimate γ (i.e. as used in $\bar{\gamma}$ and $\bar{\gamma}(\tau)$).

---

[2]Note that for this case $F_u^{-1}(0.5) = 0$ and therefore, $Q_{y_t}(0.5|\Im_{t-1}, y_{t-1} \leq 1) = Q_{y_t}(0.5|\Im_{t-1}, y_{t-1} > 1)$.

Figure 1: RMSE in estimating gamma
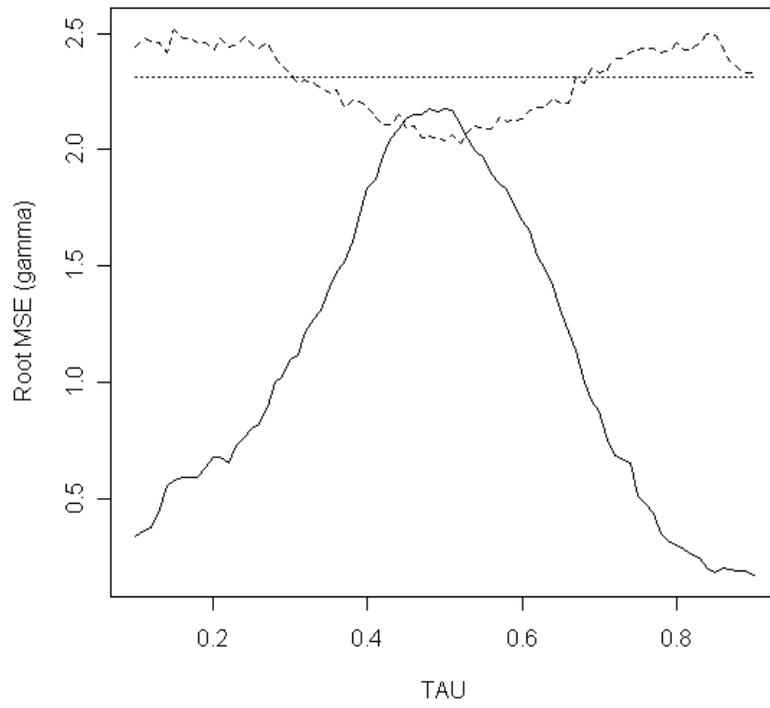
4.2 Nonlinearity tests

Now consider a test for linearity of the two-regime baseline model. If we were modeling the conditional mean this test would involve $H_0 : \theta_{01} = \theta_{02}, \theta_{11} = \theta_{12}$. An extension of the test for conditional quantile models could be $H_0 : \theta_{01}(\tau_0) = \theta_{02}(\tau_0), \theta_{11}(\tau_0) = \theta_{12}(\tau_0)$, for fixed $\tau_0 \in \mathcal{T}$ or $H_0 : \theta_{01}(\tau) = \theta_{02}(\tau), \theta_{11}(\tau) = \theta_{12}(\tau), \forall \tau \in \mathcal{T}$. In this experiment we assume that γ is not known and approximate by simulation the distribution of the Wald tests to obtain the p-value transformation. We calculate the p-value using 1000 random draws and consider a sample size of $T = 500$. The experiments are based on $R = 500$ repeated experiments. In all cases we consider theoretical sizes of 5% and 10%.

To investigate the empirical size of the tests we consider a standard AR(1) model $y_t = \theta_0 + \theta_1 y_{t-1} + u_t$, with $\theta_0 = 0.5$, $\theta_1 = 0.5$ and $u \sim iid\ N(0,1)$, with simulation results reported in table 1. We first compute the heteroskedasticity-robust Wald test for least-squares described in Hansen (1996). As previously found in the literature, the supremum Wald test ($sup\mathcal{W}_{LS}$) approach shows considerable over-rejection. These results are similar to Hansen (1996) or Martinez and Olmo (2008) that showed similar problems of this test in related nonlinear contexts. Moreover, the corresponding average Wald test ($ave\mathcal{W}_{LS}$) reports better size properties, although even in this case, rejection rates are above the theoretical size. Over-rejection is also observed for the extreme quantiles (0.1 and 0.9) in QR based Wald tests ($sup\mathcal{W}_{QR}$ and $ave\mathcal{W}_{QR}$). The results support the evidence that the average Wald test performs better that the supremum, also in the QR context. In this method, the rejection rates are closer to the theoretical size as the quantile moves towards the median. Overall, the tests developed here show encouraging signs to provide a better alternative to the least-squares based Hansen statistics. Kolmogorov-Smirnov tests (KS $sup\mathcal{W}_{QR}$ and KS $ave\mathcal{W}_{QR}$) are mainly driven by the extreme quantiles, and as a result, they have high rejection rates.

In order to evaluate these tests in terms of power we calculate the empirical size for the two-regime baseline model (32). The results for this experiment are reported in table 2. As expected, it can be observed that Hansen tests show little power against the null hypothesis of linearity, because the mean process is the same in both regimes. Rejection rates are only slightly above those in the size experiments, and therefore, it can be concluded that they do not identify the presence of different regimes. Moreover, QR based tests have good power properties, which increase towards the extremes. In this case, tests for $\tau = 0.5$ should not reject, but this power may be due to the effect of contiguous quantiles. As before, Kolmogorov-Smirnov tests have empirical rejection rates

Table 1: Size - Gaussian AR(1) process

τ	Theoretical size			
	0.05	0.10	0.05	0.10
	$sup\mathcal{W}_{LS}$		$ave\mathcal{W}_{LS}$	
	0.200	0.304	0.095	0.147
	$sup\mathcal{W}_{QR}$		$ave\mathcal{W}_{QR}$	
0.10	0.304	0.408	0.182	0.278
0.25	0.078	0.121	0.026	0.069
0.50	0.008	0.017	0.017	0.069
0.75	0.052	0.086	0.060	0.113
0.9	0.200	0.304	0.156	0.208
	KS $sup\mathcal{W}_{QR}$		KS $ave\mathcal{W}_{QR}$	
	0.252	0.400	0.174	0.234

similar to those in the extreme quantiles.

Overall, the average Wald statistics for non-extreme quantiles have the best performance in terms of size and power.

Table 2: Power

τ	Theoretical size			
	0.05	0.10	0.05	0.10
	$sup\mathcal{W}_{LS}$		$ave\mathcal{W}_{LS}$	
	0.321	0.379	0.264	0.367
	$sup\mathcal{W}_{QR}$		$ave\mathcal{W}_{QR}$	
0.10	1.000	1.000	1.000	1.000
0.25	0.988	0.988	0.988	0.988
0.50	0.632	0.643	0.505	0.574
0.75	1.000	1.000	1.000	1.000
0.9	1.000	1.000	1.000	1.000
	KS $sup\mathcal{W}_{QR}$		KS $ave\mathcal{W}_{QR}$	
	1.000	1.000	1.000	1.000

5   Empirical Application

Most macroeconomic series are affected by booms and bursts in economic activity that produce in turn expansionary and contractionary business cycles. It is well known in empirical macroeconometrics that the dynamics of these series have an asymmetric behavior which extent depends on the nature of the macroeconomic series studied. Examples can be found in Beaudry and Koop (1993) that study the effects of negative and positive shocks in US GDP; Sarno, Valente, and Leon

(2006) that use regime switching processes to model asymmetries in exchange rates, and Koenker and Xiao (2006) that study the asymmetries in US unemployment rates with quantile autoregressive models. In particular these authors conclude that there exists heterogeneity in this series that cannot be simply reflected with processes modeling the mean and estimated with ordinary least squares methods.

Our analysis extends this study by contemplating the possibility of nonlinearities in the quantile process. In this way we postulate that the heterogeneity found in series on unemployment growth, *i.e.* growth in the number of people unemployed, can be due to asymmetric dynamics in the quantile process that, for some quantiles of the process, can potentially depend on different manners on previous values of the variable and lead us, therefore, to propose different TAR processes for different quantiles. The choice of this series is due to the clear stationary character, in contrast to US unemployment rates series. Moreover, this is an important variable from the policymaker perspective, as it clearly anticipates periods of social calm or distress. Figure 2 reports the monthly unemployment growth series from February 1948 to June 2007.

A preliminary exploratory analysis of this data inclines us to choose autoregressive processes of order one, which allow us to compare with Monte Carlo experiments above. The nonlinear model that we explore is

$$Q_{y_t}(\tau|y_{t-1}) = \begin{cases} \theta_{01} + \theta_{11}y_{t-1}, & y_{t-1} \leq \gamma(\tau), \\ \theta_{02} + \theta_{12}y_{t-1}, & y_{t-1} > \gamma(\tau). \end{cases}$$

The ordinary least squares estimates for this model appear in table 3. The Hansen's (1996) supremum and average Wald-test statistics cannot reject the linearity of the mean process at the 15% significance level. Therefore the data suggest that the mean process follows a linear process with autoregressive parameter 0.116 (standard error 0.037).

Table 3: OLS estimates

Regime 1 $y_{t-1} \leq \gamma$		Regime 2 $y_{t-1} > \gamma$		threshold	Wald-test (sup)	Wald-test (ave)
θ_{01}	θ_{11}	θ_{02}	θ_{12}	γ	p-value	p-value
0.723	0.119	−0.278	0.242	-1.348	0.267	0.148
(0.423)	(0.110)	(0.235)	(0.113)			

Note: Standard deviations in parenthesis.

Alternatively, the estimates from the QR method are reported in table 4 and figure 3. The results from this table are illuminating. As the quantile increases towards unity we find statistical evidence

Figure 2: Unemployment growth series (in %)
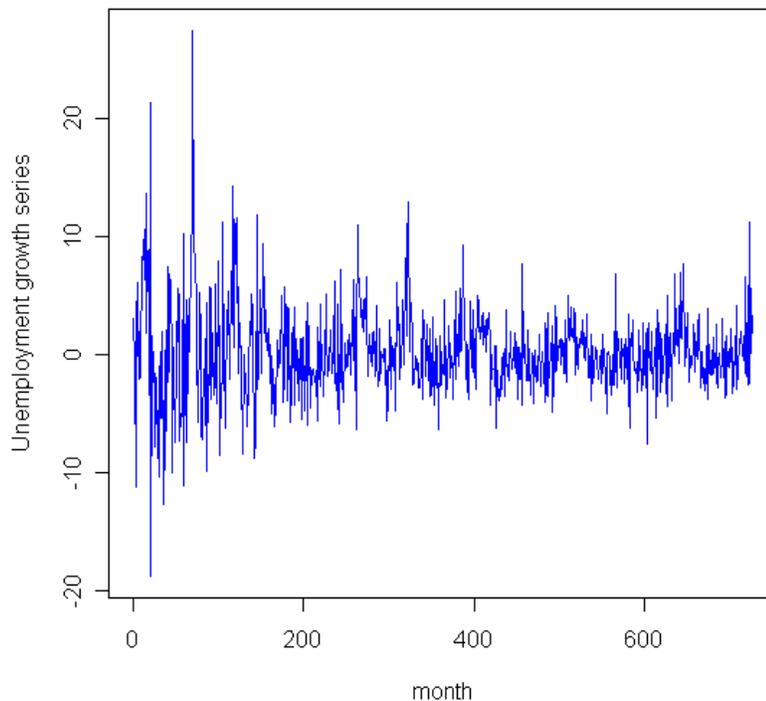
Notes: Monthly unemployment growth for US spanning the period February 1948 to June 2007.
Source: freelunch.com

Table 4: QR estimates

	Regime 1 $y_{t-1} \leq \gamma(\tau)$		Regime 2 $y_{t-1} > \gamma(\tau)$		threshold	Wald-test (sup)	Wald-test (ave)
τ	$\theta_{01}(\tau)$	$\theta_{11}(\tau)$	$\theta_{02}(\tau)$	$\theta_{12}(\tau)$	$\gamma(\tau)$	p-value	p-value
0.10	-0.624 (0.775)	0.595 (0.237)	-3.942 (0.259)	0.025 (0.105)	-1.983	0.019	0.189
0.25	-1.327 (0.166)	-0.0097 (0.064)	-4.147 (1.697)	0.380 (0.238)	3.915	0.622	0.344
0.50	0.944 (0.963)	0.034 (0.158)	0.210 (0.179)	0.196 (0.067)	-2.897	0.041	0.006
0.75	1.928 (0.235)	0.019 (0.073)	-0.587 (1.161)	0.770 (0.268)	1.969	0.028	0.002
0.9	4.529 (0.373)	0.087 (0.088)	1.053 (1.327)	0.876 (0.279)	1.969	0.027	0.003
Kolmogorov-Smirnov						0.040	0.008

Note: Standard deviations in parenthesis.

Figure 3: Parameter estimates: OLS and QR
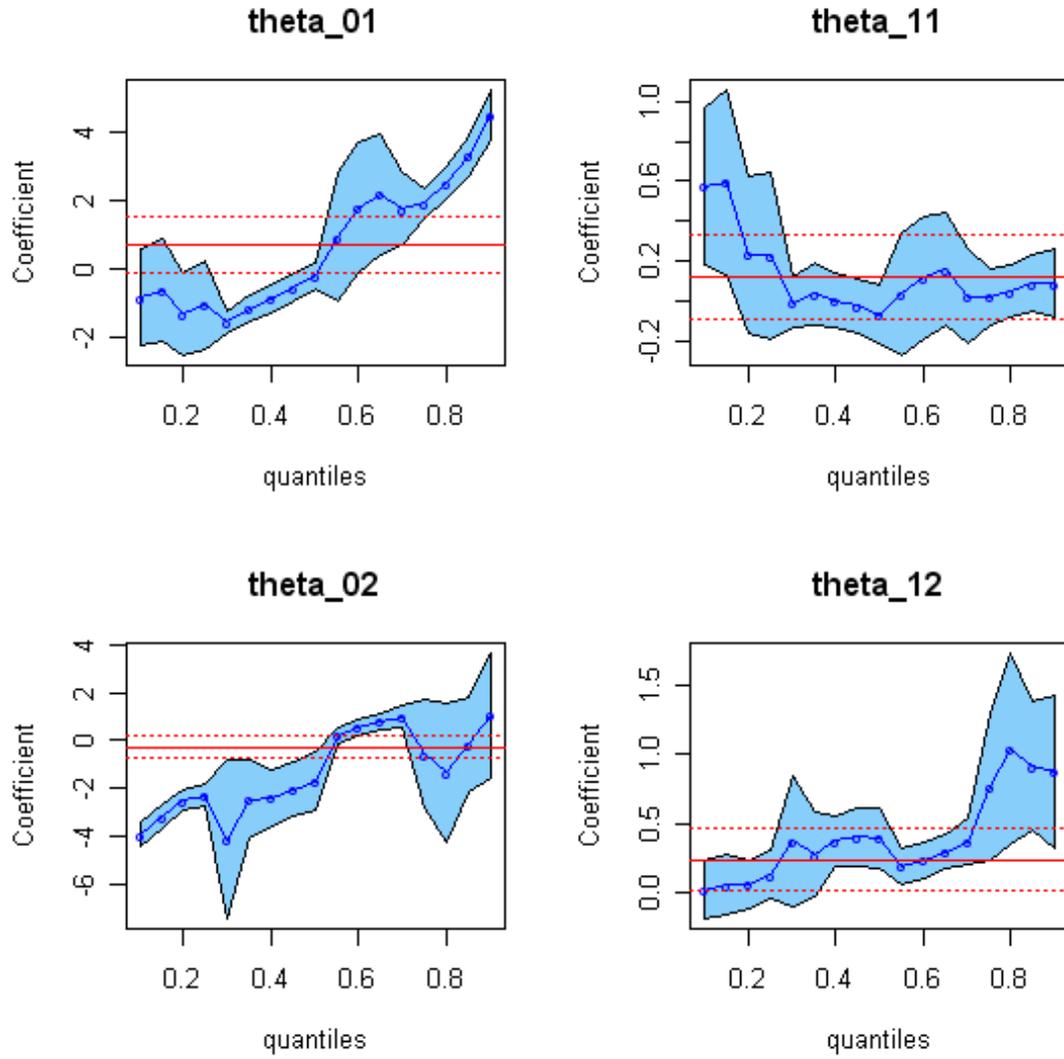
to reject the autoregressive linear specification above. In particular we observe two regimes in the autoregressive parameter for $\tau \geq 0.50$. In these cases, the autoregressive parameter for the regime determined by $y_{t-1} \leq \gamma(\tau)$ is not statistically significant. On the other hand, the autoregressive parameters corresponding to the upper regime (*i.e.* $y_{t-1} > \gamma(\tau)$) are strongly significant and monotonically increasing as τ goes to 1.

Overall, we can conclude that in calm periods of the economy characterized by levels of unemployment growth below 2% unemployment growth is serially independent, pointing out that increases and declines in the level of unemployment have no persistence and are due to instantaneous shocks affecting the economy. On the contrary, when unemployment growth reaches extreme levels, defined by increments of roughly 2% monthly, we conclude that the higher quantiles of the response variable have quite a strong persistence and depend heavily on the level of unemployment growth in the previous period. This is particularly remarkable for $\tau \geq 0.75$. For moderately low quantiles, on the other hand, we observe no persistence of the variable and unemployment growth can be considered driven by serially independent shocks.

6   Conclusions

This paper provides useful insights on estimation procedures to identify threshold parameters in threshold autoregressive models. Moreover, it shows that the threshold quantile autoregressive (T-QAR) framework has similar asymptotic properties to those found in Chan (1993) study for OLS based models. Simulation results show that the gains in root mean squared error are considerable when QR methods are used.

This paper also shows the insensitivity of Hansen (1996, 1997) Wald tests to reject the linearity hypothesis when the nonlinearity appears in the quantile process but it is absent in the mean process. To solve this, we propose an alternative Wald test based on QR methods that contains several advantages over Hansen's device. First, it detects nonlinearities in asymmetric processes when the conditional quantile behavior varies but the mean (and variance) process are the same. Second, it has better empirical size properties in small samples, except for extreme quantiles, than the OLS based tests.

Finally, in the empirical application we extend the study in Koenker and Xiao (2006) to the macroeconomic series of unemployment growth. In our study we find statistical evidence of nonlinearity in the higher quantiles of the quantile process that the standard linearity test of Hansen fails to detect. We conclude from our experiment that the higher quantiles of the conditional distribu-

tion of the dynamics in unemployment growth exhibit an strong persistence when the economy is in distress periods, that can be well described by processes of autoregressive type. The conditional median of this variable seems, however, not to be affected by the growth in unemployment produced in previous periods.

These models have several other potential applications, which are left for future research, but that we believe highlight the usefulness of this methodology. For instance, following CAPM methodologies we may measure the risk premium (in probability) of a risky asset in terms of the risk premium of the market portfolio ($E[R_M] - R_f$), and see whether this depends on the state of the market (booming or falling markets). Another potential application of T-QAR models is in exchange rates. These time series are usually modeled by nonlinear processes that assume the presence of some fluctuation bands. Once these bands are exceeded the market forces (arbitrage conditions) make the series to be mean-reverting. It can be interesting to analyze the quantile processes corresponding to these models, and see whether these bands are different across τ, and their economic implications.

7 Appendix

First, we prove Lemma 2. The approach to prove consistency will be to use the argmax theorem, Corollary 3.2.3 van der Vaart and Wellner (1996).

Proof. We first need to establish that $M_T \Rightarrow M$ (where $\Rightarrow$ means weak convergence) in $l^\infty(B)$ for all compact set $B \subset \Theta \times \mathcal{G}$, where $M(\theta) \equiv \mathbb{P}_T m_\theta$, $\mathbb{P}_T = T^{-1} \sum_{i=1}^T \delta_{X_i}$, and δ_x assigns mass 1 at x and zero elsewhere. Then, the argmax theorem will yield the consistency.

For simplicity, define α and β as the subsets of parameters in θ that correspond to the regimes $q_{t-1} \le \gamma$ and $q_{t-1} > \gamma$ respectively, where the dependence on τ is omitted. We will use the notation $x_t = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p_y})$. For the particular SETAR(1) case, $q_{t-1} = y_{t-1}$.

Fix a compact set B. We now verify that $\Im_B \equiv \{m_\theta : \theta \in B\}$ is Glivenko-Cantelli. Note that

$$m_\theta = \rho_\tau(\xi(\tau) - x_t'\alpha + x_t'\alpha_0)1(q_{t-1} < \gamma \wedge \gamma_0) + \rho_\tau(\xi(\tau) - x_t'\beta + x_t'\alpha_0)1(\gamma < q_{t-1} \le \gamma_0)$$
$$+ \rho_\tau(\xi(\tau) - x_t'\alpha + x_t'\beta_0)1(\gamma_0 < q_{t-1} \le \gamma) + \rho_\tau(\xi(\tau) - x_t'\beta + x_t'\beta_0)1(q_{t-1} > \gamma \vee \gamma_0).$$

In addition, $\{\rho_\tau(\xi(\tau) - x_t'\alpha - x_t'\alpha_0) : \theta \in B\}$ and $1\{q_{t-1} < \gamma \wedge \gamma_0 : \theta \in B\}$ are separately Glivenko-Cantelli classes. For the first part, let

$$\mathscr{P} = \{g_T(\theta, \gamma) : (\theta, \gamma) \in \Theta \times \mathcal{G}\},$$

where $g_T \equiv \rho_\tau(y_t - x_t(\gamma)'\theta)$. Note that by assumption A2 $\Theta \times \mathcal{G}$ is compact, $\mathscr{P}$ is continuous and uniformly Lipschitz over $\Theta \times \mathcal{G}$. Therefore by Lemma 3.10 in van de Geer (2000) we have that $H_{1,B}(\delta, \mathscr{P}, P) < \infty$, that is, the δ-entropy with bracketing of $\mathscr{P}$ is finite. Hence, it satisfies a uniform law of large numbers. The second part follows directly from assumption A2 and monotonicity of the indicator function (Theorem 2.7.5 in van der Vaart and Wellner (1996)).

Thus, by corollary 9.26 in Kosorok (2008), the product of the components is integrable, the product of the two classes is also Glivenko-Cantelli. Similar arguments reveal that the remaining terms of the sum are also Glivenko-Cantelli and the same theorem yields that $\Im_B$ is Glivenko-Cantelli. Thus $M_T \Rightarrow M$ in $l^\infty(B)$ for all compact B. Hence, $(\hat{\theta}(\tau), \hat{\gamma}(\tau)) \xrightarrow{p} (\theta_0(\tau), \gamma_0(\tau))$. ∎

Now we prove Theorem 1.

Proof.

We want to show that $\sqrt{T}\|\hat{\theta}(\tau) - \theta_0(\tau)\| = O_p(1)$ and $T|\hat{\gamma}(\tau) - \gamma_0(\tau)| = O_p(1)$. For simplicity, we omit the dependence on τ and denote $\hat{\theta}(\tau) = \hat{\theta}_T$ and $\hat{\gamma}(\tau) = \hat{\gamma}_T$.

Theorem 14.4 in Kosorok (2008) (Rate of convergence) Let M_T be a sequence of stochastic processes indexed by a semimetric space (Θ, d) and $M : \Theta \to \mathbb{R}$ a deterministic function such that for every θ in a neighborhood of θ_0, there exists a $c_1 > 0$ such that

$$M(\theta) - M(\theta_0) \leq -c_1 \tilde{d}^2(\theta, \theta_0),$$

where $\tilde{d} : \Theta \times \Theta \mapsto [0, \infty)$ satisfies $\tilde{d}(\theta_T, \theta) \to 0$ as $d(\theta_T, \theta) \to 0$. Suppose that for all n large enough and sufficiently small δ, the centered process $M_T - M$ satisfies

$$E^* \sup_{\tilde{d}(\theta_T, \theta) < \delta} \sqrt{T} |(M_T - M)(\theta) - (M_T - M)(\theta_0)| \leq c_2 \phi_T(\delta),$$

for $c_2 < \infty$ and functions ϕ_T such that $\delta \mapsto \phi_T(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$ not depending on n. Let

$$r_T^2 \phi(r_T^{-1}) \leq c_3 \sqrt{T}$$

for every n and some $c_3 < \infty$. If the sequence $\hat{\theta}_T$ satisfies $M_T(\hat{\theta}_T) \geq \sup_{\theta \in \Theta} M_T(\theta) - O_p(r_T^{-2})$ and converges in outer probability to θ_0, then $r_T \tilde{d}(\hat{\theta}_T, \theta_0) = O_p(1)$.

We will use Theorem 14.4 in Kosorok (2008) to obtain the convergence rates via the discrepancy function $\tilde{d}(\theta, \theta_0) \equiv \|\theta - \theta_0\| + \sqrt{|\gamma - \gamma_0|}$. Note that this is not a norm since it does not satisfy the triangle inequality. Nevertheless, $\tilde{d}(\theta, \theta_0) \to 0$ if and only if $\|\theta - \theta_0\| \to 0$. Moreover, we have that

$$
\begin{aligned}
M(\theta) - M(\theta_0) &= P\{q_{t-1} < \gamma \wedge \gamma_0\} \left[\rho_\tau(\xi(\tau) - x_t'\alpha + x_t'\alpha_0) - \rho_\tau(\xi(\tau)) \right] \\
&\quad + P\{\gamma < q_{t-1} \leq \gamma_0\}\} \left[\rho_\tau(\xi(\tau) - x_t'\beta + x_t'\alpha_0) - \rho_\tau(\xi(\tau)) \right] \\
&\quad + P\{\gamma_0 < q_{t-1} \leq \gamma\}\} \left[\rho_\tau(\xi(\tau) - x_t'\alpha + x_t'\beta_0) - \rho_\tau(\xi(\tau)) \right] \\
&\quad + P\{q_{t-1} > \gamma \vee \gamma_0\} \left[\rho_\tau(\xi(\tau) - x_t'\beta + x_t'\beta_0) - \rho_\tau(\xi(\tau)) \right] \\
&\leq P\{q_{t-1} < a\}|x_t'(\alpha - \alpha_0)| + P\{q_{t-1} > b\}|x_t'(\beta - \beta_0)| + k_1(1 + o(1))\sqrt{|\gamma - \gamma_0|} \\
&\leq P\{q_{t-1} < a\}\|x_t\|\|\alpha - \alpha_0\| + P\{q_{t-1} > b\}\|x_t\|\|\beta - \beta_0\| + k_1(1 + o(1))\sqrt{|\gamma - \gamma_0|} \\
&\leq (k_1 \wedge \delta_1 - o(1))\tilde{d}(\theta, \theta_0)
\end{aligned}
$$

where the first inequality follows from the fact that that $\rho_\tau(x + y) - \rho_\tau(x) \leq 2|y|$, and the product of the density of q_{t-1} and $|\alpha_0 - \beta_0|$ is bounded below by some $k_1 > 0$. The second inequality follows from Cauchy-Schwarz inequality. The third inequality follows from both $P(q_{t-1} < a)$ and $P(q_{t-1} > b)$ being bounded below by some $\delta_1 > 0$ and A3 (using A3 $\|x_t\| = o_p(1)$). Thus $M(\theta) - M(\theta_0) \leq \tilde{d}(\theta, \theta_0)$ for all $\|\theta - \theta_0\|$ small enough, as desired.

Consider now the class of functions $\mathcal{M}_\delta \equiv \{m_\theta - m_{\theta_0} : \tilde{d}(\theta, \theta_0) < \delta\}$. Using previous calculations, we have

$$\begin{aligned}
m_\theta - m_{\theta_0} &= \left[\rho_\tau(\xi(\tau) - x_t'\alpha + x_t'\alpha_0) - \rho_\tau(\xi(\tau)) \right] 1\{q_{t-1} < \gamma \wedge \gamma_0\} \\
&+ \left[\rho_\tau(\xi(\tau) - x_t'\beta + x_t'\alpha_0) - \rho_\tau(\xi(\tau)) \right] 1\{\gamma < q_{t-1} \le \gamma_0\} \\
&+ \left[\rho_\tau(\xi(\tau) - x_t'\alpha + x_t'\beta_0) - \rho_\tau(\xi(\tau)) \right] 1\{\gamma_0 < q_{t-1} \le \gamma\} \\
&+ \left[\rho_\tau(\xi(\tau) - x_t'\beta + x_t'\beta_0) - \rho_\tau(\xi(\tau)) \right] 1\{q_{t-1} > \gamma \vee \gamma_0\} \\
&\le \|x_t\|\|\alpha - \alpha_0\| 1\{q_{t-1} < \gamma \wedge \gamma_0\} + \|x_t\|\|\beta - \alpha_0\| 1\{\gamma < q_{t-1} \le \gamma_0\} \\
&+ \|x_t\|\|\alpha - \beta_0\| 1\{\gamma_0 < q_{t-1} \le \gamma\} + \|x_t\|\|\beta - \beta_0\| 1\{q_{t-1} > \gamma \vee \gamma_0\} \\
&\equiv A_1(\theta) + A_2(\theta) + A_3(\theta) + A_4(\theta).
\end{aligned}$$

Consider first $A_1(\theta)$. Since $\{1\{q_{t-1} \le t\} : t \in [a,b]\}$ is in the class of indicator functions, it is a VC class. So, as a consequence of Lemma 8.17 in Kosorok (2008), it is possible to compute

$$E^* \sup_{\tilde{d}(\theta, \theta_0) < \delta} |\mathbb{G}_T A_1(\theta)| \le \delta$$

where $\mathbb{G}_T = \sqrt{T}(\mathbb{P}_T - P)$, and $\mathbb{P}_T m_\theta = M_T(\theta)$. Similar calculations apply to $A_4(\theta)$.

Now we consider A_2. An envelope for the class $\mathcal{F} = \{A_2(\theta) : \tilde{d}(\theta, \theta_0) < \delta\}$ is $F = (\|x_t\|\|\beta_0 - \alpha_0\| + \delta)1\{\gamma_0 - \delta^2 < q_{t-1} \le \gamma_0\}$, then it is possible to verify that

$$\log N_{[]}(\eta\|F\|_{P,2}, \mathcal{F}, L_2(P)) \le \log(1/\eta).$$

Now Theorem 11.4 in Kosorok (2008) yields

$$E^* \sup_{\tilde{d}(\theta, \theta_0) < \delta} |\mathbb{G}_T A_2(\theta)| = E^* \|\mathbb{G}_T\|_\mathcal{F} \times \|F\|_{P,2} \le \delta^2.$$

Similar calculations apply to A_3. Combining all the results with the fact that $O(\delta^2) = O(\delta)$, we obtain $E^* \sup_{\tilde{d}(\theta, \theta_0) < \delta} \|\mathbb{G}_T\|_{\mathcal{M}_\delta} \le \delta$.

Now when $\delta \mapsto \phi(\delta) = \delta$, $\phi(\delta)/\delta^\alpha$ is decreasing for any $\alpha \in (1,2)$. Thus, the condition of the Theorem 3.2.5 in van der Vaart and Wellner (1996) is satisfied with $\phi(\delta) = \delta$. Since $r_T^2 \phi(1/r_T) = r_T$, we obtain that $\sqrt{T}\tilde{d}(\hat{\theta}, \theta_0) = O_p(1)$. By the form of $\tilde{d}$, this implies that $\sqrt{T}\|\hat{\theta}_T - \theta_0\| = O_p(1)$ and $T|\hat{\gamma}_T - \gamma_0| = O_p(1)$. ∎

For $\tau \in \mathcal{T}$ and $\gamma \in \mathcal{G}$ define

$$S_T(\tau, \gamma, b) = \frac{1}{\sqrt{T}} \sum_{t=1}^T x_{t\gamma} \psi_\tau(y_t - F_{y_t}^{-1}(\tau | \Im_{t-1})) = T^{-1/2} \sum_{t=1}^T x_{t\gamma} \psi_\tau(u_{t\tau}),$$

25

where $u_{t\tau} = y_t - x'_{t\gamma}b$ and $x_{t\gamma} = x_t(\gamma)$.

The next lemma shows Donskerness of the process, and will help in the derivation of Theorem 2.

Lemma A1 *Suppose assumptions A1-A7 hold. Then, the process $S_T(\tau, \gamma)$ is Donsker.*

Proof. Let

$$
\begin{aligned}
S_T(\tau, \gamma, \theta(\tau)) &= T^{-1/2} \sum_{t=1}^{T} x_{t\gamma}[1(y_t \leq x'_{t\gamma}\theta(\tau)) - \tau] \\
&= T^{-1/2} \sum_{t=1}^{T} x_{t\gamma}[1(F_t(y_t) \leq \tau) - \tau]
\end{aligned}
$$

where $F_t(\cdot)$ is the conditional distribution function of y_t, and the last equality follows because Assumption A4 implies $F(\cdot)$ is absolute continuous and strictly increasing almost everywhere. Define $u_t = F_t(y_t)$, then u_t has a standard uniform distribution. Hence,

$$
\begin{aligned}
S_T(\tau, \gamma, \theta(\tau)) &= T^{-1/2} \sum_{t=1}^{T} x_{t\gamma}[1(u_t \leq \tau) - \tau] \\
&= T^{-1/2} \sum_{t=1}^{T} (x_t 1(y_{t-1} \leq \gamma), \quad x_t 1(y_{t-1} > \gamma))[1(u_t \leq \tau) - \tau]
\end{aligned}
$$

. Therefore by Assumption A2-A3, uniformly boundedness, monotonicity of the indicator function and Theorem 2.7.5 in van der Vaart and Wellner (1996), the bracketing entropy integral is finite and the Donsker property holds. ∎

Then, we can finally prove Theorem 2.

Proof. Let us fix $\gamma \in \mathcal{G}$, for certain $\tau \in \mathcal{T}$ given. From the Bahadur representation of the QR model

$$
\sqrt{T}\left(\hat{\theta}_\gamma(\tau) - \theta_\gamma(\tau)\right) = \widehat{\Omega}_1^{-1}(\tau, \gamma)S_T(\tau, \gamma) + o_p(1),
$$

where $S_T(\tau, \gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t(\gamma)\psi_\tau(y_t - F_t^{-1}(\tau))$ is the score function and $\psi_\tau(u) = \tau - I(u < 0)$ is the influence function in the quantile regression models. By the law of iterated expectations $E[x_t(\gamma)\psi_\tau(y_t - F_t^{-1}(\tau))|\Im_t] = 0$. Now, by the central limit theorem, Slutsky's theorem, and A1-A7, we have

$$
\frac{1}{\sqrt{T}}\widehat{\Omega}_1^{-1}(\tau, \gamma(\tau)) \sum_{t=1}^{T} x_t(\gamma)\psi_\tau(y_t - F_t^{-1}(\tau)) \Rightarrow N\left(0, \tau(1-\tau)\Sigma(\tau, \gamma)\right), \tag{33}
$$

26

with

$$\Sigma(\tau, \gamma) = \Omega_1(\tau, \gamma)^{-1}\Omega_0(\gamma, \gamma)\Omega_1(\tau, \gamma)^{-1}. \tag{34}$$

Now, we can extend this result to the corresponding functional process indexed by τ and γ, with τ and γ dense in $\mathcal{T}$ and $\mathcal{G}$ respectively. By Lemma A1, this is possible given that the class of functions we are interested belongs to the Donsker class. Then, this process converges in distribution in the Skorohod space $D(\mathcal{T}, \mathcal{G})$, equipped with the uniform norm, to a bivariate Gaussian process with zero mean and covariance kernel

$$K((\tau_i, \gamma_i), (\tau_j, \gamma_j)) = (\tau_i \wedge \tau_j - \tau_i\tau_j)\,\Omega_1(\tau_i, \gamma_i)^{-1}\Omega_0(\gamma_i, \gamma_j)\Omega_1(\tau_j, \gamma_j)^{-1},$$

for every $i, j = 1, \ldots, n$ with $\tau_i, \tau_j \in \mathcal{T}$ and $\gamma_i, \gamma_j \in \mathcal{G}$.

Finally, under $H_0 : R\theta_\gamma(\tau) = 0$, we obtain

$$\sqrt{T}R\left(\hat{\theta}_\gamma(\tau) - \theta_\gamma(\tau)\right) = \Omega_1^{-1}(\tau, \gamma)RS_T(\tau, \gamma) + o_p(1), \tag{35}$$

that converges in distribution to the mean zero bivariate Gaussian process $B(\tau, \gamma)$ with covariance kernel $K_\tau((\tau_i, \gamma_i), (\tau_j, \gamma_j))$. ∎

References

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

BEAUDRY, P., AND G. KOOP (1993): "Do recessions permanently change output?," *Journal of Monetary Economics*, 31, 149–163.

CAI, Y., AND J. STANDER (2008): "Quantile-Self Exciting Threshold Time Series Models," *Journal of Time Series Analysis*, 29, 187–202.

CANER, M. (2002): "A Note On Least Absolute Deviation Estimation of Threshold Model," *Econometric Theory*, 18, 800–814.

CHAN, K. S. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *Annals of Statistics*, 21, 520–533.

CHAN, K. S., AND H. TONG (1986): "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, 7, 179–194.

CHAN, K. S., AND R. TSAY (1998): "Limiting properties of the conditional least squares estimator of a continuous TAR model," *Biometrika*, 85, 413–426.

CLARIDA, R., L. SARNO, M. TAYLOR, AND G. VALENTE (2006): "The Role of Asymmetries and Regime Shifts in the Term Structure of Interest Rates," *Journal of Business*, 79, 1193–1224.

DAVIES, R. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 64, 247–254.

——— (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 74, 33–43.

GLOSTEN, L. R., R. JAGANNATTHAN, AND D. RUNKLE (1993): "On the relation between the expected value and the volatility of the nominal excess returns on stocks," *Journal of Finance*, 48, 1779–1801.

HANSEN, B. E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64, 413–430.

——— (1997): "Inference in TAR models," *Studies in Nonlinear Dynamics and Econometrics*, 2, 1–14.

——— (2000): "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575–603.

KOENKER, R. (2005): *Quantile Regression*. Cambridge University Press, New York, New York.

KOENKER, R., AND G. W. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–49.

KOENKER, R., AND Z. XIAO (2006): "Quantile Autoregression," *Journal of American Statistical Association*, 101, 980–990.

KOSOROK, M. R. (2008): *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag Press, New York, New York.

MARTINEZ, O., AND J. OLMO (2008): "A Nonlinear Threshold Model for the Dependence of Extremes of Stationary Sequences," Mimeo.

NELSON, D. (1991): "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, 59, 347–370.

SARNO, L., G. VALENTE, AND H. LEON (2006): "Nonlinearity in Deviations from Uncovered Interest Parity: An Explanation of the Forward Bias Puzzle," *Review of Finance*, 10, 443–482.

TERASVIRTA, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of American Statistical Association*, 89, 208–218.

TONG, H. (1983): "Threshold Models in Non-Linear Time Series Analysis," in *Lecture Notes in Statistics, 21*. Springer-Verlag Press, Berlin.

——— (1990): *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.

TONG, H., AND K. LIM (1980): "Threshold autoregression, limit cycles and cyclical data," *Journal of Royal Statistical Society*, 4, 245–292.

TSAY, R. (1989): "Testing and modeling threshold autoregressive processes," *Journal of American Statistical Association*, 18, 931–955.

VAN DE GEER, S. (2000): *Empirical Processes in M-Estimation*. Cambridge University Press, Cambridge.

VAN DER VAART, A., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag Press, New York, New York.

ZAKOIAN, J. (1994): "Threshold heteroscedastic models," *Journal of Economic Dynamics and Control*, 18, 931–955.