A NONLINEAR THRESHOLD MODEL FOR THE DEPENDENCE OF EXTREMES OF STATIONARY SEQUENCES

Oscar Martinez *
Universitat Rovira i Virgili, Spain

Jose Olmo
City University London, UK

March 2009

Abstract

We propose a TAR(3,1)-GARCH(1,1) model able to describe two different types of extreme events: a first type generated by large uncertainty regimes and a second type where extremes come from isolated dread/joy events. The novelty of this model resides on the definition of the regimes, motivated by the occurrence of extreme values, and of the threshold variable, defined by the shock affecting the process one period lagged. The model is able to uncover dependence and clustering of extremes in high and low volatility periods. By analyzing the period around the crisis of September 11^{th}, 2001 for GM stock prices we find evidence of predictability of extremes due to correlation in the mean between these observations. This finding supports the hypothesis of runs of negative returns due to correlation between extreme events rather to an increase in volatility.

JEL classification: C12; C15; C22; C51

Keywords and Phrases: asymmetries, consistency, crises, hypothesis testing, nonlinearities, threshold models

*Corresponding author: Oscar Martinez, Economics Department, Universitat Rovira i Virgili. Avenida de la Universidad 1, 43204, Reus, Spain. E-mail: oscar.martinez@urv.net. Phone: +34 977759849, Fax: +34 977300661

1 Introduction

In stationary linear time series models the extreme values are generated by the distribution function of the error term, thereby the importance of assuming distributions with higher kurtosis than the Gaussian distribution to describe these events. On the other hand for stationary nonlinear (multiplicative) models extreme observations can be generated either by the volatility process or by the error distribution function. GARCH models are thought as natural candidates for time series exhibiting clustering of extremes for they are able to generate this feature by the structure of dependence in the conditional volatility together with the shape of the error distribution. These processes modeling the conditional volatility, see Engle (1982) or Bollerslev (1986), are not capable however of producing runs of extremes of positive or negative sign. In particular, if the error distribution is symmetric these processes satisfy that

$$P\{y_t \leq -v|\Im_{t-1}\} = P\{y_t > v|\Im_{t-1}\}, \tag{1}$$

with v some positive value and $\Im_{t-1}$ denoting the $\sigma-$algebra generated by the set of available information up to time t-1. This property also holds for more convoluted GARCH type processes as E-GARCH of Nelson (1991), T-GARCH of Glosten, Jagannathan and Runkle (1993) and Zakoian (1994), or other related models as the stochastic volatility processes of Taylor (1994) and Harvey, Ruiz and Shephard (1994). A straightforward extension of these processes are ARMA-GARCH models. These processes model the conditional mean and make allowance then for mean values different from zero that tilt the conditional distribution of the time series y_t in one or other direction making more likely extreme values of the same sign of the conditional mean. A positive mean implies, in principle, a higher likelihood of extreme values in the positive tail. This fact, however, makes difficult for the model to describe periods of runs of extremes of opposite sign than the mean. In this case one should consider distributions with heavier tails than the Gaussian distribution. Also, from a finance perspective the statistical significance of the conditional mean component makes difficult to reconcile these models with theory on market efficiency.

A widely explored alternative that extends ARMA processes is the use of nonlinear models for the mean. These models are founded on the assumption of different regimes or states of the world and are used to capture different nonlinear phenomena exhibited by time series without having to entertain error distributions different from the Gaussian probability law. Examples of these nonlinear phenomena are asymmetries, time-irreversibility, different tail behavior of the distribution of the data, etc. These models have enjoyed a great popularity since the early work of Tong and Lim (1980), Tong (1983, 1990), Tsay (1989) or the general survey of Granger and Teräsvista (1993). For alternatives contemplating the presence of unit roots for certain regimes see Gonzalez and Gonzalo (1998) and for methods for estimating and testing for the presence of threshold effects see Chan (1990), Hansen (1996, 2000) or Gonzalo and Pitarakis (2002). Other family of nonlinear models is Smooth Transition Models (STAR) characterized by an infinite number of regimes and where the variable under study changes smoothly from one state to the other, see Teräsvirta (1994) among others.

Regarding the way in which the regime evolves over time two classes of threshold models can be distinguished. In the first class regimes are determined by an observable variable, examples of this with a finite number of regimes are the initial Threshold AutoRegressive (TAR) model of Tong (1978) or self-exciting processes (SETAR) where the threshold variable is a lagged value of the time series itself. The models in the second class assume that the regime cannot be observed and are determined by an unobservable stochastic process. In this class lies the widely studied Markov Switching Models, see Hamilton (1989), the STOPBREAK model of Engle and Smith (1999) or TIMA models of Gonzalo and Martinez (2006). In the last two cases the threshold variable is the shock that is not observable although estimable.

In this paper we claim that runs of very large observations of stationary time series can be under some conditions predictable for small time periods. In order to accommodate this postulate we propose a TAR model that has ingredients from both classes of nonlinear threshold models. The threshold variable is given by the term representing upcoming information into the model but lagged one period. This variable is not observable by its nature, but can be estimated at time t. The possibility of conditional heteroscedasticity is also entertained, thus the model that we propose is a TAR(3,1)-GARCH(1,1) process defined as follows:

$$y_t = \alpha + \begin{cases} \rho_1 y_{t-1} + h_t \varepsilon_t, & \varepsilon_{t-1} \leq u_1, \\ \rho_2 y_{t-1} + h_t \varepsilon_t, & u_1 < \varepsilon_{t-1} \leq u_2, \\ \rho_3 y_{t-1} + h_t \varepsilon_t, & \varepsilon_{t-1} > u_2, \end{cases} \tag{2}$$

with ε_t denoting the shock term, u_1 and u_2 threshold values defining the TAR (3,1) model, and h_t describing the volatility dynamics of a GARCH (1,1) process driving the error term. Note that the choice of volatility process is not instrumental for our analysis. Thus, one could choose instead our TAR model for the mean with a different structure for the volatility dynamics. Nevertheless, given the popularity and tractability of GARCH models we will restrict ourselves to the GARCH(1,1) case when presenting theoretical results.

Under certain conditions, u_1 and u_2 defining the bounds of the sequence of extremes of ε_t in each tail and $\rho_2 = 0$, this process makes allowance for dependence of extremes not only produced by high volatility regimes but by mean dependence produced by the occurrence of extreme shocks. While for economic and financial time series the first class of extremes is identified with periods of high uncertainty the second one could well describe, for example, booms and sudden drops in asset prices due to financial distress periods, periods of peaks in energy prices due to sudden weather variations, or periods of underpriced/overpriced currencies due to large country-related shocks.

The paper is structured as follows. In Section 2 the model, statistical properties and conditions to ensure stationarity and geometric ergodicity are introduced. Forecasting properties in the short and long run are also studied. Section 3 derives the asymptotic theory of the proposed model. In particular we develop a new nonlinearity hypothesis test to see the statistical significance of the threshold effect, and study the consistency and inference for the quasi-maximum likelihood estimators of the model parameters. Section 4 presents a Monte-Carlo analysis of the performance of size and power of this statistical test for finite samples. Section 5 introduces an application of the methodology to measure the effect on General Motors (GM) stock prices of September 11th, 2001. Finally, Section 6 concludes. All proofs are gathered into a mathematical appendix.

2 A TAR(3,1)-GARCH(1,1) model

We consider the following threshold autoregressive model with three regimes where we make allowance for conditional heteroscedasticity. The main feature of this model is that the threshold variable is the term describing shocks but one period lagged. The model is as follows:

$$y_t = \alpha + \begin{cases} \rho_1 y_{t-1} + h_t\varepsilon_t, & \varepsilon_{t-1} \leq u_1, \\ \rho_2 y_{t-1} + h_t\varepsilon_t, & u_1 < \varepsilon_{t-1} \leq u_2, \\ \rho_3 y_{t-1} + h_t\varepsilon_t, & \varepsilon_{t-1} > u_2, \end{cases} \tag{3}$$

with u_1 and u_2 threshold values defining the TAR (3,1) model, h_t a process describing the volatility dynamics of an error term $a_t := h_t\varepsilon_t$ driven by a GARCH(1,1) process, that is,

$$h_t^2 = \beta_0 + \beta_1 a_{t-1}^2 + \beta_2 h_{t-1}^2, \tag{4}$$

and $\{\varepsilon_t\}$ is a sequence of random shocks following a distribution function (*d.f.*) $F_\varepsilon(\cdot)$ with mean zero and variance one. The corresponding density function will be denoted by $f_\varepsilon(\cdot)$. This process can be expressed more compactly as

$$y_t = \alpha + \rho_t y_{t-1} + a_t, \tag{5}$$

with $\rho_t = \rho_1 I(\varepsilon_{t-1} \leq u_1) + \rho_2 I(u_1 < \varepsilon_{t-1} \leq u_2) + \rho_3 I(\varepsilon_{t-1} > u_2)$, and where $I(A)$ denotes the indicator function that takes a value of 1 if A is true and zero otherwise. Another alternative is considering as threshold variable the error term a_{t-1}. In this case the threshold values are time varying and depend on the volatility regime:

$$y_t = \alpha + \left[\rho_1 I(a_{t-1} \leq u_{1,t}^*) + \rho_2 I(u_{1,t}^* < a_{t-1} \leq u_{2,t}^*) + \rho_3 I(a_{t-1} > u_{2,t}^*)\right] y_{t-1} + a_t, \tag{6}$$

with $u_{j,t}^* = h_{t-1}u_j$, $j = 1, 2$, threshold values that depend on the conditional volatility process. For $\rho_2 = 0$, for example, this representation of the model shows that the structure of dependence in the data is driven by the occurrence of extreme observations in the shock variable ε_t independently of the volatility regime. This is in contrast to standard SETAR methodologies where the dependence structure, and therefore the occurrence of extremes, is influenced by the volatility regime.

This process accommodates many different dependence structures and time series dynamics. Our interest is primarily in describing financial time series. In this case it can be convenient to assume $\rho_2 = 0$. Some examples are plotted below.

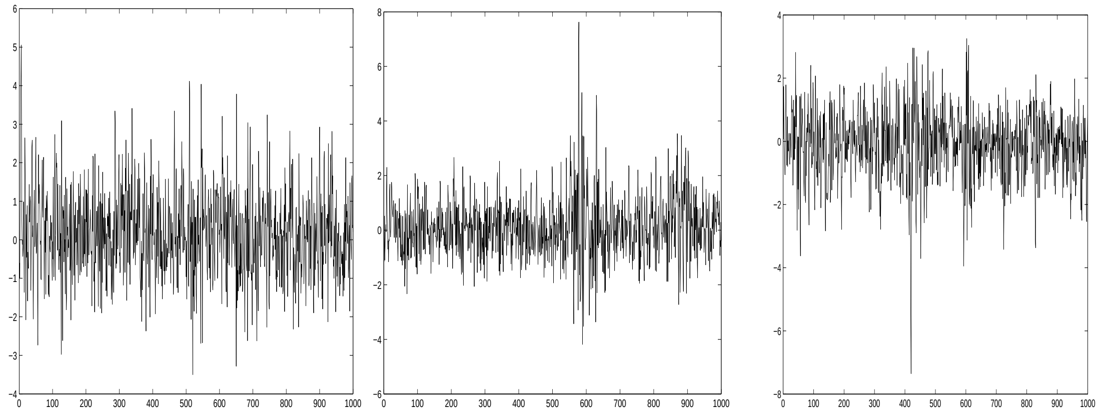

Figure 2.1. *Time series representing three different TAR(3,1)-GARCH(1,1) processes. The left panel depicts a process with* $\alpha = 0$, $(\rho_1, \rho_2, \rho_3) = (-0.7, 0, 0.7)$ *and* $(\beta_0, \beta_1, \beta_2) = (1, 0, 0)$. *Middle panel for* $\alpha = 0$, $(\rho_1, \rho_2, \rho_3) = (-0.7, 0, 0.7)$ *and* $(\beta_0, \beta_1, \beta_2) = (0.05, 0.10, 0.85)$*; and right panel for* $\alpha = 0$, $(\rho_1, \rho_2, \rho_3) = (0.7, 0, 0.7)$ *and* $(\beta_0, \beta_1, \beta_2) = (0.05, 0.10, 0.85)$. *The error follows a standard Gaussian distribution. The threshold values are* $u_1 = -1.64$ *and* $u_2 = 1.64$ *and the sample size is* $n = 1000$.

To ensure the stationarity and ergodicity of the above process we need to impose the following assumptions:

A.1 $\{\varepsilon_t\}$ is an independent and identically distributed (*iid*) sequence with mean zero, variance one, and distribution function F_ε with Lebesgue density f_ε, that is uniformly bounded and uniformly continuous.

A.2 $\beta_0 > 0$ and $\beta_i \geq 0$ for $i = 1, 2$.

A.3 $E\left[\max\left(0, log(\varepsilon_t^2)\right)\right] < \infty$, for all t.

A.4 $E\left[log(\beta_1 \varepsilon_t^2 + \beta_2)\right] < 0$, for all t.

A.5 $\beta_1 + \beta_2 < 1$.

A.6 $-\infty < E\left[log(\rho_t)\right] < 0$, for all t.

Assumptions **A.1** to **A.4** or **A.1**, **A.2** and **A.5** are conditions for the strict stationarity and ergodicity of the GARCH process a_t. Assumption **A.6** is a standard condition to show the strict stationarity and ergodicity of threshold models.

Theorem 1.- *Assume that* ***A.1*** *to* ***A.4*** *and* ***A.6*** *hold, then process* (3) *has a unique strictly stationary and ergodic solution. Equally, substituting* ***A.3*** *and* ***A.4*** *by* ***A.5****, the same result is obtained.*

The following proposition sets the conditions to ensure that the first k statistical moments of (3) are finite. First, define $\|x\|_k = \left(E\left[|x|^k\right]\right)^{1/k}$.

Proposition 1.- *Under the assumptions in Theorem 1 and the following conditions*

A.7 $\|\rho_t\|_k < 1$,

A.8 $\|\varepsilon_t^2\|_{k/2} < \infty$ *and* $E\left[\left(\beta_1 \varepsilon_t^2 + \beta_2\right)^{k/2}\right] < 1$,

the first k statistical moments of $\{y_t, a_t\}$ defined on process (3) are finite.

Proposition 2.- *Under assumptions in Proposition 1 for $k = 2$, the first statistical moment of process (3) is*

$$E[y_t] = \frac{\alpha}{1 - E[\rho_t]} + \frac{E[\rho_t a_{t-1}]}{1 - E[\rho_t]}. \tag{7}$$

If we further assume that the process has zero unconditional mean, the unconditional variance is

$$Var(y_t) = \frac{Var(a_t)}{1 - E\left[\rho_t^2\right]} + \frac{Cov\left(\rho_t^2, y_{t-1}^2\right) - E^2\left[\rho_t y_{t-1}\right]}{1 - E\left[\rho_t^2\right]}, \tag{8}$$

and the unconditional correlation of order one is

$$Corr(y_t, y_{t-1}) = E\left[\rho_t\right] + \frac{Cov(\rho_t, y_{t-1}^2)}{Var(y_t)}. \tag{9}$$

Note that the randomness of the autoregressive parameter adds one extra term compared to the corresponding standard AR(1) process: $\frac{E[\rho_t a_{t-1}]}{1-E[\rho_t]}$ in the unconditional mean, $\frac{Cov\left(\rho_t^2, y_{t-1}^2\right) - E^2[\rho_t y_{t-1}]}{1-E[\rho_t^2]}$ in the unconditional variance and $\frac{Cov(\rho_t, y_{t-1}^2)}{Var(y_t)}$ in the autocorrelation function of order one.

The expression for the optimal forecast l-periods ahead for the TAR(3,1)-GARCH(1,1) model is also an extension of the corresponding formulas for the AR(1)-GARCH(1,1) methodology. Thus, the optimal forecasts one-period ahead of y_t using the mean square prediction error criterion are

$$E[y_{t+1}|\Im_t] = \rho_{t+1} y_t, \tag{10}$$

and

$$Var\left(y_{t+1}|\Im_t\right) = h_{t+1}^2. \tag{11}$$

Proposition 3.- *Under assumptions in Proposition 2 the optimal forecast l-periods ahead, with $l > 1$, of process (3) is*

$$E[y_{t+l}|\Im_t] = \alpha \frac{1 - E[\rho_{t+1}]^{l-2}}{1 - E[\rho_{t+1}]} + E[\rho_{t+1}]^{l-1} \rho_{t+1} y_t + \sum_{i=1}^{l-1} E[\rho_{t+i+1} a_{t+i}|\Im_t] E[\rho_{t+1}]^{l-i-1}. \tag{12}$$

Furthermore, as $l \to \infty$ the optimal conditional forecast converges to the unconditional mean,

$$E[y_{t+l}|\Im_t] \overset{L_2}{\to} \frac{\alpha}{1-E[\rho_{t+1}]} + \frac{E[\rho_{t+1}a_t]}{1-E[\rho_{t+1}]}, \tag{13}$$

with L_2 standing for convergence in square norm.

The main advantage of model (3) is its flexibility to describe the dynamics in the mean process. In contrast to standard TAR models the regimes in our specification depend on the lagged error variable, and therefore the model can accommodate asymmetries in the likelihood of positive and negative extremes and in the occurrence of runs of extremes produced by the sign and magnitude of the shock ε_{t-1}. The following proposition entertains, in particular, the probability of runs of extremes in the above TAR model.

Proposition 4.- *Under model (3),*

$$\begin{aligned} P_{t-2}\left\{y_t \le -v, y_{t-1} \le -v\right\} = &\int_{-\infty}^{x_{1t}} F_\varepsilon\left(\frac{-v-(\alpha+\rho_1\left(z_{t-2}+\varepsilon h_{t-1}\right))}{h_t\left(\varepsilon\right)}\right) f_\varepsilon\left(\varepsilon\right)\partial\varepsilon \\ &+ \int_{x_{1t}}^{x_{2t}} F_\varepsilon\left(\frac{-v-(\alpha+\rho_2\left(z_{t-2}+\varepsilon h_{t-1}\right))}{h_t\left(\varepsilon\right)}\right) f_\varepsilon\left(\varepsilon\right)\partial\varepsilon \\ &+ \int_{u_2}^{x_{3t}} F_\varepsilon\left(\frac{-v-(\alpha+\rho_3\left(z_{t-2}+\varepsilon h_{t-1}\right))}{h_t\left(\varepsilon\right)}\right) f_\varepsilon\left(\varepsilon\right)\partial\varepsilon, \end{aligned}$$

where $P_s\{A_t\} = P\{A_t|\Im_s\}$, $z_{t-2} = E\left[y_{t-1}|\Im_{t-2}\right] = \alpha + \rho_{t-1}y_{t-2}$, $x_{1t} = \min\left\{u_{1,}\frac{-v-z_{t-2}}{h_{t-1}}\right\}$, $x_{2t} = \min\left\{u_{2,}\frac{-v-z_{t-2}}{h_{t-1}}\right\}$, $x_{3t} = \max\left\{u_{2,}\frac{-v-z_{t-2}}{h_{t-1}}\right\}$, $h_t^2(\varepsilon) = E\left[h_t^2\middle|\varepsilon_{t-1}=\varepsilon,\Im_{t-2}\right] = \beta_0+\beta_1\varepsilon^2+\beta_2h_{t-1}^2$, *and v denotes a positive threshold. Equally,*

$$\begin{aligned} P_{t-2}\left\{y_t \ge v, y_{t-1} \ge v\right\} = 1 - &F_\varepsilon\left(\frac{v-z_{t-2}}{h_{t-1}}\right) \\ &- \int_{x'_{1t}}^{u_1} F_\varepsilon\left(\frac{v-(\alpha+\rho_1\left(z_{t-2}+\varepsilon h_{t-1}\right))}{h_t\left(\varepsilon\right)}\right) f_\varepsilon\left(\varepsilon\right)\partial\varepsilon \\ &- \int_{x'_{2t}}^{x'_{3t}} F_\varepsilon\left(\frac{v-(\alpha+\rho_2\left(z_{t-2}+\varepsilon h_{t-1}\right))}{h_t\left(\varepsilon\right)}\right) f_\varepsilon\left(\varepsilon\right)\partial\varepsilon \\ &- \int_{x'_{3t}}^{\infty} F_\varepsilon\left(\frac{v-(\alpha+\rho_3\left(z_{t-2}+\varepsilon h_{t-1}\right))}{h_t\left(\varepsilon\right)}\right) f_\varepsilon\left(\varepsilon\right)\partial\varepsilon, \end{aligned}$$

where $x'_{1t} = \min\left\{u_{1,}\frac{v-z_{t-2}}{h_{t-1}}\right\}$, $x'_{2t} = \max\left\{u_{1,}\frac{v-z_{t-2}}{h_{t-1}}\right\}$, $x'_{3t} = \max\left\{u_{2,}\frac{v-z_{t-2}}{h_{t-1}}\right\}$.

In a pure GARCH(1,1) model and with $F_\varepsilon(\cdot)$ being symmetric about zero these two tail probabilities are identical. In our TAR framework, on the other hand, this will depend on the value of the autoregressive parameters in each regime, accommodating therefore, the occurrence of clustering of extremes due to large uncertainty regimes as well as to runs of correlated extremes.

3 Estimation and Inference for TAR(3,1)-GARCH(1,1) processes

The section is structured as follows. First, we discuss nonlinearity tests in Hansen's (1996) spirit to determine statistically the presence of TAR effects of the type described above; and second, the section describes estimation and related asymptotic properties for the nonlinear TAR(3,1)-GARCH(1,1) model.

3.1 Nonlinearity tests

Following the literature on threshold models we will distinguish two cases. One, in which the threshold vector $u = (u_1, u_2)$ is known, and a second case, in which the vector is not identified under the null hypothesis. In both scenarios the null hypothesis corresponds to the case $\rho_1 = \rho_2 = \rho_3$, or alternatively to $H_0 : \gamma_2 = \gamma_3 = 0$, in the following model,

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 y_{t-1} I(\varepsilon_{t-1} \leq u_1) + \gamma_3 y_{t-1} I(\varepsilon_{t-1} > u_2) + h_t \varepsilon_t, \tag{14}$$

with $\gamma_0 := \alpha$, $\gamma_1 := \rho_2$, $\gamma_2 := \rho_1 - \rho_2$ and $\gamma_3 := \rho_3 - \rho_2$ in model (3). We will see below that the implementation of heteroscedasticity robust tests implies that the choice of process h_t is not instrumental for the nonlinearity test. For simplicity and consistency with previous results we assume h_t following a $GARCH(1,1)$ process.

The null hypothesis implies no different correlation regimes determined by the magnitude of the standardized lagged shocks. In this way, we entertain a process that under the null hypothesis is an AR(1)-GARCH(1,1) model:

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + h_t \varepsilon_t,$$

with h_t a GARCH(1,1) defined in (4) . Let us denote ϕ_{0,H_0} for the vector of true parameters $(\gamma_0, \gamma_1, \beta_0, \beta_1, \beta_2)$ of this model under the null hypothesis. For u known and an observable threshold variable, this composite test is standard in the literature and appropriate test statistics are heteroscedasticity-robust $F-$tests and Wald tests. The special feature of the model under the alternative, an unobserved threshold variable, implies that to implement the test we need to do a preliminary estimation. Here, we estimate the model under the null hypothesis by quasi-maximum likelihood (QML) and store the residuals. Then, in a second stage we estimate model (14) using as proxy for the threshold variables the residual process from the null AR(1)-GARCH(1,1) model. The rationale for doing this is that under the null hypothesis the residual sequence converges in probability to the error sequence in (14). It is clear in this case that the regressors involving threshold variables are not statistically significant.

Further, in order to implement a heteroscedasticity-robust Wald type test as proposed in Hansen (1996) we use the ordinary least squares (OLS) method. The corresponding OLS vector of estimators $\widehat{\gamma}(u) = \left(\sum_{t=1}^{n} y_{t-1}(u) y_{t-1}(u)'\right)^{-1} \left(\sum_{t=1}^{n} y_{t-1}(u) y_t\right)$ of $\gamma := (\gamma_0, \gamma_1, \gamma_2, \gamma_3)' \in \Gamma$, with Γ a compact set and

$y_{t-1}(u) = (1 \;\; y_{t-1} \;\; y_{t-1}I(\varepsilon_{t-1} \leq u_1) \;\; y_{t-1}I(\varepsilon_{t-1} > u_2))'$, satisfies

$$\sqrt{n}(\widehat{\gamma}(u) - \gamma) \xrightarrow{d} N(0, \Sigma(u)) \tag{15}$$

with $\Sigma(u) := M(u,u)^{-1}V(u)M(u,u)^{-1}$ defined by $M(u,u) := E[y_{t-1}(u)y_{t-1}(u)']$, $V(u) := E[s_t(u)s_t(u)']$ and $s_t(u) := y_{t-1}(u)a_t$, where $\xrightarrow{d}$ denotes convergence in distribution.

The vector $\widehat{\gamma}(u)$ is, however, an unfeasible estimator of γ that depends on the unobservable vector of regressors $y_{t-1}(u)$. Let $\widehat{\phi}_{n,H_0}$ be the QML estimator vector of ϕ_{0,H_0} and $\varepsilon_{n,t-1}$ the corresponding residual sequence obtained from estimating the null AR(1)-GARCH(1,1) process. Now, we use these residuals to construct the feasible OLS estimator of γ defined by $\widehat{\gamma}_n(u) = \left(\sum_{t=1}^n \widehat{y}_{t-1}(u)\widehat{y}_{t-1}(u)'\right)^{-1} \left(\sum_{t=1}^n \widehat{y}_{t-1}(u)y_t\right)$ with $\widehat{y}_{t-1}(u) = (1 \;\; y_{t-1} \;\; y_{t-1}I(\varepsilon_{n,t-1} \leq u_1) \;\; y_{t-1}I(\varepsilon_{n,t-1} > u_2))'$. The following result shows that $\widehat{\gamma}_n(u)$ also converges to the same asymptotic distribution as $\widehat{\gamma}(u)$, making possible to construct a Wald type test.

Proposition 5.- *Assume that*

$$\left|\widehat{\phi}_{n,H_0} - \phi_{0,H_0}\right| = O_p\left(n^{-1/2}\right).$$

Hence, under H_0 and assumptions in Proposition 1 for $k = 8$,

$$\sup_{u \in U} \left|\sqrt{n}(\widehat{\gamma}(u) - \widehat{\gamma}_n(u))\right| = o_P(1).$$

Ling and McAleer (2003, Theorem 5.1) gives sufficient conditions for the asymptotic normality of the QML estimator in AR(1)-GARCH(1,1) processes. Therefore, using the result in Proposition 5 and further simple algebra we are ready to derive the nonlinearity test for u known. The hypothesis of interest can be written as $H_0 : R\gamma = 0$ with $R = [0_2 \;\; I_2]$ a block diagonal matrix where 0_2 and I_2 are the 2×2 null and identity matrices, respectively. The heteroscedasticity-robust Wald test in this case is

$$\widehat{T}_n(u) := n(R\widehat{\gamma}_n(u))' \left[R\widehat{\Sigma}_n(u)R'\right]^{-1} R\widehat{\gamma}_n(u), \tag{16}$$

with $\widehat{\Sigma}_n(u)$ the empirical version of $\Sigma(u)$, defined by $\widehat{\Sigma}_n(u) := \widehat{M}_n(u,u)^{-1}\widehat{V}_n(u)\widehat{M}_n(u,u)^{-1}$, where $\widehat{M}_n(u,u) := \frac{1}{n}\sum_{t=1}^{n} \widehat{y}_{t-1}(u)\widehat{y}_{t-1}(u)'$ and $\widehat{V}_n(u) := \frac{1}{n}\sum_{t=1}^{n} \widehat{s}_t(u)\widehat{s}_t(u)'$ with $\widehat{s}_t(u) := \widehat{y}_{t-1}(u)\left[y_t - \widehat{\gamma}_n(u)'\widehat{y}_{t-1}(u)\right]$.

Theorem 2.- *Let $\widehat{T}_n(u)$ be the Wald test for the null hypothesis H_0, for a given vector u known. Under assumptions in Proposition 5 the Wald statistic satisfies*

$$\widehat{T}_n(u) \xrightarrow{d} \chi_2^2 \tag{17}$$

with χ_2^2 a chi-square distribution with two degrees of freedom.

For the most interesting cases, such as testing for nonlinearity when the values of the threshold vector u are not known, u_1 and u_2 are nuisance parameters that cannot be identified under the null hypothesis. In this case Hansen (1996) shows that the composite nonlinearity test is nonstandard. As proposed by this author, see also Davies (1977, 1987) or Andrews and Ploberger (1994), hypothesis tests for nonlinearity can be based on the supremum and average of the relevant Wald test statistic computed over the domain of the nuisance parameter. In our case this is defined by

$$U = \left\{ (u_1, u_2) \in R^2 \quad \text{s.t.} \quad F_\varepsilon(u_1) \in (a_1, b_1) \wedge F_\varepsilon(u_2) \in (a_2, b_2) \quad \text{with} \;\; 0 < a_1 < b_1 < a_2 < b_2 < 1 \right\},$$

and the relevant test statistics are $sup_{u \in U} \widehat{T}_n(u)$ and $Ave_{u \in U} \widehat{T}_n(u)$ with sup and Ave standing for the supremum and average functionals, respectively. Define now the score function $S_n(u) := \frac{1}{\sqrt{n}}[M(u,u)]^{-1} \sum_{t=1}^{n} s_t(u)$, and the asymptotic covariance function

$$\Sigma(u, u^*) := M(u,u)^{-1} V(u, u^*) M(u^*, u^*)^{-1},$$

with $V(u, u^*) = E[s_t(u) s_t(u^*)']$ the functional counterpart of $V(u)$. For the results below we need the following two assumptions:

Assumption A.9: $\inf\limits_{u, u^* \in U} det(\Sigma(u, u^*)) > 0.$

Assumption A.10: The empirical estimators $\widehat{M}_n(u,u)$ and $\widehat{V}_n(u, u^*)$ converge uniformly to $M(u,u)$ and $V(u, u^*)$, respectively, over $u, u^* \in U$.

In particular assumption A.9 guarantees that the covariance function is well defined, and A.10 with Proposition 5 ensure that under the null hypothesis, $\sqrt{n}\,(\widehat{\gamma}_n(u) - \gamma) = S_n(u) + o_p(1)$, uniformly on u. For u fixed, expression (15) guarantees the weak convergence of $S_n(u)$ to a normal distribution. Now, the tightness of this process on $u \in U$, shown in Hansen (1996), guarantees the weak convergence of $S_n(u)$ to a Gaussian process with covariance function $\Sigma(u, u^*)$, with $u, u^* \in U$. Under the hypothesis $R\gamma = 0$ we have that $\sqrt{n} R \widehat{\gamma}_n(u)$ converges weakly to a zero mean Gaussian process with covariance function

$$\overline{\Sigma}(u, u^*) := R \Sigma(u, u^*) R'.$$

Theorem 3.- *Under assumptions in Proposition 5, A.9-A.10, and $H_0 : R\gamma = 0$, we have*

$$\widehat{T}_n(u) \Rightarrow T_O, \tag{18}$$

with T_O a chi-square process with zero mean and covariance function $\overline{\Sigma}(u, u^*)$. Also,

$$sup_{u \in U} \widehat{T}_n(u) \Rightarrow sup_{u \in U} T_O, \quad \text{and} \quad Ave_{u \in U} \widehat{T}_n(u) \Rightarrow Ave_{u \in U} T_O. \tag{19}$$

The distributions of these asymptotic processes depend, in general, on the covariance function $\overline{\Sigma}$; hence its critical values cannot be tabulated except in special cases. To obtain the $p-values$ of these two asymptotic tests we propose two possible approximations to the asymptotic distribution: Hansen's $p-value$ transformation and a Wild bootstrap approximation. The validity of these asymptotic approximations follows trivially from Hansen (1996, 1997). In Section 4 we explore the finite-sample accuracy of these approximations for different data generating processes and for both supremum and average tests.

First, we discuss in the following subsection the asymptotic properties of the estimation procedure under the alternative hypothesis of nonlinearity.

3.2 Asymptotic Properties of the Parameter Estimators

Once the hypothesis of linearity of the data is rejected we proceed to estimate jointly the parameters of the whole model by QML. In more detail, define $u = (u_1, u_2)$, $\phi = (\alpha, \rho, \beta)$, where $\rho = (\rho_1, \rho_2, \rho_3)$ and $\beta = (\beta_0, \beta_1, \beta_2)$, and maximize the following function,

$$L_n(\phi, u) = \sum_{t=1}^{n} l_t(\phi, u),$$

with

$$\begin{aligned} l_t(\phi, u) &= -\frac{1}{2}\ln h_t^2(\phi, u) - \frac{a_t^2(\phi, u)}{2h_t^2(\phi, u)}, \\ a_t(\phi, u) &= \varepsilon_t(\phi, u)h_t(\phi, u) = y_t - (\alpha, \rho)y_{t-1}(u), \\ h_t^2(\phi, u) &= \beta_0 + \beta_1 a_{t-1}^2(\phi, u) + \beta_2 h_{t-1}^2(\phi, u), \end{aligned}$$

and $h_0^2(\phi, u) = \frac{\beta_0}{1-\beta_1-\beta_2}$. For a given $u = (u_1, u_2) \in U$, the solution is $\widehat{\phi}(u)$, then in order to find the optimal threshold vector $\widehat{u}_n$ we maximize $L_n(\widehat{\phi}(u), u)$ with respect to u. Thus, the QML estimator of (ϕ, u) is $(\widehat{\phi}(\widehat{u}_n), \widehat{u}_n)$.

Under the alternative hypothesis of threshold effect the objective function $l_t(\phi, u)$ is neither differentiable nor continuous with respect to the parameter vector. This discontinuity implies that standard asymptotic results on consistency and asymptotic normality for the parameter estimators cannot be obtained. In fact, the consistency of the threshold estimator for these models needs to be studied on a case-by-case basis. Thus, Chan (1993) shows the n-consistency of the OLS estimator of the threshold value u for standard SETAR processes. For more convoluted processes, as our TAR model, the theoretical derivation of the convergence rate of the parameter estimators is much more cumbersome. Some related work can be found in Gonzalo and Martinez (2007). Alternatively, following Politis, Romano and Wolf (1999, p. 177) and Gonzalo and Wolf (2005), the convergence rate of each estimator can be estimated using subsampling techniques. We elaborate more on this in the following paragraphs, but first we introduce some extra assumptions to derive the consistency of the vector $(\widehat{\phi}(\widehat{u}_n), \widehat{u}_n)$ of parameter estimators.

Assumption A.11: Let $f_{a|t-1}(a)$ be the density function of a_t conditioned on $\Im_{t-1}$, then $\sup_a f_{a|t-1}(a) \leq \overline{f}$. Also, there exists some constant $M < \infty$ such that $|y_t| \leq M$ almost sure $(a.s.)$, for all t.

Assumption A.12: Let Θ be the parameter space, then $(\phi, u) \in \Theta$ if $u \in U$, $|\alpha| \leq \overline{\alpha}$, $0 < \underline{\beta}_0 \leq \beta_0 \leq \overline{\beta}_0$, $0 \leq \beta_1 \leq \overline{\beta}_1, 0 \leq \beta_2$ and

$$(\pi_1 + \pi_3) + (\pi_2 - \pi_1\pi_3) < 1,$$
$$(\pi_1 + \pi_3) - (\pi_2 - \pi_1\pi_3) > -1,$$
$$(\pi_2 - \pi_1\pi_3) > -1,$$

for $\pi_1 = 2\overline{f}(\lambda_1 + \lambda_2)$, $\pi_2 = 2\overline{f}(\lambda_1 u_1 + \lambda_2 u_2)\beta_1^{1/2}, \pi_3 = \beta_2^{1/2}$, with $\lambda_1 = (|\rho_1| + |\rho_2|) M$ and $\lambda_2 = (|\rho_2| + |\rho_3|) M$.

Assumption A.13: Let (ϕ_0, u_0) be the true parameters, then: $(\phi_0, u_0) = \arg\max_{(\phi,u)\in\Theta} E[L_n(\phi, u)]$.

Assumptions A.11 and A.12 impose conditions on the memory and on the extent of discontinuity of the TAR-GARCH process. Assumption A.13 is an identifiability condition for the true parameters of the process. With these assumptions in place we are ready to introduce the next theorem.

Theorem 4.- *Let* $(\widehat{\phi}(\widehat{u}_n), \widehat{u}_n) = (\widehat{\alpha}(\widehat{u}_n), \widehat{\rho}(\widehat{u}_n), \widehat{\beta}(\widehat{u}_n), \widehat{u}_n)$ *be the vector of quasi-maximum likelihood estimators of* (ϕ, u) *for the TAR(3,1)-GARCH(1,1) process in* (3). *Then, under assumptions in Theorem 1 and A.11 to A.13,*

$$\left(\widehat{\phi}(\widehat{u}_n) - \phi(u)\right) = o_P(1). \tag{20}$$

Remark: For the particular case of a TAR(3,1)-IID, condition A.12 boils down to assuming $2\overline{f}(|\rho_1| + 2|\rho_2| + |\rho_3|) M < 1$.

Due to the nonstandard nature of the problem asymptotic theory for general threshold models has not been widely explored yet. Some few examples are Chan (1993), Hansen (2000) or Gonzalo and Pitarakis (2002). The distribution of the parameter estimators, and in particular of the estimator of the threshold value u usually depends on the continuity of the threshold model. Thus, in principle, in the standard SETAR environment the inference problem can be considered solved when the model is continuous, in this case $\widehat{u}_n$ and the rest of parameter estimators in the model converge asymptotically to a normal distribution at a $\sqrt{n}-$rate. In the discontinuous case Chan (1993) shows that $n(\widehat{u}_n - u)$ converges weakly to a non-degenerate distribution that depends on a very complicated way on a compound Poisson process and that apparently cannot be consistently estimated. Under more restrictive assumptions such as threshold effect vanishing asymptotically, the method of Hansen (2000) can be employed. As discussed before, Gonzalo and Wolf (2005) solve this problem by using subsampling techniques to estimate the convergence rate of the threshold estimator and to approximate the distribution of the parameter estimators. In particular, these authors extend this technique to situations where the discontinuity of the model is not known and the inference for the regression parameters of the model becomes very difficult.

The discontinuity of our TAR(3,1)-GARCH(1,1) process implies that the whole vector of parameter estimators $\widehat{\phi}(\widehat{u}_n)$ is consistent at a higher rate than $\sqrt{n}$, and therefore one has to rely on asymptotic results of the type derived in Chan (1993) and Hansen (2000). Also, as in the SETAR case it is not known whether a bootstrap approach would work. Thereby, following Gonzalo and Wolf (2005) we propose subsampling methods to estimate the convergence rate and to approximate the exact finite distribution of the vector $(\widehat{\phi}(\widehat{u}_n), \widehat{u}_n)$. For sake of space we only present the main results. The interested reader is referred to Politis, Romano and Wolf (1999).

To this end let $J_n(x, F) = P\{\tau_n|\widehat{\theta}_n - \theta| \leq x\}$ be the finite-sample distributions of the standardized QML estimator $\widehat{\theta}_n$ of the parameter θ, where with an abuse of notation $\widehat{\theta}_n$ and θ stand for any element of the vector $(\widehat{\phi}(\widehat{u}_n), \widehat{u}_n)$ and $(\phi(u), u)$, respectively. In our case the convergence rate τ_n is not known, and needs to be estimated. Define $\widehat{J}_n(x, F) = P\{\widehat{\tau}_n|\widehat{\theta}_n - \theta| \leq x\}$ and the corresponding subsampling approximation

$$\widehat{J}_{n,b}(x, F) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} I\left(\widehat{\tau}_b|\widehat{\theta}_{b,i} - \widehat{\theta}_n| \leq x\right),$$

where $1 < b < n$ is referred to as the block size, $\widehat{\theta}_{b,i}$ is the QML estimator computed from a subsample of consecutive observations $(y_i, \ldots, y_{i+b-1})$, and $\widehat{\tau}_b$ is the corresponding convergence rate for each subsample. Given that the rate of convergence is not known this expression cannot be computed. Instead, we define

$$\widetilde{J}_{n,b}(x, F) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} I\left(|\widehat{\theta}_{b,i} - \widehat{\theta}_n| \leq x\right),$$

and denote by $\widetilde{J}^{-1}_{n,b}(\nu, F)$ the ν-quantile of $\widetilde{J}_{n,b}(x, F)$. Now, let $b_i = \lfloor n^{\kappa_i} \rfloor$ for constants $0 < \kappa_1 < \ldots < \kappa_I < 1$, with $\lfloor \cdot \rfloor$ denoting the nearest smaller integer value, let ν_j for $j = 1, \ldots, J$ be some points in $(0.5, 1)$ and let $z_{i,j} = log(\widetilde{J}^{-1}_{n,b_i}(\nu_j, F))$. If τ_n is of the form $\tau_n = n^{\kappa}$ we can define the following estimator $\widehat{\tau}_n = n^{\widehat{\kappa}_{I,J}}$ with $\widehat{\kappa}_{I,J}$ given by

$$\widehat{\kappa}_{I,J} = -\frac{\sum_{i=1}^{I} (z_{i,\cdot} - \overline{z})(log\, b_i - \overline{log})}{\sum_{i=1}^{I} (log\, b_i - \overline{log})^2}, \tag{21}$$

where $z_{i,\cdot} = \frac{1}{J} \sum_{j=1}^{J} z_{i,j}$, $\overline{y} = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} z_{i,j}$, and $\overline{log} = \frac{1}{I} \sum_{i=1}^{I} log(b_i)$.

Politis, Romano and Wolf (1999, Theorem 8.2.1) shows that this estimator satisfies

$$\widehat{\kappa}_{I,J} = \kappa + o_P((log\, n)^{-1}). \tag{22}$$

The following proposition establishes the asymptotic validity of subsampling with an estimated rate of convergence.

Proposition 6.- *Assume that (i) $J_n(x,F)$ converges weakly to a continuous limiting distribution $J(x,F)$; (ii) $b \to \infty$ and $b/n \to 0$, as $n \to \infty$; (iii) $\tau_n = n^{\kappa}$, with $1/2 \leq \kappa \leq 1$. Let $\widehat{\tau}_n = n^{\widehat{\kappa}_{I,J}}$ defined in (21), then,*

$$\sup_{x \in \mathbb{R}} |\widehat{J}_{n,b}(x,F) - J(x,F)| = o_P(1). \tag{23}$$

Let $\nu \in (0,1)$, and let $\widehat{c}_{n,b}(\nu) = \widehat{J}^{-1}_{n,b}(\nu, F)$ be the ν-th sample quantile of $\widehat{J}_{n,b}(x,F)$. Then,

$$P\left\{\widehat{\tau}_n|\widehat{\theta}_n - \theta| \leq \widehat{c}_{n,b}(\nu)\right\} \longrightarrow \nu, \quad \text{as } n \to \infty. \tag{24}$$

Thus, the asymptotic coverage probability of the interval $\widehat{\theta}_n \pm \widehat{\tau}_n^{-1}\widehat{c}_{n,b}(\nu)$ is the nominal level ν.

The proof follows as a special case of Theorem 8.3.1 of Politis, Romano and Wolf (1999, p. 184).

The threshold nature of the TAR(3,1)-GARCH(1,1) process introduced in (3) implies that the maximum likelihood estimators of the model parameters properly standardized satisfy assumption (*i*) in proposition 6. Assumption (*ii*) is standard in subsampling, and can be achieved by a proper choice of the block size b. Finally, and again due to the threshold nature of the process, assumption (*iii*) holds with $1/2 < \kappa \leq 1$.

4 Simulation Experiment

This section consists of a Monte-Carlo simulation experiment to examine the performance of size and power of the preceding test for finite samples. For completeness, and following Hansen (1996), we explore the bootstrap and p-value transformation methods for approximating the finite-sample distribution of the supremum and average tests above discussed.

4.1 Finite-Sample Performance of Nonlinearity Tests

For the first simulation experiment we commence studying the empirical size of the test for three linear processes in the mean. These are an *iid* process, a pure GARCH(1,1) process and an AR(1)-GARCH(1,1) process:

1. $y_t = \varepsilon_t$ with ε_t *iid*(0,1),
2. $y_t = a_t = \varepsilon_t h_t$, and $h_t^2 = \beta_0 + \beta_1 a_{t-1}^2 + \beta_2 h_{t-1}^2$ with parameters $\beta_0 = 0.05$, $\beta_1 = 0.10$ and $\beta_2 = 0.85$, and ε_t defined as in the previous case.
3. $y_t = \rho y_{t-1} + a_t$ with $\rho = 0.20$, and a_t defined as in the previous case.

The error term is assumed standard Gaussian although other simulation experiments could be developed to see the robustness of the test to departures from Gaussianity. In all of the experiments the threshold regime is defined by the following space:

$$U = \{(u_1, u_2) \in R^2 \quad \text{s.t.} \quad F_\varepsilon(u_1) \in (0.10, 0.30) \wedge F_\varepsilon(u_2) \in (0.60, 0.90)\}.$$

Whereas we use $n = 250$, 500 for the Wild bootstrap approximation we consider $n = 250$, 500 and 1000 for the p-value transformation. This is due to the poor results obtained from the second method for small sample sizes. The number of Monte-Carlo simulations for all the experiments is $M = 2000$. The following table 4.1 reports empirical estimates of the size at 5% and 10% significance level for the statistic defined by the supremum of $\widehat{T}_n(u)$ over the set of possible threshold values.

$sup_{u\in U}\widehat{T}_n(u)$	n=250		n=500		n=1000	
size	0.05	0.10	0.05	0.10	0.05	0.10
IID	0.077	0.138	0.068	0.136	0.059	0.116
$GARCH(1,1)$	0.090	0.160	0.077	0.133	0.070	0.130
$AR(1) - GARCH(1,1)$	0.087	0.145	0.080	0.150	0.070	0.125

Table 4.1. *Empirical size at 5% and 10% of the $sup_{u\in U}\widehat{T}_n(u)$ test for $n = 250$, $n = 500$ and $n = 1000$ for different data generating processes derived from the Hansen p-value transformation. $M = 2000$ Monte-Carlo simulations and 500 internal simulation replications (except for the case $n = 1000$ where we use 300 internal simulation replications).*

For the statistic defined by the average of $\widehat{T}_n(u)$ the results of the simulated size are reported in table 4.2:

$Ave_{u\in U}\widehat{T}_n(u)$	n=250		n=500		n=1000	
size	0.05	0.10	0.05	0.10	0.05	0.10
IID	0.059	0.110	0.057	0.116	0.055	0.120
$GARCH(1,1)$	0.069	0.128	0.066	0.113	0.062	0.113
$AR(1) - GARCH(1,1)$	0.070	0.122	0.069	0.119	0.058	0.112

Table 4.2. *Empirical size at 5% and 10% of the $Ave_{u\in U}\widehat{T}_n(u)$ test for $n = 250$, $n = 500$ and $n = 1000$ for different data generating processes derived from the Hansen p-value transformation. $M = 2000$ Monte-Carlo simulations and 500 internal simulation replications (except for the case $n = 1000$ where we use 300 internal simulation replications)*

The Hansen $p - value$ transformation is too "liberal" for the supremum case. This can be produced by the definition of the U space. Hansen (1996) observes that the pointwise test statistics are ill-behaved for extreme values of u, that is, with $F_\varepsilon(u)$ close to 0 or 1, and proposes a $[0.2, 0.8]$ region for searching potential thresholds. Our model, however, focuses on threshold effects on the extremes of the time series, hence our interest in giving more freedom to the threshold region in order to capture this effect. Nevertheless, the empirical size seems to converge to the nominal size for the three processes and two test statistics.

This phenomenon, on the other hand, is less important for the Wild Bootstrap approximation for which we report simulations for $n = 250, 500$ and M=2000 in tables 4.3 and 4.4.

$sup_{u \in U} \widehat{T}_n(u)$	n=250		n=500	
size	0.05	0.10	0.05	0.10
IID	0.056	0.107	0.054	0.107
$GARCH(1,1)$	0.051	0.110	0.058	0.108
$AR(1) - GARCH(1,1)$	0.062	0.113	0.066	0.111

Table 4.3. *Empirical size at* 5% *and* 10% *of the* $sup_{u \in U} \widehat{T}_n(u)$ *test for* $n = 250$*,* $n = 500$ *for different data generating processes derived from the Wild bootstrap p-value approximation.* $M = 2000$ *Monte-Carlo simulations and 500 internal simulation replications.*

$Ave_{u \in U} \widehat{T}_n(u)$	n=250		n=500	
size	0.05	0.10	0.05	0.10
IID	0.043	0.096	0.052	0.101
$GARCH(1,1)$	0.044	0.094	0.049	0.100
$AR(1) - GARCH(1,1)$	0.046	0.092	0.054	0.099

Table 4.4. *Empirical size at* 5% *and* 10% *of the* $Ave_{u \in U} \widehat{T}_n(u)$ *test for* $n = 250$*,* $n = 500$ *for different data generating processes derived from the Wild bootstrap p-value approximation.* $M = 2000$ *Monte-Carlo simulations and 500 internal simulation replications.*

Finally, we present power results for the test when we use the Wild bootstrap approximation. For that, we consider two different models. In both cases the conditional mean is given by:

$$y_t = 0.20 y_{t-1} I(\varepsilon_{t-1} \leq -1.70) - 0.20 y_{t-1} I(\varepsilon_{t-1} > 1.70) + a_t.$$

In the first case, $a_t = \varepsilon_t$, in the second one $a_t = \varepsilon_t h_t$ with $h_t^2 = 0.05 + 0.10 a_{t-1}^2 + 0.85 h_{t-1}^2$. In both cases, ε_t is *iid* $N(0,1)$. The results are in tables 4.5 and 4.6.

$sup_{u \in U} \widehat{T}_n(u)$	n=250		n=500	
size	0.05	0.10	0.05	0.10
$TAR - IID$	0.332	0.476	0.664	0.779
$TAR - GARCH(1,1)$	0.257	0.382	0.527	0.657

Table 4.5. *Empirical power at* 5% *and* 10% *of the* $sup_{u\in U}\widehat{T}_n(u)$ *test for* $n = 250$, $n = 500$ *for different data generating processes derived from the Wild bootstrap p-value approximation.* $M = 2000$ *Monte-Carlo simulations and 500 internal simulation replications.*

$Ave_{u\in U}\widehat{T}_n(u)$	n=250		n=500	
$size$	0.05	0.10	0.05	0.10
$TAR - IID$	0.431	0.562	0.776	0.857
$TAR - GARCH(1,1)$	0.339	0.467	0.643	0.742

Table 4.6. *Empirical power at* 5% *and* 10% *of the* $Ave_{u\in U}\widehat{T}_n(u)$ *test for* $n = 250$, $n = 500$ *for different data generating processes derived from the Wild bootstrap p-value approximation.* $M = 2000$ *Monte-Carlo simulations and 500 internal simulation replications.*

Note that in both examples and for both test statistics the power grows with the sample size, however the average test statistic seems to be more powerful. Regarding the structure of the volatility process we observe more power of the nonlinearity test against homoscedastic alternative processes. To obtain a better insight about the power of both test statistics for the Wild bootstrap approximation method we carry out another two Monte Carlo experiments where different values of (ρ_1, ρ_2, ρ_3) and (u_1, u_2) are considered. The family of models under the alternative:

$$y_t = \rho y_{t-1} I(\varepsilon_{t-1} \leq -u) - \rho y_{t-1} I(\varepsilon_{t-1} > u) + a_t,$$

are indexed in the first case by ρ, with $\rho = i/10$ for $i = 0, 1, \ldots, 9$ and $u = 1.7$; and by u in the second case, with $u = 1.5 + i/10$ for $i = 0, 1, \ldots, 6$ and $\rho = 0.2, 0.5$. For simplicity, we only consider in these cases $a_t = \varepsilon_t h_t$ with $h_t^2 = 0.05 + 0.10a_{t-1}^2 + 0.85h_{t-1}^2$ and a sample size $n = 250$. As in previous cases, ε_t is *iid* $N(0, 1)$. The empirical size of the test is reported in Figure 4.1. As expected, the power of the test is increasing with ρ, although it stabilizes close to one after $\rho = 0.60$. However the power is quite stable respect to the threshold parameter u.

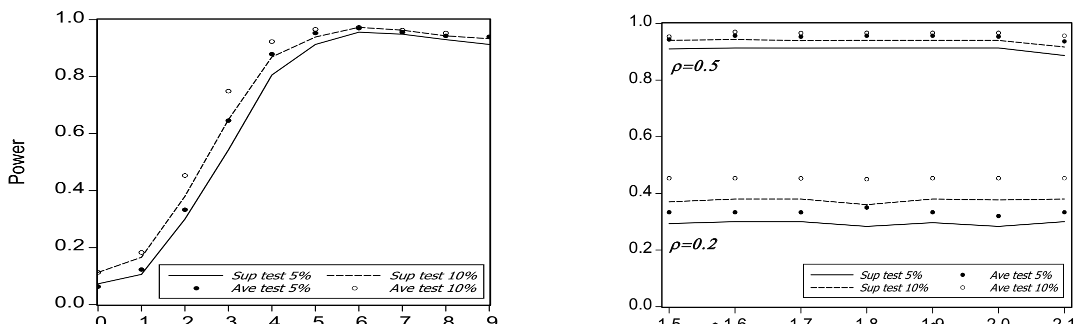

Figure 4.1. *Empirical power at* 5% *and* 10% *of the* $\sup_{u\in U}\widehat{T}_n(u)$ *and* $Ave_{u\in U}\widehat{T}_n(u)$ *tests for* $n = 250$ *and different data generating processes: on the left the power is indexed by* ρ, *on the right by* u. *The power is derived from the Wild bootstrap p-value approximation with* $M = 300$ *Monte-Carlo simulations and 200 internal simulation replications.*

The following section explores the suitability of these nonlinear processes for modeling financial returns.

5 Empirical application: Predicting in crises episodes

After the bombing attacks that shook the US in September 11th, 2001 the stock exchanges all around the world fell dramatically not only that day but during a short period of time after the attack. It is striking however to observe that this drop in asset prices worldwide elapsed only a short period of time, five to ten days and then markets went back to normal. The following plot represents the sequences of prices and log-returns $r_t = 100\,(lnP_t - lnP_{t-1})$ of General Motors (GM) stocks for this episode. The data are collected from Yahoo-Finance.

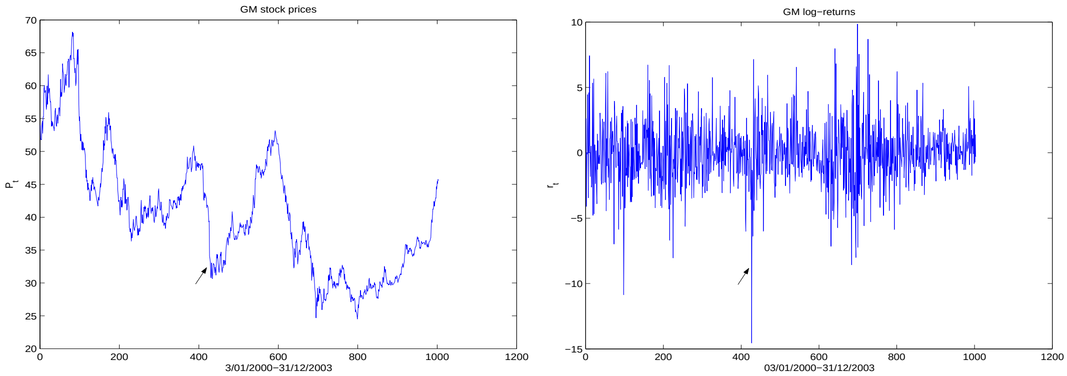

Figure 5.1. *The left and right panel depict the prices and log-returns, respectively, of GM stocks for the period* $03/01/2000 - 31/12/2003$*. The arrows points to September 11th, 2001.*

In this section the TAR-GARCH methodology is applied to determine statistically if the returns on the days following this event were predictable or not. If the events simply sparked an increase in volatility as stated by the leverage effect investors were better off by conserving their assets than to exposing to adverse movements of prices before the realization of the buy/sale order. In contrast, if these events were sparked by an extreme shock investors could have predicted future returns just after the shocks.

The following table reports the estimates of the different candidate models above studied. The analysis comprises 1004 observations.

$Model$	$GARCH(1,1)$	$AR(1)-GARCH(1,1)$	$TAR(3,1)-GARCH(1,1)$
α	0.104 (0.069)	0.108 (0.070)	0.183 (0.072)
ρ_1	-	-0.052 (0.031)	0.072 (0.043)
ρ_2	-	-0.052 (0.031)	-0.031 (0.061)
ρ_3	-	-0.052 (0.031)	1.225 (0.027)
β_0	0.117 (0.033)	0.117 (0.033)	0.087 (0.069)
β_1	0.079 (0.012)	0.076 (0.012)	0.080 (0.050)
β_2	0.902 (0.014)	0.904 (0.014)	0.908 (0.105)
Log lkl	2252.3	2251.1	2659.8

Table 5.1. $u_1 = -0.919$, $u_2 = 0.785$. *Estimates for September 2001 subsample (01/01/2000-31/12/2003), n=1004. p-value of Hansen test (*$sup_{u\in U}T_n(u)$*=0.046, p-value of* $Ave_{u\in U}T_n(u)$*=0.064). The standard errors of the different estimates are in brackets. In the TAR(3,1)-GARCH(1,1) column these values correspond to the subsampling exercise for* $b = 500$ *and* $M = n - b + 1$.

While the linear AR(1)-GARCH(1,1) model points towards a negative conditional mean process the nonlinear TAR-GARCH model also reflects this effect for the middle regime but describes as well two outer regimes where observations have a different and stronger dependence structure. The number of extremes in the sequence of shocks is 200 for the lower threshold and 128 for u_2. Hence, there is sufficient information in the samples to believe that there is positive dependence between series of positive extremes and between runs of negative extremes. The case of positive extremes is more significant. There is statistical evidence of nonlinearity and thereby of the presence of different regimes for the conditional mean process. Both supremum and average Hansen tests and the corresponding bootstrap counterpart tests ($sup_{u\in U}T_n(u)$=0.024, and $Ave_{u\in U}T_n(u)$=0.048) are found significant at 10%, and the likelihood function of the TAR model is substantially larger than that of the GARCH and AR-GARCH models. Further, this is also observed from the standard deviation estimates obtained from the subsampling approximation with $b = 500$, see Politis, Romano and Wolf (1999, p. 95) for details on how to estimate the standard deviation via subsampling methods.

Therefore, although the magnitude of the lower regime autoregressive parameter is small we believe that the nonlinearity of this model supports the presence of dependence in both extreme regimes, and therefore provides evidence to claim that the sequence of extreme observations after the bombing attacks of September 11^{th} were positively correlated. It is also worth mentioning that these effects could have been more significant if *NYSE* would have not interrupted trading in the floor for one week after the attack.

Finally, a simple visual inspection of the histogram of the residuals, figure 5.2, also supports the statistical significance of the $TAR(3,1) - GARCH(1,1)$ model.

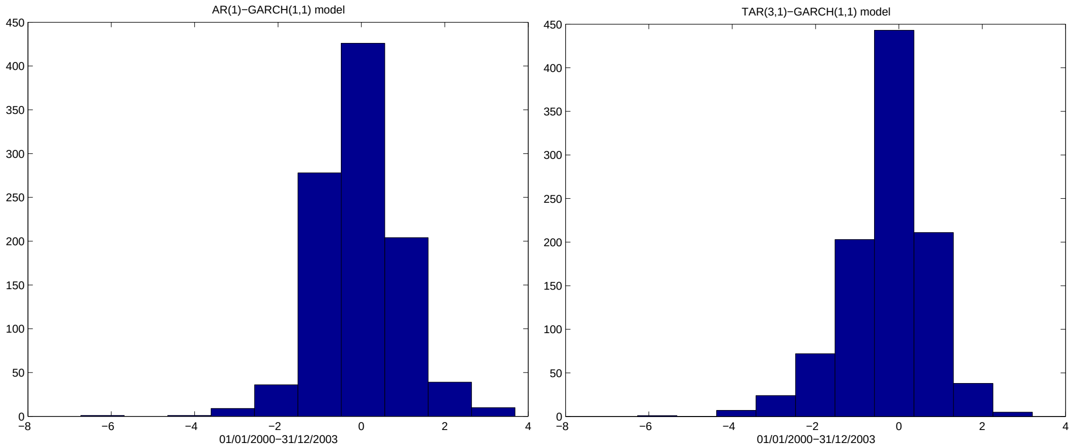

Figure 5.2. *Histograms for the residuals sequence from AR(1)-GARCH(1,1) (left panel) and from TAR(3,1)-GARCH(1,1) (right panel) for the period* $03/01/2000 - 31/12/2003$.

6 Conclusions

This paper introduces a new class of nonlinear threshold models. Its novelty resides on two features of the model that make it different from previous TAR methodologies. First, the regimes are motivated by the occurrence of extreme values, and second, the threshold variable determining the regime is defined by the shock affecting the process in the preceding period. In this way this process is able to describe two types of dependence, linear dependence derived from the occurrence of extreme shocks and clustering of large observations derived from the occurrence of high volatility periods.

The model is flexible in what is able to describe a variety of dependence structures; asymmetries in the probabilities in the tails, in the sequences of runs of extremes, etc. This is particularly interesting for modeling financial time series for this model is able to replicate in a parsimonious way the stylized facts commonly encountered in these series, including the absence of linear correlation, but accommodating at the same time the possibility of linear correlation in the extremes. This fact led us in the empirical application to study the consequences of the September 11^{th} terrorist attack in GM stocks. Using our TAR-GARCH method we find evidence of predictability of extremes after this event due to correlation in the extremes and not to an increase in the underlying volatility.

MATHEMATICAL APPENDIX:

Proof of Theorem 1: The strict stationarity and ergodicity of a_t and ρ_t together with assumption **A.6** are sufficient conditions for the unique strictly stationary and ergodic solution of (3). This is shown in Brandt (1986). In order to prove the strict stationarity and ergodicity of a_t Kristensen (2006) uses assumptions A.1 to A.4 and Ling and McAleer (2003) A.1, A.2 and A.5. □

Proof of Proposition 1: From equation (5) and Theorem 1,

$$y_t = \rho_t y_{t-1} + a_t = \sum_{j=0}^{\infty} a_{t-j} \prod_{i=0}^{j-1} \rho_{t-i}.$$

with $\prod_{i=0}^{-1} \rho_{t-i} = 1$ by convention. Denote $\|\rho_t\|_k = \lambda_1$. Then, from the Minkowsky inequality, independence of ε_t and strict stationarity of a_t:

$$\begin{aligned} \|y_t\|_k &\leq \sum_{j=0}^{\infty} \|a_{t-j} \prod_{i=0}^{j-1} \rho_{t-i}\|_k = \|a_t\|_k + \sum_{j=0}^{\infty} \|\rho_{t-j} a_{t-j-1}\|_k \| \prod_{i=0}^{j-1} \rho_{t-i}\|_k \\ &\leq \|a_t\|_k + \sum_{j=0}^{\infty} \lambda_1^j \|\rho_{t-j} a_{t-j-1}\|_k \leq \frac{(1 - \lambda_1 + \max_i |\rho_i|) \|a_t\|_k}{(1-\lambda_1)}, \end{aligned}$$

with $\lambda_1 < 1$ by assumption A.7. Then, it is sufficient to show that $\|a_t\|_k < \infty$, to prove Proposition 1. For that,

$$\begin{aligned} \|a_t\|_k &= \|a_t^2\|_{k/2}^{1/2} \\ \|a_t^2\|_{k/2} &= \|\varepsilon_t^2 h_t^2\|_{k/2} = \|\varepsilon_t^2\|_{k/2} \|h_t^2\|_{k/2} \\ \|h_t^2\|_{k/2} &\leq \beta_0 + \|\beta_1 \varepsilon_{t-1}^2 + \beta_2\|_{k/2} \|h_{t-1}^2\|_{k/2} \leq \frac{\beta_0}{(1-\lambda_2)} \end{aligned}$$

with $\lambda_2 = \|\beta_1 \varepsilon_t^2 + \beta_2\|_{k/2} < 1$ and $\|\varepsilon_t^2\|_{k/2} < \infty$ by assumption A.8, which proves that $\|a_t\|_k < \infty$. □

Proof of Proposition 2: Now we derive the first two moments of the process $y_t = \alpha + \rho_t y_{t-1} + a_t$, with $a_t = h_t \varepsilon_t$, when the assumptions in Proposition 1 hold for $k = 2$.

$$E[y_t] = \alpha + E[\rho_t y_{t-1}]. \tag{25}$$

Replacing y_{t-1} in the above expression we obtain

$$E[y_t] = \alpha + E[\rho_t y_{t-1}] = \alpha + E[\rho_t (\alpha + \rho_{t-1} y_{t-2} + a_{t-1})].$$

Now, using the stationarity of the process $\rho_t y_{t-1}$ and the independence of ρ_t from $\rho_{t-1} y_{t-2}$ it is simple to see that

$$E[y_t] = \frac{\alpha}{1 - E[\rho_t]} + \frac{E[\rho_t a_{t-1}]}{1 - E[\rho_t]}.$$

Now in order to obtain $E[y_t] = 0$ the intercept must be $\alpha = -E[\rho_t a_{t-1}]$. For that case, the unconditional variance is

$$Var(y_t) = E[y_t^2].$$

Also,

$$Var(y_t) = Var(\rho_t y_{t-1}) + Var(a_t).$$

Note that

$$Var\left(\rho_t y_{t-1}\right) = E\left[\rho_t^2 y_{t-1}^2\right] - E^2\left[\rho_t y_{t-1}\right],$$

and

$$E\left[\rho_t^2 y_{t-1}^2\right] = Cov(\rho_t^2, y_{t-1}^2) + E\left[\rho_t^2\right] E\left[y_{t-1}^2\right].$$

Then

$$E\left[y_t^2\right] = Cov(\rho_t^2, y_{t-1}^2) + E\left[\rho_t^2\right] E\left[y_{t-1}^2\right] - E^2\left[\rho_t y_{t-1}\right] + Var\left(a_t\right),$$

and by stationarity $(Var\left(y_t\right) = E\left[y_t^2\right] = E\left[y_{t-1}^2\right])$ we obtain

$$Var\left(y_t\right) = \frac{Var\left(a_t\right)}{1 - E\left[\rho_t^2\right]} + \frac{Cov\left(\rho_t^2, y_{t-1}^2\right) - E^2\left[\rho_t y_{t-1}\right]}{1 - E\left[\rho_t^2\right]}.$$

Finally, for the first order autocorrelation we compute first the autocovariance of order one that yields,

$$Cov(y_t, y_{t-1}) = E\left[\rho_t\right] Var\left(y_t\right) + Cov(\rho_t, y_{t-1}^2),$$

given that $E[\rho_t y_{t-1}^2] = Cov(\rho_t, y_{t-1}^2) + E[\rho_t] E[y_{t-1}^2]$. Hence the first order autocorrelation is

$$Corr(y_t, y_{t-1}) = E\left[\rho_t\right] + \frac{Cov(\rho_t, y_{t-1}^2)}{Var\left(y_t\right)}.$$

□

Proof of Proposition 3: Under assumptions in Proposition 2 the optimal forecast l-periods ahead, with $l > 1$, of the process $y_t = \alpha + \rho_t y_{t-1} + h_t \varepsilon_t$ are

$$\begin{gathered} E[y_{t+l}|\Im_t] = \alpha\left(1 + E[\rho_{t+l}|\Im_t] + E[\rho_{t+l}\rho_{t+l-1}|\Im_t] + \ldots + E[\rho_{t+l}\cdots\rho_{t+2}|\Im_t]\right) + \\ E[\rho_{t+l}\rho_{t+l-1}\cdots\rho_{t+2}|\Im_t]\rho_{t+1}y_t + E[\rho_{t+l}a_{t+l-1}|\Im_t] + \\ E[\rho_{t+l}\rho_{t+l-1}a_{t+l-2}|\Im_t] + \ldots + E[\rho_{t+l}\rho_{t+l-1}\cdots\rho_{t+2}a_{t+1}|\Im_t]. \end{gathered}$$

This expression can be simplified given that the shock sequence ε_t and in turn ρ_t are *iid*. Also, using that ρ_{t+1} and $\rho_{t+2}a_{t+1}$ are stationary sequences and a_{t+1} a martingale difference sequence the preceding expression reads as

$$E[y_{t+l}|\Im_t] = \alpha \sum_{i=0}^{l-1} E[\rho_{t+1}]^i + E[\rho_{t+1}]^{l-1}\rho_{t+1}y_t + \sum_{i=1}^{l-1} E[\rho_{t+i+1}a_{t+i}|\Im_t] E[\rho_{t+1}]^{l-i-1}.$$

Thus,

$$E[y_{t+l}|\Im_t] = \alpha \tfrac{1 - E[\rho_{t+1}]^{l-2}}{1 - E[\rho_{t+1}]} + E[\rho_{t+1}]^{l-1}\rho_{t+1}y_t + \sum_{i=1}^{l-1} E[\rho_{t+i+1}a_{t+i}|\Im_t] E[\rho_{t+1}]^{l-i-1}.$$

As $l \to \infty$ the optimal conditional forecast converges to the unconditional mean in L_2.

$$E[y_{t+l}|\Im_t] \xrightarrow{L_2} \frac{\alpha}{1 - E[\rho_{t+1}]} + \frac{E[\rho_{t+1}a_t]}{1 - E[\rho_{t+1}]}. \tag{26}$$

This is equivalent to show that $\left\| E[y_{t+l}|\Im_t] - \frac{\alpha + E[\rho_{t+1}a_t]}{1 - E[\rho_{t+1}]} \right\|_2 \to 0$. Note that it is sufficient to prove that

$$\sum_{i=1}^{l-1} \left\| E[h_{t+i}|\Im_t] - E\left[h_{t+i}\right] \right\|_2 \left| E[\rho_{t+1}] \right|^{l-i-1} \to 0, \tag{27}$$

when $l \to \infty$, given that $E[\rho_{t+1}a_t] = E[\rho_{t+1}\varepsilon_t] E\left[h_t\right]$. To prove this result we use asymptotic theory for $L_2 - mixingales$, defined as follows: a zero mean process x_t is a $L_2 - mixingale$ if $\left\| E\left[x_t|\Im_{t-m}\right] \right\|_2 \leq \left\| c_t \right\|_2 \gamma\left(m\right),$ with $\gamma\left(m\right) \to 0$, see Davidson (1994) or McLeish (1975). Using A.8 for $k = 2$ it can be proved that $x_t = h_t - E[h_t]$ is a L_2-Near Epoch Dependence (L_2-NED, see Davidson (1994) or McLeish (1975)) on ε_t of size $-\infty$. Further, from Theorem 17.5 of Davidson (1994) and A.8, the process x_t is a L_2-mixingale with $\|c_t\|_2 \leq c_h < \infty$ and $\gamma(m) = \lambda_3^m$ with $\lambda_3 = (\beta_1 + \beta_2) < 1$. Then, expression (27) has the following upper bound that satisfies

$$\sum_{i=1}^{l-1} \gamma(i) \|c_t\|_2 |E[\rho_{t+1}]|^{l-i-1} \leq \sum_{i=1}^{l-1} \lambda_4^{l-1} \|c_t\|_2 \leq (l-1) \lambda_4^{l-1} c_h,$$

with $\lambda_4 = \max\{\lambda_3, E(\rho_t)\}$. By A.7, $0 \leq \lambda_4 < 1$. Then, the upper bound goes to 0 and (26) immediately follows. □

Proof of Proposition 4: By Bayes' theorem

$$\begin{aligned} P\{y_t \leq -v, y_{t-1} \leq -v\} =& P\{y_t \leq -v \cap y_{t-1} \leq -v \cap \varepsilon_{t-1} \leq u_1\} \\ &+ P\{y_t \leq -v \cap y_{t-1} \leq -v \cap u_1 < \varepsilon_{t-1} \leq u_2\} \\ &+ P\{y_t \leq -v \cap y_{t-1} \leq -v \cap \varepsilon_{t-1} > u_2\}, \end{aligned}$$

where v denotes a positive threshold. By operating on the first expression on the right term we obtain the following result:

$$P_{t-2}\{y_t \leq -v \cap y_{t-1} \leq -v \cap \varepsilon_{t-1} \leq u_1\} = \int_{-\infty}^{x_{1t}} F_\varepsilon\left(\frac{-v - (\alpha + \rho_1(\alpha + \rho_{t-1}y_{t-2} + \varepsilon h_{t-1}))}{h_t(\varepsilon)}\right) f_\varepsilon(\varepsilon)\, \partial\varepsilon.$$

Operating in the same way with the other summands we obtain the first part of Proposition 4. The second part of the proposition is obtained in the same way, taking into account that

$$\begin{aligned} P_{t-2}\{y_t \geq v, y_{t-1} \geq v\} =& P_{t-2}\{y_{t-1} \geq v\} \\ &- P_{t-2}\{y_t \leq v | y_{t-1} \geq v, \varepsilon_{t-1} \leq u_1\} P_{t-2}\{y_{t-1} \geq v, \varepsilon_{t-1} \leq u_1\} \\ &- P_{t-2}\{y_t \leq v | y_{t-1} \geq v, u_1 < \varepsilon_{t-1} \leq u_2\} P_{t-2}\{y_{t-1} \geq v, u_1 < \varepsilon_{t-1} \leq u_2\} \\ &- P_{t-2}\{y_t \geq v | y_{t-1} \geq v, \varepsilon_{t-1} > u_2\} P_{t-2}\{y_{t-1} \geq v, \varepsilon_{t-1} > u_2\}, \end{aligned}$$

given that $P_{t-2}\{y_{t-1} \geq v, \varepsilon_{t-1} \leq u_1\} + P_{t-2}\{y_{t-1} \geq v, u_1 < \varepsilon_{t-1} \leq u_2\} + P_{t-2}\{y_{t-1} \geq v, \varepsilon_{t-1} > u_2\} = P_{t-2}\{y_{t-1} \geq v\}$. □

Proof of Proposition 5: To obtain the result in Proposition 5 it is sufficient to study the convergence in probability to zero of $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t [y_{t-1}(u) - \widehat{y}_{t-1}(u)]$. In particular we concentrate on showing this for one element of the vector, namely, $a_t\,[y_{t-1} I(\varepsilon_{t-1} \leq u) - y_{t-1} I(\varepsilon_{n,t-1} \leq u)]$; for the remaining terms the result is analogous. This process, using further notation in Koul and Ling (2006), can be expressed as

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t y_{t-1} \left[I(\varepsilon_{t-1} \leq u) - I(\varepsilon_{t-1} \leq u + u b_{n,t-1}(\widehat{x}) + c_{n,t-1}(\widehat{x}))\right]$$

where $\widehat{x} = n^{1/2}\left(\widehat{\phi}_{n,H_0} - \phi_{0,H_0}\right)$, $b_{n,t}(\widehat{x}) = \frac{h_t\left(\phi_{0,H_0} + n^{-1/2}\widehat{x}\right) - h_t\left(\phi_{0,H_0}\right)}{h_t\left(\phi_{0,H_0}\right)}$, $c_{n,t}(\widehat{x}) = \frac{\mu_t\left(\phi_{0,H_0} + n^{-1/2}\widehat{x}\right) - \mu_t\left(\phi_{0,H_0}\right)}{\mu_t\left(\phi_{0,H_0}\right)}$, and $\mu_t(\phi_{0,H_0})$ the $AR(1)$ mean process. In order to prove this result we need to show the weak convergence of the empirical process

$$K_n(u, x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t y_{t-1} \Psi_{t-1}(\varsigma), \quad \varsigma \in H,$$

where $\varsigma = (u, x)$; $H := (\underline{u}, \overline{u}) \times \left(\underline{b}, \overline{b}\right)^5$ and $\Psi_{t-1}(\varsigma) := I(\varepsilon_{t-1} \leq u + ub_{n,t-1}(x) + c_{n,t-1}(x))$, with $b_{n,t}$and $c_{n,t}$ random variables which are $\Im_{t-1}$ measurable. For the rest of the proof we define the following distance, $d(X, Y) = \max_{1 \leq i \leq k} |X_i - Y_i|^{1/4}$ with $X, Y \in R^k$.

Lemma A1.- If the following assumptions W.1 and W.2 hold,

W.1 For each $n \geq 1$ and $\varsigma \in H$, $\{(a_t, y_{t-1}, b_{n,t-1}(x), c_{n,t-1}(x)) : 1 \leq t \leq n\}$ is a strictly stationary and ergodic process. The sequence $\{a_t y_{t-1} \Psi_{t-1}(\varsigma), \Im_{t-1}, 1 \leq t \leq n\}$ is a square-integrable martingale difference sequence for each $\varsigma \in H$. Also, there exists a function $C(\varsigma_1, \varsigma_2)$ on $H \times H$ to R such that uniformly in $(\varsigma_1, \varsigma_2) \in H \times H$,

$$n^{-1} \sum_{t=1}^{n} a_t^2 y_{t-1}^2 \Psi_{t-1}(\varsigma_1) \Psi_{t-1}(\varsigma_2) = C(\varsigma_1, \varsigma_2) + o_P(1). \tag{28}$$

W.2 For every $\delta > 0$ there exists a finite partition $B_\delta = \{H_k; 1 \leq k \leq N(\delta, H, d)\}$ of H with $N_\delta := N(\delta, H, d)$ being the elements of such partition, such that

$$\int_0^\infty \sqrt{\log(N_\delta)} d\delta < \infty, \tag{29}$$

and

$$\sup_{\delta \in (0,1) \cap Q} \frac{CV_n(B_\delta)}{\delta^2} = O_P(1), \tag{30}$$

with

$$CV_n(B_\delta) = \max_k n^{-1} \sum_{t=1}^{n} E\left[\sup_{(\varsigma_1, \varsigma_2) \in H_k \times H_k} |a_t y_{t-1} \Psi_{t-1}(\varsigma_1) - a_t y_{t-1} \Psi_{t-1}(\varsigma_2)|^2 \middle| \Im_{t-1} \right], \tag{31}$$

then it follows that

$$K_n(\varsigma) \Longrightarrow K_\infty(\varsigma),$$

with $K_\infty(\varsigma)$ a Gaussian process with zero mean and covariance function given by $C(\varsigma_1, \varsigma_2)$.

Proof of Lemma A1. It follows from Theorem A1 in Delgado and Escanciano (2007). □

A consequence of the asymptotic tightness of $K_n(\varsigma)(= K_n(u, x))$ is that if $\widetilde{x}$ converges in distribution to x_0, then

$$\sup_{u \in U} |K_n(u, \widetilde{x}) - K_n(u, x_0)| = o_P(1).$$

Therefore, for $\widetilde{x} = n^{1/2}\left(\widehat{\phi}_{n,H_0} - \phi_{0,H_0}\right)$, under H_0, conditions W.1 and W.2, and given that $\sup_{1 \leq t \leq n} b_{n,t}(x_0) = o_P(1)$ and $\sup_{1 \leq t \leq n} c_{n,t}(x_0) = o_P(1)$, it follows from Lemma A1 that

$$\sup_{u \in U} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t y_{t-1} \left[I(\varepsilon_{t-1} \leq u) - I(\varepsilon_{t-1} \leq u + ub_{n,t-1}(x_0) + c_{n,t-1}(x_0)) \right] \right| = o_P(1).$$

This and the previous results prove that

$$\sup_{u \in U} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t y_{t-1} \left[I(\varepsilon_{t-1} \leq u) - I(\varepsilon_{n,t-1} \leq u) \right] \right| = o_P(1),$$

and

$$\sup_{u \in U} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t [y_{t-1}(u) - \widehat{y}_{t-1}(u)] \right| = o_p(1)$$

which proves Proposition 5. We only need to show that conditions W.1 and W.2 are satisfied in our case.

In particular, the proof of (28) follows from the assumptions in Theorem 1 and a uniform law of large numbers, see Jennrich (1969, Theorem 2). The proof of W.2 is more cumbersome. From the compactness of H we have

$$\int_0^\infty \sqrt{\log\left(N\left(\delta, H, d\right)\right)} d\delta < \infty.$$

Now, let B_δ be a partition of H $\delta-$balls with respect to d. Thus (29) holds for such partition. We shall prove that (30) is also satisfied for such partition. Given that ε_t does not depend on ς and is iid, $CV_n\left(B_\delta\right)$ in (31) is upper bounded by

$$\max_{1\leq k\leq N_\delta} n^{-1}\sum_{t=2}^{n} E\left(a_t^2 y_{t-1}^2\middle|\Im_{t-1}\right) \sup_{(\varsigma_1,\varsigma_2)\in H_k}\left(\Psi_t\left(\varsigma_1\right)-\Psi_t\left(\varsigma_2\right)\right)^2 \leq$$
$$\left(n^{-1}\sum_{t=2}^{n} E^2\left(a_t^2 y_{t-1}^2\middle|\Im_{t-1}\right)\right)^{1/2}\left(\max_{1\leq k\leq N_\delta}\sup_{(\varsigma_1,\varsigma_2)\in H_k} n^{-1}\sum_{t=2}^{n}\left(\Psi_t\left(\varsigma_1\right)-\Psi_t\left(\varsigma_2\right)\right)^4\right)^{1/2}.$$

From conditions in Proposition 1 for $k=8$ we obtain that $E\left[E^2\left(a_t^2 y_{t-1}^2\middle|\Im_{t-1}\right)\right]<\infty$, and therefore, $n^{-1}\sum_{t=2}^{n} E^2\left(a_t^2 y_{t-1}^2\middle|\Im_{t-1}\right)=O_P\left(1\right)$. Now, for the second right term let us define the following difference of indicators:

$$\begin{aligned}\widetilde{\Psi}_t\left(\varsigma_k,\delta\right) = {} & I(\varepsilon_t \leq u_k+\delta^4\left(1+b_{n,t-1}\left(x_k\right)+v_{n,t-1}\left(x_k\right)\right)+u_k b_{n,t-1}\left(x_k\right)+u_k v_{n,t-1}\left(x_k\right)+c_{n,t-2}\left(x_k\right)+w_{n,t-1}\left(x_k\right))- \\ & -I(\varepsilon_t \leq u_k-\delta^4\left(1+b_{n,t-1}\left(x_k\right)-v_{n,t-1}\left(x_k\right)\right)+u_k b_{n,t-1}\left(x_k\right)-u_k v_{n,t-1}\left(x_k\right)+c_{n,t-2}\left(x_k\right)-w_{n,t-1}\left(x_k\right)),\end{aligned}$$

where $v_{n,t}\left(x\right) := \sup_{d(x,s)<\delta}\left|b_{n,t}\left(x\right)-b_{n,t}\left(s\right)\right|$ and $w_{n,t}\left(x\right) := \sup_{d(x,s)<\delta}\left|c_{n,t}\left(x\right)-c_{n,t}\left(s\right)\right|$. We also introduce the conditional distribution function $F_{t-1}\left(\varsigma_k,\delta\right)=E\left(\widetilde{\Psi}_t\left(\varsigma_k,\delta\right)\middle|\Im_{t-1}\right)$. Without loss of generality suppose that $u_k\geq 0$ (for the opposite case the expression can be bounded in a similar way). Using the monotonicity of the indicator function we obtain

$$\begin{aligned}\max_{1\leq k\leq N_\delta}\sup_{(\varsigma_1,\varsigma_2)\in H_k} n^{-1}\sum_{t=2}^{n}\left(\Psi_t\left(\varsigma_1\right)-\Psi_t\left(\varsigma_2\right)\right)^4 & \leq \max_{1\leq k\leq N_\delta} n^{-1}\sum_{t=2}^{n}\widetilde{\Psi}_t\left(\varsigma_k,\delta\right) \\ & \leq \max_{1\leq k\leq N_\delta} n^{-1}\sum_{t=2}^{n}\left[\widetilde{\Psi}_t\left(\varsigma_k,\delta\right)-F_{t-1}\left(\varsigma_k,\delta\right)\right]+\max_{1\leq k\leq N_\delta} n^{-1}\sum_{t=2}^{n}F_{t-1}\left(\varsigma_k,\delta\right).\end{aligned}$$

Now the first term of the last inequality is the sum of a martingale difference sequence and can be easily proved that is $o_p\left(1\right)$. The second term, using that $\max_e f_\varepsilon\left(e\right)\leq\overline{f}<\infty$, by $A.1$, and the mean value theorem, can be upper bounded as

$$\begin{aligned}\max_{1\leq k\leq N_\delta} n^{-1}\sum_{t=2}^{n}F_{t-1}\left(\varsigma_k,\delta\right) & \leq 2\overline{f}\max_{1\leq k\leq N_\delta} n^{-1}\sum_{t=2}^{n}\left[\delta^4\left(1+b_{n,t-1}\left(x_k\right)\right)+u_k v_{n,t-1}\left(x_k\right)+w_{n,t-1}\left(x_k\right)\right] \\ & \leq 2\overline{f}\delta^4+o_p(1),\end{aligned}$$

where the last inequality follows from assuming that

$$E\left(\max_{1\leq k\leq N_\delta} b_{n,t-1}\left(x_k\right)\right)=o_P\left(1\right)$$
$$E\left(\max_{1\leq k\leq N_\delta} v_{n,t-1}\left(x_k\right)\right)=o_P\left(1\right)$$
$$E\left(\max_{1\leq k\leq N_\delta} w_{n,t-1}\left(x_k\right)\right)=o_P\left(1\right).$$

These last three conditions are shown in Koul (2002) in a similar context. With this we have proved that

$$CV_n(B_\delta) = \delta^2 O_P(1),$$

and therefore (30) is satisfied. □

Proof of Theorem 2: From expression (15) and the result in Proposition 5 we have that

$$\sqrt{n}(\widehat{\gamma}_n(u) - \gamma) \xrightarrow{d} N(0, \Sigma(u)). \tag{32}$$

By a law of large numbers and arguments similar to those in Proposition 5, $\widehat{\Sigma}_n(u)$ is a consistent estimator of the asymptotic variance $\Sigma(u)$. Therefore, $\sqrt{n}\widehat{\Sigma}^{-1/2}(u)(\widehat{\gamma}_n(u) - \gamma) \xrightarrow{d} N(0,1)$. It is immediate to see now that $\widehat{T}_n(u)$ converges weakly to a centered χ^2_2 distribution. □

Proof of Theorem 3: The proof immediately follows from Theorem 2, assumptions A.9 and A.10, and the tightness of the process $S_n(u)$ that follows from the uniform convergence in Proposition 5. The weak convergence of the supremum and average tests follows from the continuous mapping theorem. □

Proof of Theorem 4: Given assumption $A.13$ the consistency of $\left(\widehat{\phi}, \widehat{u}_n\right)$ is given by the following result

$$\sup_{\phi, u \in \Theta} |L_n(\phi, u) - E[L_n(\phi, u)]| \xrightarrow{p} 0, \tag{33}$$

where Θ is the parametric space defined in $A.12$. Define $B(\phi, u, \delta) = \left\{\widetilde{\phi}, \widetilde{u} \in \Theta : d\left((\phi, u), \left(\widetilde{\phi}, \widetilde{u}\right)\right) \leq \delta\right\}$ with $d(.)$ any distance, and

$$l_t^*(\phi, u) \geq \sup\left\{l_t\left(\widetilde{\phi}, \widetilde{u}\right) : \widetilde{\phi}, \widetilde{u} \in B(\phi, u, \delta)\right\}$$
$$l_{*t}(\phi, u) \leq \inf\left\{l_t\left(\widetilde{\phi}, \widetilde{u}\right) : \widetilde{\phi}, \widetilde{u} \in B(\phi, u, \delta)\right\}.$$

The result in (33) follows from the following conditions (see Theorem of Andrews 1987):

C.1 Θ is a compact metric space.

C.2

a) $l_t(\phi, u), l_t^*(\phi, u)$ and $l_{*t}(\phi, u)$ are random variables, $\forall \phi, u \in \Theta, \forall t$ and $\forall \delta$ sufficiently small (where δ may depend on ϕ, u).

b) $l_t^*(\phi, u)$ and $l_{*t}(\phi, u)$ satisfy pointwise strong (weak) laws of large numbers, $\forall \phi, u \in \Theta$ and $\forall \delta$ sufficiently small (where δ may depend on ϕ, u).

C.3 For all $\phi, u \in \Theta$

$$\lim_{\delta \to 0} \sup_{n \geq 1} \left| n^{-1} \sum_{t=1}^{n} \left[E\left(l_t^*(\phi, u)\right) - E\left(l_t(\phi, u)\right)\right] \right|$$

and likewise with $l_t^*(\phi, u)$ replaced by $l_{*t}(\phi, u)$.

Now we prove conditions $C.1 - C.3$ for our model using the assumptions of Theorem 4. Condition $C.1$ can be proved straightforward from the definition of Θ. Moreover, it can be proved that $\max_{1 \leq j \leq 3} |\rho_j| <$

$\overline{\rho} < \infty$ and $0 < \beta_2 < 1$. Now we define $l_t^* (\phi, u)$ and $l_{*t} (\phi, u)$ in our case. For that, we need the following processes:

$$\begin{aligned}
d_{at} (\phi, u, \delta) &:= \delta C + M (|\rho_1| + |\rho_2|) \bar{I}_{1t} (\phi, u) + M (|\rho_2| + |\rho_3|) \bar{I}_{2t} (\phi, u) ; \\
d_{ht} (\phi, u, \delta) &:= \frac{\delta C}{1 - \beta_2^{1/2}} + \sum_{i=1}^{t-1} \beta_2^{\frac{i-1}{2}} \beta_1^{1/2} d_{at-i} (\phi, u, \delta) , \\
\overline{I}_{it+1} (\phi, u) &:= 1 (u_i h_t (\phi, u) - \Lambda (\phi, u, \delta) < a_t (\phi, u) < u_i h_t (\phi, u) + \Lambda (\phi, u, \delta)) ; \\
\Lambda (\phi, u, \delta) &:= u_i d_{ht} (\phi, u, \delta) + \delta C \overline{h} + d_{at} (\phi, u, \delta) ; \\
\overline{h}^2 &:= \frac{\overline{\beta}_0 + \overline{\beta}_1 (\overline{\alpha} + 2M)^2}{1 - \overline{\beta}_2},
\end{aligned}$$

with C any finite constant independent of ϕ, $\overline{\beta}_2 = \sup \left\{ \left| \widetilde{\beta}_2 \right| : \widetilde{\phi} \in B (\phi, u, \delta) \right\}$. Then it follows from above that there exists some $C < \infty$ such that

$$\sup_{\widetilde{\phi}, \widetilde{u} \in B(\phi,u,\delta)} \left| a_t (\phi, u) - a_t \left(\widetilde{\phi}, \widetilde{u} \right) \right| \leq d_{at} (\phi, u, \delta) \quad \text{and} \quad \sup_{\widetilde{\phi}, \widetilde{u} \in B(\phi,u,\delta)} \left| h_t (\phi, u) - h_t \left(\widetilde{\phi}, \widetilde{u} \right) \right| \leq d_{ht} (\phi, u, \delta) .$$

Using the law of iterated expectation (LIE) and the Mean Value Theorem it is easy to prove that

$$\left\| \overline{I}_{it} (\phi, u) \right\|_1 \leq 2\overline{f} \left[\delta C_1 + u_i \left\| d_{ht-1} (\phi, u, \delta) \right\|_1 + \left\| d_{at-1} (\phi, u, \delta) \right\|_1 \right]$$

with $C_1 < \infty$. Using again the LIE and the previous result

$$\begin{aligned}
\left\| d_{at} (\phi, u, \delta) \right\|_1 &\leq \delta C_2 + 2\overline{f} (\lambda_1 u_1 + \lambda_2 u_2) \left\| d_{ht-1} (\phi, u, \delta) \right\|_1 + 2\overline{f} (\lambda_1 + \lambda_2) \left\| d_{at-1} (\phi, u, \delta) \right\|_1 \\
\left\| d_{ht} (\phi, u, \delta) \right\|_1 &\leq \frac{\delta C_3}{1 - \beta_2^{1/2}} + \sum_{i=1}^{t-1} \beta_2^{\frac{i-1}{2}} \beta_1^{1/2} \left\| d_{at-i} (\phi, u, \delta) \right\|_1
\end{aligned}$$

with $C_2, C_3 < \infty$. Therefore

$$\begin{aligned}
\left\| d_{at} (\phi, u, \delta) \right\|_1 &\leq \delta C_4 + 2\overline{f} (\lambda_1 + \lambda_2) \left\| d_{at-1} (\phi, u, \delta) \right\|_1 + \sum_{i=2}^{t-1} 2\overline{f} (\lambda_1 u_1 + \lambda_2 u_2) \beta_2^{\frac{i-2}{2}} \beta_1^{1/2} \left\| d_{at-i} (\phi, u, \delta) \right\|_1 \\
&\leq \delta C_5 + \pi_1 \left\| d_{at-1} (\phi, u, \delta) \right\|_1 + \sum_{i=2}^{t-1} \pi_2 \pi_3^{i-2} \left\| d_{at-i} (\phi, u, \delta) \right\|_1
\end{aligned}$$

with $C_4, C_5 < \infty$. Rewriting the previous equation

$$\left\| d_{at} (\phi, u, \delta) \right\|_1 \leq \delta C_5 (1 - \pi_3) + (\pi_1 + \pi_3) \left\| d_{at-1} (\phi, u, \delta) \right\|_1 + (\pi_2 - \pi_1 \pi_3) \left\| d_{at-2} (\phi, u, \delta) \right\|_1 ,$$

and therefore, under the conditions of assumption $A.12$ it can be shown that there exits $C_a, C_h < \infty$ such that

$$\begin{aligned}
\left\| d_{at} (\phi, u, \delta) \right\|_1 &\leq \delta C_a \\
\left\| d_{ht} (\phi, u, \delta) \right\|_1 &\leq \delta C_h .
\end{aligned}$$

Now define

$$
\begin{aligned}
l_t^*(\phi,u) &:= -\frac{1}{2}\ln h_t^2(\phi,u) + 2\frac{d_{ht}(\phi,u,\delta)}{\underline{\beta}_0} - \frac{1}{2}e_t^2(\phi,u) + 2(2M+\overline{\alpha})\overline{h}\frac{d_{at}(\phi,u,\delta)+(2M+\overline{\alpha})d_{ht}(\phi,u,\delta)}{\underline{\beta}_0^2} \\
&= l_t(\phi,u) + 2(2M+\overline{\alpha})\overline{h}\underline{\beta}_0^{-2}d_{at}(\phi,u,\delta) + 2\left[(2M+\overline{\alpha})^2\overline{h}+\underline{\beta}_0\right]\underline{\beta}_0^{-2}d_{ht}(\phi,u,\delta) \\
l_{*t}(\phi,u) &:= -\frac{1}{2}\ln h_t^2(\phi,u) - 2\frac{d_{ht}(\phi,u,\delta)}{\underline{\beta}_0} - \frac{1}{2}e_t^2(\phi,u) - 2(2M+\overline{\alpha})\overline{h}\frac{d_{at}(\phi,u,\delta)+(2M+\overline{\alpha})d_{ht}(\phi,u,\delta)}{\underline{\beta}_0^2} \\
&= l_t(\phi,u) - 2(2M+\overline{\alpha})\overline{h}\underline{\beta}_0^{-2}d_{at}(\phi,u,\delta) - 2\left[(2M+\overline{\alpha})^2\overline{h}+\underline{\beta}_0\right]\underline{\beta}_0^{-2}d_{ht}(\phi,u,\delta).
\end{aligned}
$$

Thus, there exists $C^* < \infty$ such that

$$
\begin{aligned}
\lim_{\delta\to 0}\sup_{n\geq 1}\left|n^{-1}\sum_{t=1}^{n}\left[E\left(l_t^*(\phi,u)\right)-E\left(l_t(\phi,u)\right)\right]\right| &\leq \frac{2\left[(2M+\overline{\alpha})^2\overline{h}+\underline{\beta}_0\right]}{\underline{\beta}_0^2}\lim_{\delta\to 0}\sup_{n\geq 1}\left|n^{-1}\sum_{t=1}^{n}E\left(d_{ht}(\phi,u,\delta)\right)\right| + \\
&\quad \frac{2(2M+\overline{\alpha})\overline{h}}{\underline{\beta}_0^2}\lim_{\delta\to 0}\sup_{n\geq 1}\left|n^{-1}\sum_{t=1}^{n}E\left(d_{at}(\phi,u,\delta)\right)\right| \\
&\leq \frac{2\left[\left((2M+\overline{\alpha})+1\right)(2M+\overline{\alpha})\overline{h}+\underline{\beta}_0\right]}{\underline{\beta}_0^2}\lim_{\delta\to 0}\delta C^* = 0.
\end{aligned}
$$

The same result can be shown for $l_{*t}(\phi,u)$. With this result condition $C.3$ is proven. Finally, to prove $C.2$ we prove that $l_t^*(\phi,u)$ and $l_{*t}(\phi,u)$ with mean μ_t^* and μ_{*t}, respectively, are $L_1 - NED$ of size $-\infty$ with constants $d_t^* \ll \|l_t^*(\phi,u)-\mu_t^*\|$ and $d_{*t} \ll \|l_{*t}(\phi,u)-\mu_{*t}\|$ on ε_t, which is $\alpha - mixing$ of size $-\infty$ (given that these are iid). The details of the proof are similar to those in Davidson (2002) and hence omitted for sake of space. With this result we can apply Theorem 20.19 of Davidson (1994) for $a_t = t$ and $q = 2$. Note, however, that to use this theorem we must prove

$$
\sum_{t=1}^{\infty}\left\|\frac{l_t^*(\phi,u)-\mu_t^*}{t}\right\|_2^{4/3} < \infty,
$$

and the same result for $l_{*t}(\phi,u)-\mu_{*t}$. This inequality is upper bounded as follows:

$$
\sum_{t=1}^{\infty}\left\|\frac{l_t^*(\phi,u)-\mu_t^*}{t}\right\|_2^{4/3} \leq \sum_{t=1}^{\infty}t^{-4/3}2\left\|l_t^*(\phi,u)\right\|_2^{4/3} \leq \max_t 2\left\|l_t^*(\phi,u)\right\|_2^{4/3}\sum_{t=1}^{\infty}t^{-4/3} < \infty,
$$

given that $\max_t 2\left\|l_t^*(\phi,u)\right\|_2 < \infty$ from previous results. The same result applies to $l_{*t}(\phi,u)-\mu_{*t}$. Then, from Theorem 20.19 of Davidson (1994)

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left(l_t^*(\phi,u)-\mu_t^*\right) &\xrightarrow{a.s.} 0, \\
\frac{1}{T}\sum_{t=1}^{T}\left(l_{*t}(\phi,u)-\mu_{*t}\right) &\xrightarrow{a.s.} 0,
\end{aligned}
$$

satisfying both processes a strong law of large numbers and proving $C.2$ b). □

Acknowledgments

The authors thank participants of Workshop in Computational and Financial Econometrics held in Geneva, 2007, NBER Time Series Conference held in Iowa, 2007 and participants of XXXII Spanish Economic Simposio in Granada. and anonymous referees for their useful and stimulating suggestions and comments that led to a considerably improved and more focused version of the article. Any remaining errors are the authors' own. Oscar Martinez acknowledges financial support from the Spanish Ministerio de Educación y Ciencia, reference numbers SEJ2007-63098/ECON and SEJ2007-64605/ECON, and Jose Olmo from the Spanish Ministerio de Educación y Ciencia, reference number SEJ2007-63098/ECON.

References

[1] Andersen, T.G., Bollerslev, T., (1998). Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review* 39, 885-905.

[2] Andrews, D.W.K., (1987). Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers. *Econometrica* 55, 1465-1471.

[3] Andrews, D.W.K., Ploberger, W., (1994). Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative. *Econometrica* 62, 1383-1414.

[4] Bollerslev, T., (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* 31, 307-327.

[5] Brandt, A., (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Probab.,* 18, 211–220.

[6] Chan, K. S., (1990). Testing for Threshold Autoregression. *Annals of Statistics* 18, 1886-1894.

[7] Chan, K. S., (1993). Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model. *Annals of Statistics* 21, 520-533.

[8] Davidson, J.E.H., (1994). Stochastic Limit Theory. Oxford: Oxford University Press.

[9] Davidson, J.E.H., (2002). Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics* 106, 243-269.

[10] Davies, R.B., (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247-254.

[11] Davies, R.B., (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33-43.

[12] Delgado, M.A., Escanciano, J.C., (2007). Nonparametric Tests for Conditional Symmetry in Dynamic Models. *Journal of Econometrics* 141, 652-682.

[13] Engle, R.F., (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1007.

[14] Engle, R.F., Smith, A.D., (1999). Stochastic Permanent Breaks. *Review of Economics and Statistics* 84, 553-574.

[15] Glosten, L.R., Jagannatthan, R., and Runkle, D.E., (1993). On the relation between the expected value and the volatility of the nominal excess returns on stocks. *Journal of Finance* 48, 1779-1801.

[16] Gonzalez, M., Gonzalo, J., (1998) Threshold Unit Root Processes. Working Paper.

[17] Gonzalo, J., Martinez, O., (2006). Large shocks vs. small shocks. Or does size matter? May be so. *Journal of Econometrics* 127, 311-347.

[18] Gonzalo, J., Martinez, O., (2007). SHOCKS. Do you want to identify them? Working Paper.

[19] Gonzalo, J., Pitarakis, J., (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics* 110, 319 − 352.

[20] Gonzalo, J., Wolf, M., (2005). Subsampling Inference in Threshold Autoregressive Models. *Journal of Econometrics* 127, 201 − 224.

[21] Granger, C.W.J., Terasvirta, T., (1993). Modelling Nonlinear Economic Relationships. Oxford University Press, Oxford.

[22] Hamilton, J.D., (1989). A new approach to the economic analysis of nonstationarity time series and the business cycle. *Econometrica* 57, 357-384.

[23] Hansen, B.E., (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413-430.

[24] Hansen, B.E., (1997). Inference in TAR Models. *Studies in Nonlinear Dynamics & Econometrics* 2 (1), 1-14.

[25] Hansen, B.E., (2000). Sample Splitting and Threshold Estimation. *Econometrica* 68, 575-603.

[26] Harvey, A.C., Ruiz, E., and Shepard, N., (1994). Multivariate stochastic variance models. *Review of Economic Studies* 61, 247-264.

[27] Jennrich, R.I., (1969). Asymptotic Properties of Nonlinear Least Square Estimators. *Annals of Mathematical Statistics* 40, 633-643.

[28] Koul, L.H., (2002). Weighted Empirical Processes in Dynamic Nonlinear Models. Lecture Notes in Statistics. Springer.

[29] Koul, L.H., Ling, S., (2006). Fitting an error distribution in some heterocedastic time series models. *The Annals of Statistics* 34, 994-1012.

[30] Kristensen, D., (2006). On Stationarity and Ergodicity of the Bilinear Model with Applications to GARCH Models. Mimeo.

[31] Ling, S., McAleer, M., (2003). Asymptotic Theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 280-310.

[32] McLeish, D.L., (1975). A maximal inequality and dependent strong laws. *Annals of Probability 3, 329-339.*

[33] Nelson, D.B., (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347-370.

[34] Politis, D. N., Romano, J.P., and Wolf., M., (1999). Subsampling. Springer-Verlag, New York.

[35] Taylor, S.J., (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance* 4, 183-204.

[36] Teräsvirta, T., (1994). Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association* 89, 208-218.

[37] Tong, H., (1978). On a threshold model. In C.H. Chen (ed.), Pattern Recognition and Signal Processing. Sijhoff & Noordhoff, Amsterdam.

[38] Tong, H., (1983). Threshold Models in Non-Linear Time Series Analysis. Lecture Notes in Statistics, 21. Springer, Berlin.

[39] Tong, H., (1990). Non-Linear Time Series: A Dynamical System Approach. Oxford University Press, Oxford.

[40] Tong, H., Lim, K.S., (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B,* 4, 245-292.

[41] Tsay, R.S., (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association* 84, 231-240.

[42] Zakoian, J.M., (1994). Threshold heteroscedastic models. *Journal of Economic Dynamics and Control* 18, 931-955.