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

This draft, August 2007

Abstract

One of the main implications of the efficient market hypothesis (EMH) is that expected future returns on financial assets are not predictable if investors are risk neutral. In this paper we argue that financial time series offer more information than that this hypothesis seems to supply. In particular we postulate that runs of very large returns can be predictable for small time periods. In order to prove this we propose a TAR(3,1)-GARCH(1,1) model that is able to describe two different types of extreme events: a first type generated by large uncertainty regimes where runs of extremes are not predictable and a second type where extremes come from isolated dread/joy events. This model is new in the literature in nonlinear processes. Its novelty resides on two features of the model that make it different from previous TAR methodologies. The regimes are motivated by the occurrence of extreme values and the threshold variable is defined by the shock affecting the process in the preceding period. In this way this model is able to uncover dependence and clustering of extremes in high as well as in low volatility periods. This model is tested with data from General Motors stocks prices corresponding to two crises that had a substantial impact in financial markets worldwide; the Black Monday of October 1987 and September 11th, 2001. By analyzing the periods around these crises we find evidence of statistical significance of our model and thereby of predictability of extremes for September 11^{th} but not for Black Monday. These findings support the hypotheses of a big negative event producing runs of negative returns in the first case, and of the burst of a worldwide stock market bubble in the second example.

JEL classification: C12; C15; C22; C51

Keywords and Phrases: asymmetries, crises, extreme values, hypothesis testing, leverage effect, nonlinearities, threshold models

*Research of both authors is funded by the Spanish Ministerio de Educación y Ciencia, reference number SEJ2004-04101/ECON.

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 filtration 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 a higher likelihood of extreme values in the positive tail for example. This property however is challenging for example for describing periods where the sequence exhibits runs of extremes of opposite sign than the mean. In this case these ARMA-GARCH processes should consider distributions with heavier tails than the Gaussian. Despite of this allowance, these models are not good candidates either to describe the data if there is no observed correlation in the mean of the sequence under study. This is particularly relevant for financial data.

An alternative widely explored that extends the 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 distribution errors 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 Granger and Teräsvista (1993) that provide a general survey. 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, 1999) or Gonzalo and Pitarakis (2002). Other possibility is Smooth Transition Models (STAR) that have an infinite number of regimes and 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 we propose is a TAR(3,1)-GARCH(1,1) process defined as follows:

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

with ε_t denoting the shock term, u_1 and u_2 threshold values defining the TAR (3,1) model, and h_t describing a GARCH (1,1) process for the volatility dynamics in the error term.

If the thresholds u_1 and u_2 define the lower and upper bound respectively of the sequence of extreme values of ε_t this model is able to distinguish two different types of extreme events in y_t. A first type generated by large uncertainty regimes where runs of extremes are not predictable (middle regime) and a second type where extremes come from isolated dread/joy events (extreme shocks). In this way the 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 bursts in financial markets, periods of peaks in energy prices due to sudden weather variations as for example cold snaps, or periods of underpriced/overpriced currencies due to large country-related shocks and reflected in extreme values of the sequence of exchange rates.

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 discusses the estimation of the parameters of the model and of the threshold variable, and hypothesis testing, this understood as statistical significance of the threshold effect against pure GARCH(1,1) and AR(1)-GARCH(1,1) models. This section also presents a Monte-Carlo analysis of the performance of size and power of the statistical test for finite samples. Section 4 discusses the suitability of these models to describe the so-called stylized facts of financial returns. Section 5 introduces an application of the methodology to measure the effect on General Motors (GM) stock prices of two crises that had a substantial impact in financial markets worldwide; the Black Monday of October 1987 and 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. A general model can be described as follows:

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

with u_1 and u_2 threshold values defining the TAR (3,1) model, and h_t describing a GARCH (1,1) process for the volatility dynamics in the error term. This is denoted by a_t and defined as $a_t = h_t \varepsilon_t$ with

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

and $\{\varepsilon_t\}$ a sequence of random shocks following a distribution function $F_\varepsilon(\cdot)$ with mean zero and variance one.

For the analysis of financial returns we focus on the following model:

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

with $\rho_t = \rho_1 I(\varepsilon_{t-1} < u_1) + \rho_3 I(\varepsilon_{t-1} > u_2)$, 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. In this case the threshold values are time varying and depend on the volatility regime. The model is as follows

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

with $u^*_{j,t} = h_{t-1} u_j$, $j = 1, 2$, threshold values that depend on the conditional volatility process. Note from this representation of the model that this process allows to identify extreme observations in low volatility regimes for the size of the preceding shocks determines the occurrence of extreme values and not previous observations of the sequence y_t as in SETAR methodologies.

This process accommodates many different dependence structures. Some examples are plotted below. Note that these processes could well describe time series of financial returns.

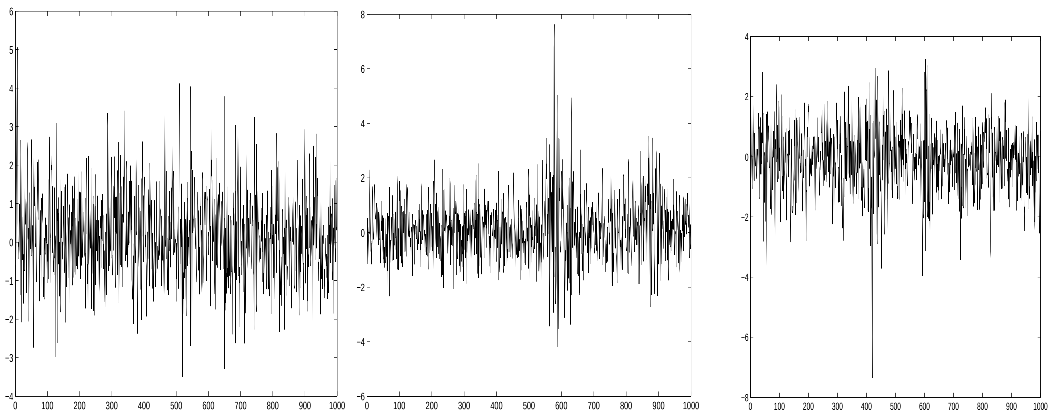

Figure 2.1. *Time series representing three different TAR(3,1)-GARCH(1,1) processes. The left panel depicts a process with* $\alpha = 0$, $(\rho_1, \rho_2, \rho_3) = (-0.7, 0, 0.7)$ *and* $(\beta_0, \beta_1, \beta_2) = (1, 0, 0)$. *The parameters of the middle panel are* $\alpha = 0$, $(\rho_1, \rho_2, \rho_3) = (-0.7, 0, 0.7)$ *and* $(\beta_0, \beta_1, \beta_2) = (0.05, 0.10, 0.85)$, *and for the right panel* $\alpha = 0$, $(\rho_1, \rho_2, \rho_3) = (0.7, 0, 0.7)$ *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* $n = 1000$.

Other assumptions on the model are

Assumptions.-

A.1 $\{\varepsilon_t\}$ is *IID*

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

A.3 $E\left[\max\left(0, log|\beta_1\varepsilon_t^2 + \beta_2|\right)\right] < \infty$

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

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

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

The 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. Then we are prepared to introduce Theorem 1.

Theorem 1.- *Assume that **A.1** to **A.4** and **A.6** hold, then the 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.*

All proofs are gathered in the mathematical appendix. The following proposition sets the conditions to ensure that the first k statistical moments of (3) are finite.

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

A.7 $E\left[\|\rho_t\|^k\right] < 1,$

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.*

Now we present the first two moments of model (3).

Proposition 2.- *Under assumptions in Proposition 1 for $k = 2$, the first two statistical moments of the process $y_t = \alpha + \rho_t y_{t-1} + h_t\varepsilon_t$ are*

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

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

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

Note that the randomness of the autoregressive parameter adds one extra term $\frac{E[\rho_t a_{t-1}]}{1-E[\rho_t]}$ in the unconditional mean, and $\frac{Cov\left(\rho_t^2, y_{t-1}^2\right) - E^2[\rho_t y_{t-1}]}{1-E[\rho_t^2]}$ in the unconditional variance. Further, in contrast to standard AR, TAR processes a zero intercept does not necessarily imply a zero unconditional mean.

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-GARCH methodology. Thus, the optimal forecasts one-period ahead of y_t using the mean square prediction error criterion are

$$E[y_{t+1}|\Im_t] = \rho_{t+1} y_t, \tag{9}$$

and

$$V[y_{t+1}|\Im_t] = h_{t+1}^2. \tag{10}$$

For longer forecast horizons these expressions take more involved forms.

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*

$$E[y_{t+l}|\Im_t] = \alpha \frac{1 - E[\rho_{t+1}]^{l-2}}{1 - E[\rho_{t+1}]} + E[\rho_{t+1}]^{l-1} \rho_{t+1} y_t + \sum_{i=1}^{l-1} E[\rho_{t+i+1} a_{t+i}|\Im_t] E[\rho_{t+1}]^{l-i-1}. \tag{11}$$

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

$$E[y_{t+l}|\Im_t] \xrightarrow{L_2} \frac{\alpha}{1 - E[\rho_{t+1}]} + \frac{E[\rho_{t+1} a_t]}{1 - E[\rho_{t+1}]}. \tag{12}$$

These forecasts depend on the linear model parameters of both extreme regimes. Note that in the case of both the distribution of the shock and the threshold values being symmetric the long term forecast depends more on the contribution of the regime exhibiting higher *extreme* dependence. Also, for short forecast horizons not only $F_\varepsilon(\cdot)$ but also the value of the latest recorded shock have an effect on the forecasted value.

The main advantage of this model is its flexibility to describe the dynamics in the mean of the process. In contrast to standard TAR models the regimes depend on the lagged error variable and hence the model can accommodate asymmetries in the likelihood of positive and negative extremes and in the sequence of runs of extremes depending on the nature of the preceding shock. The case of asymmetric tails already mentioned in (1) is illustrated as follows:

$$P\{y_t \leq -v|\Im_{t-1}\} = F_\varepsilon \left(\frac{-v - (\alpha + \rho_t y_{t-1})}{h_t} \right), \tag{13}$$

and

$$P\{y_t > v|\Im_{t-1}\} = 1 - F_\varepsilon\left(\frac{v-(\alpha+\rho_t y_{t-1})}{h_t}\right), \tag{14}$$

with $v > 0$ a threshold value determining the tail regions. If $F_\varepsilon(\cdot)$ is symmetric about zero

$$P\{y_t > v|\Im_{t-1}\} = F_\varepsilon\left(\frac{-v+(\alpha+\rho_t y_{t-1})}{h_t}\right), \tag{15}$$

that is different from (13) unless the conditional mean process is zero.

The possibility of runs of extremes is also entertained, the corresponding probabilities are calculated in the following proposition.

Proposition 4.- Let $P_{t-2}\{A_t\} = P\{A_t|\Im_s\}$, with $\Im_s$ denoting information available up to time s. Then

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

where $x_{1t} = \min\left\{u_{1,} \frac{-v-(\alpha+\rho_{t-1}y_{t-2})}{h_{t-1}}\right\}, x_{2t} = \min\left\{u_{2,} \frac{-v-(\alpha+\rho_{t-1}y_{t-2})}{h_{t-1}}\right\}, x_{3t} = \max\left\{u_{2,} \frac{-v-(\alpha+\rho_{t-1}y_{t-2})}{h_{t-1}}\right\}$ *and* v *denotes a positive threshold. Equally,*

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

where $x_{1t} = \min\left\{u_{1,} \frac{v-(\alpha+\rho_{t-1}y_{t-2})}{h_{t-1}}\right\}, x_{2t} = \min\left\{u_{2,} \frac{v-(\alpha+\rho_{t-1}y_{t-2})}{h_{t-1}}\right\}, x_{3t} = \max\left\{u_{2,} \frac{v-(\alpha+\rho_{t-1}y_{t-2})}{h_{t-1}}\right\}.$

In a pure GARCH(1,1) model where $F_\varepsilon(\cdot)$ is symmetric about zero and $u_1 = -u_2$ 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. Thus, this property of these TAR models is not without importance for time series of financial returns where the clustering of extremes is usually attributed to large uncertainty regimes in financial markets and as such nothing is said about the possibility of predicting runs of large observations. We devote a further section below to study in more detail the implications of the TAR-GARCH model for these series.

3 Estimation and Testing of a TAR(3,1)-GARCH(1,1)

This section describes the estimation and testing of a TAR(3,1)-GARCH(1,1) process. The estimation is done by quasi-maximum likelihood (QML) where both mean and variance parameters are estimated jointly. Although the threshold variable is not observed it can be estimated from the data given its lagged character.

The main hypothesis, the existence of different correlation regimes between extremes, can be done by testing the statistical significance of the threshold effect. This test involves nuisance parameters which are not identified under the null hypothesis of no threshold effect. This is carried out by following the methodology proposed in Hansen (1996), adapted in our example to the case of an unobservable threshold variable. The following subsection describes the test.

3.1 Threshold Effect Test

The null hypothesis corresponds to the case $\rho_1 = \rho_2 = \rho_3$ in model (3) which implies that there is no different correlation regimes between the extremes determined by the sequence of standardized shocks. In this way, we entertain a process that under the null, $H_0 : \rho_1 = \rho_2 = \rho_3 = \rho$, is an AR(1)-GARCH(1,1) model,

$$y_t = \alpha + \rho y_{t-1} + h_t \varepsilon_t, \tag{16}$$

with h_t defined in (4). Under the alternative, y_t follows process (3). Clearly, under the null, u_1 and u_2 are not identified and thereby the testing problem involves nuisance parameters. As in the related literature, Davies (1977, 1987), Andrews and Ploberger (1994) and Hansen (1996), we propose a supremum and an average type test. Given that the threshold variable is not observable, the testing strategy proposed is a two-step procedure. In the first step we estimate the model under the null hypothesis. Following Ling and McAleer (2003), let $\widehat{\lambda} = \left(\widehat{\alpha}, \widehat{\rho}, \widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2,\right)$, then the QML estimator of $\lambda^0 = \left(\alpha^0, \rho^0, \beta_0^0, \beta_1^0, \beta_2^0,\right)$ is defined as:

$$\arg\max_{\lambda} L_n(\lambda) = \arg\max_{\lambda} \sum_{t=1}^{n} l_t(\lambda) \quad \text{with} \quad l_t(\lambda) = \log \left(\frac{1}{h_t(\lambda)\sqrt{2\pi}} e^{-\frac{\varepsilon_t^2(\lambda)}{2h_t^2(\lambda)}} \right)$$

with $\varepsilon_t^2(\lambda)$ and $h_t^2(\lambda)$ the error term and conditional variance process from an ARMA-GARCH(1,1) model. Theorem 5.1 of Ling and McAleer (2003) shows that

$$n^{1/2}(\widehat{\lambda}_i - \lambda_i^0) = O_p(1), \tag{17}$$

with n denoting the sample size. From this estimation procedure we obtain

$$\widehat{\varepsilon}_t = \frac{\widehat{a}_t}{\widehat{h}_t},$$

with $\widehat{a}_t(\widehat{\lambda}) = y_t - \widehat{\alpha} - \widehat{\rho} y_{t-1}$, and $\widehat{h}_t^2(\widehat{\lambda}) = \widehat{\beta}_0 + \widehat{\beta}_1 \widehat{a}_{t-1}^2 + \widehat{\beta}_2 \widehat{h}_{t-1}^2$. For ease of notation we will use throughout $\widehat{a}_t \equiv \widehat{a}_t(\widehat{\lambda})$ and $\widehat{h}_t^2 \equiv \widehat{h}_t^2(\widehat{\lambda})$.

In the second step, the model to be estimated is an alternative of (3) where the error term is replaced by the residual derived from step 1,

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 y_{t-1} I(\widehat{\varepsilon}_{t-1} < u_1) + \gamma_3 y_{t-1} I(\widehat{\varepsilon}_{t-1} > u_2) + a_t.$$

The model can be expressed in a more compact way as,

$$y_t = \gamma y_{t-1}(u) + a_t,$$

with $u = (u_1, u_2)$, $y_{t-1}(u) = (1, y_{t-1}, y_{t-1} I(\widehat{\varepsilon}_{t-1} < u_1), y_{t-1} I(\widehat{\varepsilon}_{t-1} > u_2))$ and $\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$. Thus, under the null hypothesis of no threshold effect, $\gamma_2 = \gamma_3 = 0$ for all possible values of the threshold vector $u \in U = \{(u_1, u_2) \in R^2 \quad \text{s.t.} \quad F_\varepsilon(u_1) \in (a_1, b_1) \wedge F_\varepsilon(u_2) \in (a_2, b_2) \quad \text{with} \quad 0 < a_1 < b_1 < a_2 < b_2 < 1\}$.

In order to test $H_0 : \gamma_2 = \gamma_3 = 0$ in the preceding model we propose to estimate $\widehat{\gamma}(u)$ by ordinary least squares and subsequently to use a conventional Wald type test. This is

$$T_n(u) = n\widehat{\gamma}(u)' R (R' \widehat{V}_n^*(u) R)^{-1} R' \widehat{\gamma}(u), \tag{18}$$

where $R = (0, I_2)'$, $\widehat{V}_n^*(u) = M_n(u,u)^{-1} \widehat{V}_n(u) M_n(u,u)^{-1}$, $\widehat{V}_n(u) = \frac{1}{n}\sum_{t=1}^n \widehat{s}_t(u)\widehat{s}_t(u)'$, $\widehat{s}_t(u) = y_{t-1}(u)\widehat{v}_t$, $M_n(u, u^*) = \frac{1}{n}\sum_{t=1}^n y_{t-1}(u) y_{t-1}(u^*)'$. Obviously, $T_n(u)$ depends on the value of the vector u, which is unknown. As it was mentioned above, following the related literature, two are the proposed statistics, the first one defined by the supremum on u of the set of Wald statistics, namely $sup_{u \in U} T_n(u)$; and the second test given by the average of the different test statistics on u, namely $Ave_{u \in U} T_n(u)$. The asymptotic distribution is provided in the next result, where $\Rightarrow$ denote weak convergence with respect to the Skorohod metric.

To state the asymptotic result (18) we need some additional notation. Let $S(u)$ be a mean zero Gaussian process with covariance kernel $K(u,u^*) = E\left(S(u)S(u^*)'\right)$, and let $\overline{S}(u) = R'M(u,u)^{-1}S(u)$ be another Gaussian process of zero mean where $M(u,u^*) = E\left(y_{t-1}(u)y_{t-1}(u^*)'\right)$, and covariance matrix $\overline{K}(u,u^*) = R'M(u,u^*)^{-1}K(u,u^*)M(u,u^*)^{-1}R$.

Theorem 2.- *Let $T_n(u)$ be the Wald test for the null $H_0 : \gamma_2 = \gamma_3 = 0$ for a given u. Consider that the following conditions hold:*

B.1 *ε_t are iid with zero mean and d.f. F, which admits a uniformly continuous density function f, $f > 0$.*

B.2 *Assumptions of Proposition 1 for $k = 6$.*

B.3 *Identifiability conditions in Ling and McAleer (2003).*

Then under the null,

$$T_n(u) \Rightarrow T^0(u) = \overline{S}(u)'\overline{K}(u,u)^{-1}\overline{S}(u).$$

$T^0(u)$ is the $chi-square\ process$ obtained by Hansen (1996), and the null distribution of $g^0 = g(T^0(u))$ depends, in general, on the covariance function $\overline{K}$. The critical values of this distribution cannot be tabulated except in special cases. To obtain the $p-values$ of this asymptotic test we propose two possible approximations to the asymptotic distribution: the first one is the $p-value$ approximation of Hansen (1996), and the second one the Wild bootstrap approximation. The validity of these $p-value$ approximations is shown from Theorem 2 and the Monte-Carlo simulation experiment in the next section.

A different alternative to test the correlation in the mean produced by the dependence in the extremes is to test the null hypothesis of martingale difference sequence for the residuals obtained from the model under the null. In particular, after the first step,

$$\widehat{\varepsilon}_t = \theta_0 + \theta_1 y_{t-1} I(\widehat{\varepsilon}_{t-1} < u_1) + \theta_2 y_{t-1} I(u_1 \leq \widehat{\varepsilon}_{t-1} \leq u_2) + \theta_3 y_{t-1} I(\widehat{\varepsilon}_{t-1} > u_2) + a_t,$$

we can regress

$$\begin{aligned} \widehat{\varepsilon}_t \widehat{h}_t &= y_t - \widehat{\alpha} - \widehat{\rho} y_{t-1} \\ \widehat{h}_t^2 &= \widehat{\beta}_0 + \widehat{\beta}_1 \widehat{\varepsilon}_{t-1}^2 \widehat{h}_{t-1}^2 + \widehat{\beta}_2 \widehat{h}_{t-1}^2. \end{aligned}$$

In this case, under the null, $H_0 : \theta_i = 0$ for $i = 1, 2, 3$ and for all $(u_1, u_2) \in U$. To test this hypothesis we can use a Wald or F-type test which both will depend on $u = (u_1, u_2)$ as in the previous case. Thereby we will need to use a supremum and/or an average type test as in the test discussed above.

Once the hypothesis of linearity of the data can be rejected we proceed to estimate jointly the parameters of the whole model by QML. In more detail, define $u = (u_1, u_2)$, $\phi = (\alpha, \rho, \beta)$, $\alpha = (\alpha_1, \alpha_2, \alpha_3)$,$\rho = (\rho_1, \rho_2, \rho_3)$ and $\beta = (\beta_0, \beta_1, \beta_2)$ and maximize the following function,

$$L(\phi, u) = \sum_{t=1}^{T} l_t(\phi, u),$$

with

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

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 $\widehat{u}$ we maximize $L(\widehat{\phi}(u), u)$ with respect to u. Thus, the $QMLE$ of (ϕ, u) is $(\widehat{\phi}(\widehat{u}), \widehat{u})$.

The study of statistical consistency and asymptotic distribution for the parameter estimators in this case is cumbersome for the asymptotic properties of the standard TAR and GARCH models do not generalize to this case that need of more convoluted techniques, see for example Gonzalo and Martinez (2007). Hence we leave these features of the model for future research.

3.2 Simulation Experiment

This subsection examines the performance of the preceding test for threshold effects through some Monte Carlo experiments for finite samples. In order to do this we will study 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 $\beta_0 = 0.05$, $\beta_2 = 0.1$ and $\beta_2 = 0.85$ and ε_t defined as in the previous case.
3. $y_t = \rho y_{t-1} + a_t$ with $\rho = 0.2$ 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 the experiments the threshold regime is defined by the following space:

$U = \{(u_1, u_2) \in R^2 \quad \text{s.t.} \quad F_\varepsilon(u_1) \in (0.1, 0.3) \wedge F_\varepsilon(u_2) \in (0.6, 0.9)\}$.

In practice F_ε is unknown and must be estimated by $\widehat{\varepsilon}_t$ from the model under the null. For the Hansen $p-value$ approximation we consider n=250, 500, 1000 and M=1000 Monte Carlo simulations, to show the validity of the asymptotic result for moderate sample sizes given the poor results obtained for smaller data sets. The following table 3.1 reports empirical estimates of the size at a 5% and 10% significance level for the statistic defined by the supremum of $T_n(u)$ over the set of possible threshold values.

$sup_{u \in U} T_n(u)$	n=250		n=500		n=1000	
size	0.05	0.1	0.05	0.1	0.05	0.1
IID	0.084	0.155	0.076	0.141	0.053	0.110
$GARCH(1,1)$	0.091	0.169	0.085	0.139	0.066	0.123
$AR(1) - GARCH(1,1)$	0.102	0.167	0.080	0.153	0.067	0.129

Table 3.1. *Empirical size at* 5% *and* 10% *of the* $sup_{u \in U} T_n(u)$ *test for* $n = 250$, $n = 500$ *and* $n = 1000$ *for different data generating processes derived from the Hansen p-value approximation.* $M = 1000$ *Monte-Carlo simulations and 300 internal simulation replications.*

For the statistic defined by the average of $T_n(u)$ the results of the simulated size are reported in table 3.2:

$Ave_{u \in U} T_n(u)$	n=250		n=500		n=1000	
size	0.05	0.1	0.05	0.1	0.05	0.1
IID	0.062	0.124	0.066	0.127	0.060	0.122
$GARCH(1,1)$	0.070	0.132	0.071	0.117	0.065	0.115
$AR(1) - GARCH(1,1)$	0.073	0.12	0.066	0.119	0.057	0.120

Table 3.2. *Empirical size at* 5% *and* 10% *of the* $Ave_{u \in U} T_n(u)$ *test for* $n = 250$, $n = 500$ *and* $n = 1000$ *for different data generating processes derived from the Hansen p-value approximation.* $M = 1000$ *Monte-Carlo simulations and 300 internal simulation replications.*

The Hansen $p - value$ approximation 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, thereby 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.

On the other hand this phenomenon is less important for the Wild Bootstrap approximation for which we report simulations for $n = 250, 500$ and M=500 in tables 3.3 and 3.4.

$sup_{u \in U} T_n(u)$	n=250		n=500	
size	0.05	0.1	0.05	0.1
IID	0.062	0.108	0.056	0.086
$GARCH(1,1)$	0.052	0.118	0.054	0.112
$AR(1) - GARCH(1,1)$	0.046	0.090	0.058	0.114

Table 3.3. *Empirical size at* 5% *and* 10% *of the* $sup_{u \in U} T_n(u)$ *test for* $n = 250$, $n = 500$ *for different data generating processes derived from the Wild bootstrap p-value approximation.* $M = 500$ *Monte-Carlo simulations and 300 internal simulation replications.*

$Ave_{u \in U} T_n(u)$	n=250		n=500	
size	0.05	0.1	0.05	0.1
IID	0.050	0.110	0.042	0.086
$GARCH(1,1)$	0.052	0.100	0.040	0.096
$AR(1) - GARCH(1,1)$	0.032	0.074	0.050	0.088

Table 3.4. *Empirical size at* 5% *and* 10% *of the* $Ave_{u \in U} T_n(u)$ *test for* $n = 250$, $n = 500$ *for different data generating processes derived from the Wild bootstrap p-value approximation.* $M = 500$ *Monte-Carlo simulations and 300 internal simulation replications.*

Finally we present the 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:

$$y_t = 0.2 y_{t-1} I(\varepsilon_{t-1} < -1.7) - 0.2 y_{t-1} I(\varepsilon_{t-1} > 1.7) + a_t$$

In the first case, $a_t = \varepsilon_t$, in the second one $a_t = \varepsilon_t h_t$ with $h_t^2 = 0.05 + 0.1 a_{t-1}^2 + 0.85 h_{t-1}^2$. In both cases, ε_t is *iid* $N(0,1)$. The results are in tables 3.5 and 3.6.

$sup_{u \in U} T_n(u)$	n=250		n=500	
size	0.05	0.1	0.05	0.1
$TAR - IID$	0.336	0.476	0.662	0.758
$TAR - GARCH(1,1)$	0.276	0.404	0.540	0.664

Table 3.5. *Empirical power at* 5% *and* 10% *of the* $sup_{u \in U} T_n(u)$ *test for* $n = 250$, $n = 500$ *for different data generating processes derived from the Wild bootstrap p-value approximation.* $M = 500$ *Monte-Carlo simulations.*

$Ave_{u\in U}T_n(u)$	n=250		n=500	
$size$	0.05	0.1	0.05	0.1
$TAR-IID$	0.456	0.584	0.768	0.858
$TAR-GARCH(1,1)$	0.350	0.464	0.638	0.750

Table 3.6. *Empirical power at* 5% *and* 10% *of the* $Ave_{u\in U}T_n(u)$ *test for* $n = 250$, $n = 500$ *for different data generating processes derived from the Wild bootstrap p-value approximation.* $M = 500$ *Monte-Carlo simulations.*

Note that in both examples and for both test statistics the power grows with the sample size. Also, the nonlinearity test is more powerful against homoscedastic alternative hypotheses. It is important to notice that the presence of conditional heteroscedasticity can lurk the existence of nonlinearity in the mean process. This is also studied in the next subsection.

3.3 The effect of misspecifying a TAR(3,1)-GARCH(1,1) model

This section studies the effects of misspecifying a TAR(3,1)-GARCH(1,1) model, that is, estimating a pure GARCH(1,1) or an AR(1)-GARCH(1,1) model where the true data generating process (DGP) is nonlinear. In order to measure the impact of this misspecification we carry out a Monte-Carlo simulation analysis in the following way. We generate $M = 500$ sequences of $n = 1000$ observations of a homoscedastic TAR and of a TAR-GARCH with threshold values given by $u_1 = -1.64$ and $u_2 = 1.64$. Also, in order to show the consistency of the estimators when the true DGP is estimated and to assess the accuracy of these estimates for $n = 1000$ we generate a linear AR-GARCH process. The error term for all of the three processes is standard Gaussian.

The following table (3.7) reports the sample mean and standard deviation of the estimates for the three models.

	Estimated Model						
	GARCH			*AR − GARCH*			
DGP	β_0	β_1	β_2	ρ	β_0	β_1	β_2
AR-GARCH	0.609	0.331	0.305	0.498	0.274	0.148	0.630
	(0.183)	(0.055)	(0.142)	(0.032)	(0.112)	(0.037)	(0.112)
TAR	0.876	0.246	0.027	0.264	0.578	0.113	0.380
	(0.115)	(0.042)	(0.104)	(0.032)	(0.375)	(0.065)	(0.384)
$TAR-GARCH_1$	0.645	0.314	0.152	0.264	0.491	0.208	0.358
	(0.164)	(0.045)	(0.145)	(0.036)	(0.191)	(0.047)	(0.195)
$TAR-GARCH_2$	0.642	0.314	0.154	0.024	0.622	0.272	0.209
	(0.149)	(0.046)	(0.134)	(0.045)	(0.176)	(0.043)	(0.161)

Table 3.7. *Estimates of: 1) an AR(1)-GARCH(1,1) process with parameters* $\alpha = 0$, $(\rho_1, \rho_2, \rho_3) = (0.50, 0.50, 0.50)$ *and* $(\beta_0, \beta_1, \beta_2) = (0.25, 0.10, 0.65)$*; 2) a TAR process with parameters* $\alpha = 0$, $(\rho_1, \rho_2, \rho_3) = (0.70, 0.00, 0.70)$ *and* $(\beta_0, \beta_1, \beta_2) = (1, 0, 0)$*; 3) a TAR process with parameters* $\alpha = 0$, $(\rho_1, \rho_2, \rho_3) = (0.70, 0.00, 0.70)$ *and* $(\beta_0, \beta_1, \beta_2) = (0.25, 0.10, 0.65)$*; 4) a TAR process with parameters* $\alpha = 0$, $(\rho_1, \rho_2, \rho_3) = (0.70, 0.00, -0.70)$ *and* $(\beta_0, \beta_1, \beta_2) = (0.25, 0.10, 0.65)$*. The error term is standard Gaussian. The number of Monte-Carlo simulations is* $M = 500$ *and the sample size is* $n = 1000$*. The standard errors of the different estimates are in brackets.*

The right section of the first row of the table shows the consistency of the estimators under correct specification of the model, and the left panel the known effect of misspecifying the mean when estimating a GARCH model. The other three rows are more interesting for our analysis. In all of the cases the estimates of the volatility parameters are very misleading. Note that for the estimate of the autoregressive parameter in the mean the AR-GARCH model provides very similar values for the TAR as well as for the TAR-GARCH process and that this value is neither zero or the expected value of the random autoregressive coefficient of the true model. This process depicts spurious serial correlation in the mean process. The misspecification in the case of the conditional variance is similar to the first row, that is, to the GARCH and AR-GARCH case.

4 The TAR(3,1)-GARCH(1,1) for financial returns

Under risk neutrality the efficiency market hypothesis (*EMH*) guarantees that the autocorrelation function of the sequence of returns is not significant for any lag, otherwise markets would be predictable, or is so small that makes investment opportunities derived from exploiting that dependence worthless. This is the rationale for using pure GARCH(1,1) processes or AR(1)-GARCH(1,1) with a very small although statistically significant autoregressive parameter for modeling the sequence of returns. By the same token SETAR models are not suitable to describe asset prices in an efficient market. Model (5) however can be devised to satisfy some forms of the market efficiency hypothesis but making allowance at the same time for describing periods of autocorrelation in the outer regimes. Suppose a process like (5) with unconditional zero mean and with error term following a symmetric distribution.

The first order autocovariance is

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

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

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

Note again that there is an extra term $\frac{Cov(\rho_t, y_{t-1}^2)}{Var(y_t)}$ in the latter expression compared to the case of a linear AR(1) process. If $E\left[y_t\right] \neq 0$ the extra term must be completed with functions of $E\left[y_t\right]$. Also, the TAR(3,1)-GARCH(1,1) model that we propose can reflect under some restrictions on the parameters a weak form of market efficiency (described by a zero correlation in the mean process) and also, in contrast to AR-GARCH processes, a strong linear dependence in the extremes if ρ_1 and ρ_3 take high values.

A vast majority of time series describing financial returns share some empirical features such as leptokurtic tails, volatility clustering, negative skewness and the leverage effect. This latter stylized fact given by the presence of more volatility after a large price fall than after a price rise of the same magnitude, is in particular, not well described in the standard GARCH framework. Nelson (1991) proposed the E-GARCH model to allow for asymmetric effects of previous positive and negative observations in the volatility process. In this line Glosten, Jagannathan and Runkle (1993) and Zakoian (1994) introduced the T-GARCH model, a variation of a GARCH(1,1) model where the sign of the previous observation produces an asymmetric effect in the volatility process.

The TAR(3,1)-GARCH(1,1) albeit different in spirit shares characteristics of these models. It is able to reflect the increase in the likelihood of extreme observations for the sequence y_t if the previous observation is negative as the E-GARCH and T-GARCH do. In contrast to these models however, for the TAR-GARCH this probability does not increase only because of a shift in the conditional volatility process but because of a nonlinear change in the conditional mean produced by the magnitude of the previous observation and shock. In this way this model is able to refine the insights of the leverage effect phenomenon by permitting to test whether negative returns are followed by an increase in the probability of future negative returns or simply by an increase in future volatility. Proposition 4 for model (5) with $v = 0$ yields the following probability conditional on information up to time $t-2$.

$$\begin{aligned} P_{t-2}\left(y_t \leq 0, y_{t-1} \leq 0\right) = & \int_{-\infty}^{x_{1t}} F_{\varepsilon}\left(\frac{-\left(\alpha+\rho_1\left(\alpha+\rho_{t-1} y_{t-2}+\varepsilon h_{t-1}\right)\right)}{h_t(\varepsilon)}\right) f_{\varepsilon}(\varepsilon)\, \partial \varepsilon \\ & +\int_{x_{1t}}^{x_{2t}} F_{\varepsilon}\left(\frac{-\alpha}{h_t(\varepsilon)}\right) f_{\varepsilon}(\varepsilon)\, \partial \varepsilon \\ & +\int_{u_2}^{x_{3t}} F_{\varepsilon}\left(\frac{-\left(\alpha+\rho_3\left(\alpha+\rho_{t-1} y_{t-2}+\varepsilon h_{t-1}\right)\right)}{h_t(\varepsilon)}\right) f_{\varepsilon}(\varepsilon)\, \partial \varepsilon, \end{aligned}$$

where $x_{1t} = \min\left\{u_1, \frac{-(\alpha+\rho_{t-1}y_{t-2})}{h_{t-1}}\right\}$, $x_{2t} = \min\left\{u_2, \frac{-(\alpha+\rho_{t-1}y_{t-2})}{h_{t-1}}\right\}$, $x_{3t} = \max\left\{u_2, \frac{-(\alpha+\rho_{t-1}y_{t-2})}{h_{t-1}}\right\}$.

For values of the parameters resulting in a probability different from 0.5 our TAR process describes asymmetry in the effect of previous losses and gains, or in other words, $P\{y_t \leq 0|y_{t-1} \leq 0\} \neq P\{y_t > 0|y_{t-1} \leq 0\}$. This is in clear contrast to pure GARCH(1,1), E-GARCH(1,1) and T-GARCH processes.

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. Another example of crisis in financial markets was the black Monday in October 1987, that day prices dropped by an average of 20% in US and UK stock markets and even more in other economies worldwide, and triggered a period of widespread decline in financial markets. While the first example is clearly attributed to a very negative and unexpected sequence of extreme events during that week of September of 2001 that had such disastrous effects, the nature of the second event cannot be clearly attributed to a certain cause. In the latter case markets started to respond to drops in prices with a cascade of sale orders that were followed by action of speculators that swamped the markets with buy and sell orders producing high uncertainty in financial markets during long after that Monday of October 1987.

The following plots represent the sequences of prices and log-returns $r_t = 100\,(lnP_t -$

lnP_{t-1}) of General Motors (GM) stocks for each crisis episode. The data are collected from Yahoo-Finance.

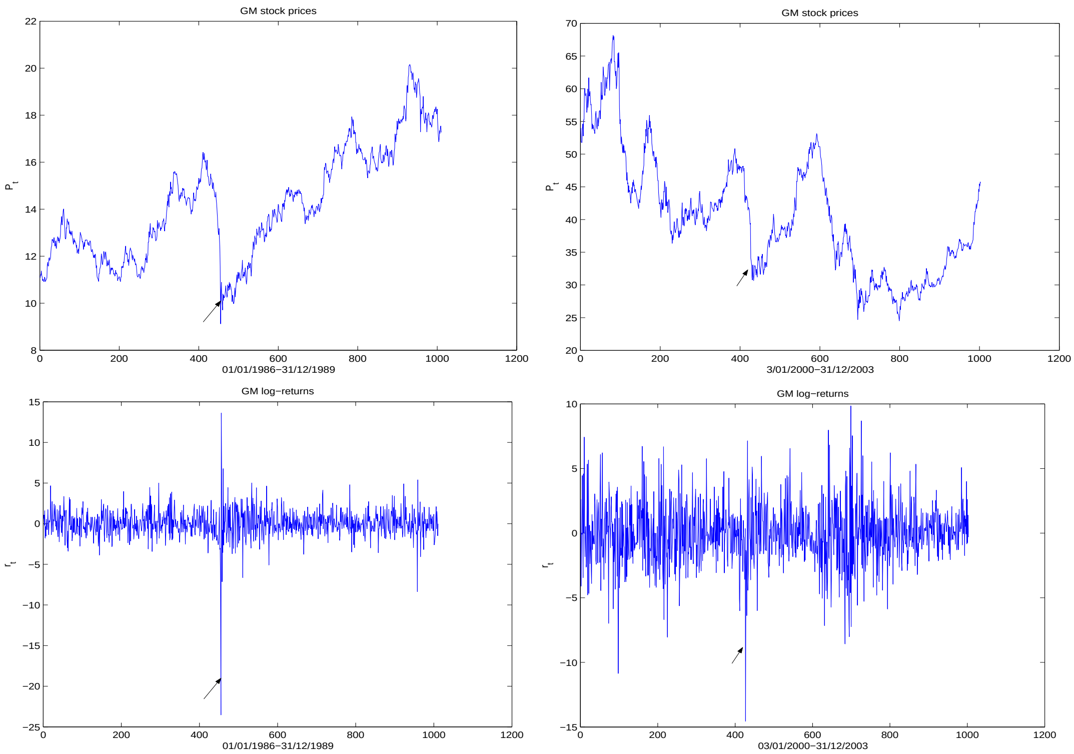

Figure 5.1. *The upper-left panel depicts the prices of GM for the period* 01/01/1986 − 31/12/1989, *and the upper-right panel those for the period* 03/01/2000 − 31/12/2003. *The lower panels plot the corresponding sequences of log-returns. The Black Monday of October 1987 and September 11th, 2001 are signaled with an arrow.*

It is interesting to observe the very different patterns in prices and returns described around each crisis. The volatilities also seem to differ very much across plots.

In this section the TAR-GARCH methodology is applied to determine statistically if the returns on the days following these events 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.

Table 5.1 reports the estimates of a GARCH(1,1), an AR(1)-GARCH(1,1), and a TAR(3,1)-GARCH(1,1) model for daily log-returns from GM stock covering the period January 1986-December 1989. The analysis comprises 1011 observations.

$Model$	$GARCH(1,1)$	$AR(1) - GARCH(1,1)$	$TAR(3,1) - GARCH(1,1)$
α	0.052 (0.046)	0.052 (0.047)	0.027
ρ_1	-	-0.017 (0.041)	-0.087
ρ_2	-	-0.017 (0.041)	0.164
ρ_3	-	-0.017 (0.041)	1.225
β_0	0.867 (0.171)	0.874 (0.173)	0.813
β_1	0.231 (0.016)	0.232 (0.016)	0.289
β_2	0.438 (0.081)	0.435 (0.082)	0.431
Log lkl	-1867.6	-1868.1	-1874.9

Table 5.1. $u_1 = -0.578$, $u_2 = 1.218$. *Estimates for October 1987 subsample (01/01/1986-31/12/1989), n=1011. p-value of Hansen test ($sup_{u \in U} T_n(u)$=0.979, p-value of $Ave_{u \in U} T_n(u)$=0.987). The standard errors of the different estimates are in brackets.*

The results from this table are somehow expected. The dependence in the extremes is found not to be statistically significant. Although the parameter for the upper regime in the TAR(3,1)-GARCH(1,1) is high the two versions of the Hansen test do not reject the null hypothesis. In line with this the likelihood functions yield very similar results indicating scarce significance also of the conditional mean parameter in the AR(1) component. These results can be interpreted as those expected in an efficient stock market. By doing so, we accept that the crisis of October 1987 was sparked by an increase in volatility and that posterior strong fluctuations in *GM* stock price were produced by the volatility process and not by dependence in the extremes.

The analysis of the *GM* stock prices corresponding to the period January 2000-December 2003 is diametrically different. The following table reports the estimates of the different models. 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
ρ_1	-	-0.052 (0.031)	0.072
ρ_2	-	-0.052 (0.031)	-0.031
ρ_3	-	-0.052 (0.031)	1.225
β_0	0.117 (0.033)	0.117 (0.033)	0.087
β_1	0.079 (0.012)	0.076 (0.012)	0.080
β_2	0.902 (0.014)	0.904 (0.014)	0.908
Log lkl	-2252.3	-2251.1	-2659.8

Table 5.2. $u_1 = -0.919$, $u_2 = 0.785$. *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.*

Stock returns on GM for this period exhibit very different characteristics from the previous period studied. 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. 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.

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.

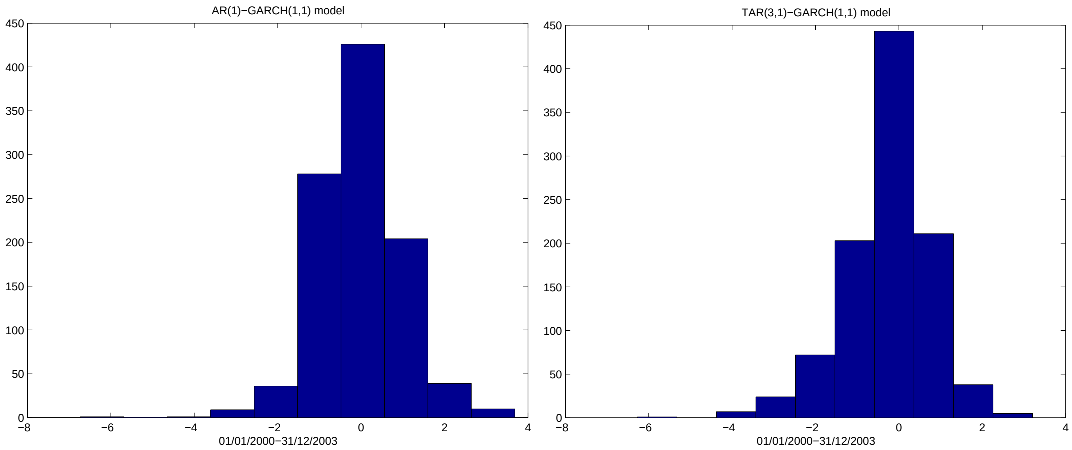

Figure 5.2. *Histograms for the residuals sequence from AR(1)-GARCH(1,1) (left panel) and from TAR(3,1)-GARCH(1,1) (right panel) for the period* $03/01/2000 - 31/12/2003$.

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 structures of dependence; asymmetries in the probabilities in the tails, in the sequences of runs of extremes, etc. This is particularly interesting for modelling 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 offering at the same time the possibility of describing linear dependence in the extremes. This fact led us in the empirical application to study the consequences of sound worldwide financial crises: Black Monday of October 1987 and September 11^{th} terrorist attacks. Using our TAR-GARCH method we find evidence of predictability of extremes after September 11^{th} but not after the Black Monday event of October 1987. This indicates that while the former event produced a sequence of predictable large negative returns, the latter event only produced a shift in the volatility process and no evidence of dependence between extremes. This is in line with literature in the topic attributing the worldwide plummeting of stocks in October 1987 to a market correction and the burst of a global stock market bubble.

Extensions of this methodology to describe the conditional volatility process are ongoing research. Other extensions include empirical applications to data in international finance and energy.

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 (5). This is shown in Brandt (1986). In order to prove the strict stationarity and ergodicity of a_t, Kristensen (2006) use 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,

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

Define, $\|x\|_k = \left(E\left[x^k\right]\right)^{1/k}$, and denote $\|\rho_t\|_k = \lambda$. Then, from the Minkowsky inequality, independence of ε_t and strict stationarity of a_t:

$$\begin{aligned}\|y_t\|_k &\leq \sum_{j=0}^{\infty} \| \prod_{i=0}^{j} \rho_{t-i} a_{t-j} \|_k = \sum_{j=0}^{\infty} \| \prod_{i=0}^{j-2} \rho_{t-i} \|_k \|\rho_{t-j+1}\rho_{t-j} a_{t-j}\|_k \\ &\leq \sum_{j=0}^{\infty} \lambda^{j-2} \|\rho_{t-j+1}\rho_{t-j} a_{t-j}\|_k \leq \frac{\max_i \rho_i^2 \|a_t\|_k}{\lambda^2(1-\lambda)}\end{aligned}$$

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

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

with $\lambda < 1$ and $\|\varepsilon_t^2\|_{k/2} < \infty$ by assumption A.8, which proves that $\|a_t\|_k < \infty$.

Proof of Proposition 2: Now we present 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 $k = 2$.

$$E[y_t] = \alpha + E[\rho_t y_{t-1}]. \tag{19}$$

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

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

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

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

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

$$V\left[y_t\right] = E\left[y_t^2\right].$$

Also,

$$V\left[y_t\right] = V\left[\rho_t y_{t-1}\right] + V\left[a_t\right].$$

Note that

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

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

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

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

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

Proof of Proposition 3: Under assumptions in Proposition 1 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 diference sequence the preceding expression reads as

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

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.

$$E[y_{t+l}|\Im_t] \xrightarrow{L_2} \frac{\alpha}{1-E[\rho_{t+1}]} + \frac{E[\rho_{t+1}a_t]}{1-E[\rho_{t+1}]}. \tag{20}$$

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

$$\sum_{i=1}^{l-1} \|E[h_{t+i}|\Im_t] - E\left[h_{t+i}\right]\|_2 \left|E[\rho_{t+1}]^{l-i-1}\right| \to 0 \tag{21}$$

when $l \to \infty$, given that $E[\rho_{t+1}a_t] = E[\rho_{t+1}\varepsilon_t]E\left[h_t\right]$.

A sufficient condition to prove (21) is to see that $x_t = h_t - E\left(h_t\right)$ is a $L_2 - mixingale$, that is, $\|E\left(x_t|\Im_{t-m}\right)\|_2 \leq \|c_t\|_2 \gamma\left(m\right)$, with $\gamma\left(m\right) \to 0$, see Davidson (1994) or McLeish (1975). This expression allows us to construct an upper bound for (21) that satisfies

$$\sum_{i=1}^{l-1} \gamma\left(i\right) \|c_t\|_2 \left|E[\rho_{t+1}]^{l-i-1}\right| \to 0.$$

Now, using A.8 for $k = 4$ it can be proved that x_t is a L_2-Near Epoch Dependence on ε_t of size $-\infty$ (L_2-NED, see Davidson (1994) or McLeish (1975)). Finally, from Theorem 17.5 of Davidson (1994), we prove that x_t is a L_2-mixingale with $\|c_t\|_2 < \infty$ and $\gamma\left(m\right) = \lambda^m$ for $0 < \lambda < 1$, and the proof of the mixingale property immediately follows.

Proof of Proposition 4: By Bayes' theorem

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

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

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

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

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

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

Proof of Theorem 2: Before proceeding with the proof of Theorem 2 we need the following definition:

$$\Upsilon_n(u, \lambda) = n\widehat{\gamma}(u, \lambda)' R \left(R' \widehat{V}_n^*(u, \lambda) R \right)^{-1} R' \widehat{\gamma}(u, \lambda),$$

with $\lambda = (\alpha, \rho, \beta_0, \beta_1, \beta_2)$, $\widehat{\gamma}(u, \lambda)$ and $\widehat{V}_n^*(u, \lambda)$ the same variables as defined in Section 3 but using $\varepsilon_t(\lambda)$ instead of $\widehat{\varepsilon}_t$.

It is easy to prove for the statistic $\Upsilon_n(u, \lambda^0)$ obtained from replacing the residual variable $\widehat{\varepsilon}_t$ by the error term ε_t in $T_n(u)$ that the assumptions 1 to 3 in Hansen (1996) hold. Thus, $\Upsilon_n(u, \lambda^0) \Rightarrow T^0(u) = \overline{S}(u)' \overline{K}(u, u)^{-1} \overline{S}(u)$. Then, the proof of Theorem 2 follows from this result:

$$\sup_{u \in U} \Lambda_n(u, \widehat{\lambda}) = \sup_{u \in U} \left| \Upsilon_n(u, \lambda^0) - \Upsilon_n(u, \widehat{\lambda}) \right| = \sup_{u \in U} \left| \Upsilon_n(u, \lambda^0) - T_n(u) \right| = o_p(1), \qquad (22)$$

If this condition holds the process $T_n(u)$ and $\Upsilon_n(u, \lambda^0)$ have the same asymptotic distribution, provided that $\Upsilon_n(u, \lambda^0)$ converges to $T^0(u)$ as shown before.

To prove (22) we use the result of Lemma 1. To state this result, let $m \geq 1$ be a fixed integer, τ_{n1}, ξ_{n1} be measurable functions from $\Re^m$ to $\Re$, such that $(\eta_{nt}, \gamma_{nt}, \tau_{nt}(\lambda), \xi_{nt}(\lambda)), 1 \leq t \leq n$, are an array of 4-tuple random variables defined on a probability space where $\{\eta_{nt}, 1 \leq t \leq n\}$ are i.i.d. according to a d.f. F, and for every $\lambda \in \Re^m$, η_{nt} is independent of $(\gamma_{nt}, \tau_{nt}(\lambda), \xi_{nt}(\lambda))$, $1 \leq t \leq n$. Furthermore, let $\{A_{nt}\}$ be and array of sub σ-fields such that $A_{nt} \subset A_{nt+1}, 1 \leq t \leq n, n \geq 1$; $(\gamma_{n1}, \tau_{n1}(\lambda), \xi_{n1}(\lambda))$ is A_{n1} measurable, the r.v.'s $\{\eta_{n1}, \ldots, \eta_{nj-1}; (\gamma_{nt}, \tau_{nt}(\lambda), \xi_{nt}(\lambda)), 1 \leq t \leq j\}$ are A_{nj} measurable, $2 \leq j \leq n$; and η_{nj} is independent of $A_{nj}, 1 \leq j \leq n$. Finally, $\gamma_{nt} \geq 0$. Define por $u \in \overline{\Re} = [0, \overline{u}]$, with $\overline{u} < \infty$,

$$\mathfrak{V}_n(u,\lambda) = n^{-1/2}\sum_{t=1}^{n-1}\eta^{+}_{nt+1}g\left(\eta^{+}_{nt}\right)\gamma_{nt}I\left(\eta_{nt}\le u+u\tau_{nt}(\lambda)+\xi_{nt}(\lambda)\right),$$

$$\mathfrak{J}(u,\lambda) = n^{-1/2}\sum_{t=1}^{n-1}\eta^{+}_{nt+1}\gamma_{nt}H\left(u+u\tau_{nt}(\lambda)+\xi_{nt}(\lambda)\right),$$

$$\mathfrak{U}_n(u,\lambda) = \mathfrak{V}_n(u,\lambda)-\mathfrak{J}(u,\lambda),$$

$$\mathfrak{U}^{*}_n(u,\lambda) = n^{-1/2}\sum_{t=1}^{n-1}\eta^{+}_{nt+1}\gamma_{nt}\left[g\left(\eta^{+}_{nt}\right)I\left(\eta_{nt}\le u\right)-H(u)\right].$$

We also need the following assumptions:

(C.0) F has a.e. positive density f with $\|f\|_\infty = \sup_{u\in\Re} f(u) < \infty$.

(C.1) $\|g(e)f(e)\|_\infty = \sup_{e\in\Re} g(e)f(e) < \infty$.

(C.2) $n^{-1}\sum_{t=1}^{n-1}\gamma^2_{nt} = O_p(1), \quad \max_{1\le t\le n} n^{-1/2}\left|\eta_{nt+1}\gamma_{nt}\right| = o_p(1)$.

For each $\lambda\in\Re^m$,

(C.3) $\max_{1\le t\le n}\left\{\left|\tau_{nt}(\lambda)\right|+\left|\xi_{nt}(\lambda)\right|\right\} = o_p(1)$,

(C.4) $n^{-1}\sum_{t=1}^{n-1}\left|\eta_{nt+1}\gamma_{nt}\right|\left[\left|\tau_{nt}(\lambda)\right|+\left|\xi_{nt}(\lambda)\right|\right] = O_p(1)$.

$\forall\epsilon>0, \exists\delta>0$, and an $n_1<\infty, \forall 0<b<\infty, \forall\|s\|\le b, \forall n>n_1$

(C.5) $$P\left(n^{-1/2}\sum_{t=1}^{n-1}\left|\eta_{nt+1}\gamma_{nt}\right|\left\{\sup_{\|\lambda-s\|<\delta}\left|\tau_{nt}(\lambda)-\tau_{nt}(s)\right|+\sup_{\|\lambda-s\|<\delta}\left|\xi_{nt}(\lambda)-\xi_{nt}(s)\right|\right\}\le\epsilon\right)>1-\epsilon.$$

The following lemma gives the needed results.

Lemma 1 *Under the above setup and under the assumptions C.1-C.5, for every* $0<b<\infty$,

$$\sup_{u\in\overline{\Re},\|\lambda\|\le b}\left|\mathfrak{U}_n(u,\lambda)-\mathfrak{U}^{*}_n(u,\lambda)\right| = o_p(1),$$

$$\sup_{u\in\overline{\Re},\|\lambda\|\le b}\left|n^{-1/2}\sum_{t=1}^{n-1}\eta_{nt+1}\gamma_{nt}\left[H\left(u+u\tau_{nt}(\lambda)+\xi_{nt}(\lambda)\right)-H(u)\right]\right| = o_p(1).$$

The proof of this lemma is similar to that of Lemma 4.1 in Koul and Ling (2006), but using now $h(x)=g\left(\eta^{+}_{nt}\right)I(\eta_{nt})$ instead of $h(x)=I(\eta_{nt})$. Hence the proof is omitted although a complete version is found in the working paper version of this article.

To see the importance of this lemma in the proof of Theorem 2 note that

$$\sup_{u\in U}\left|\Upsilon_n(u,\lambda^0)-\Upsilon_n(u,\widehat{\lambda})\right| \leq \sup_{u\in U,|\lambda^0-\lambda|=O_p(n^{-1/2})}\left|\Upsilon_n(u,\lambda^0)-\Upsilon_n(u,\lambda)\right|.$$

The right hand side of the preceding equation can be expressed as the following type of processes,

$$\sup_{u\in U}\frac{1}{n^{1/2}}\sum_{t=2}^{n}\varepsilon_t\varepsilon_{t-1}h_t\gamma_{t-2}\left[I\left(a_{t-1}(\lambda)\leq uh_{t-1}(\lambda)\right)-I\left(a_{t-1}\leq uh_{t-1}\right)\right].$$

For our case, AR-GARCH(1,1) processes, Koul and Ling (2006) prove that the preceding process can be written as

$$\sup_{u\in U}\frac{1}{n^{1/2}}\sum_{t=2}^{n}\varepsilon_t\varepsilon_{t-1}h_t\gamma_{t-2}\left[I\left(\varepsilon_t\leq u+u\tau_{nt}\left(\lambda\right)+\xi_{nt}\left(\lambda\right)\right)-I\left(\varepsilon_{t-1}\leq u\right)\right],$$

where conditions (C.2)-(C.5) are satisfied in our case by assumptions B.1-B.3, and conditions (C.0) and (C.1) are assumed on B.1. Therefore this process is $o_P(1)$ and Lemma 1 is proven. Hence the proof of Theorem 2 also follows.

References

[1] Andrews, D W K & Ploberger, W., (1994). Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative. *Econometrica, Econometric Society,* vol. 62(6), 1383-1414.

[2] Bai, J., (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *The Annals of Statistics* 22, 2051-2061.

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

[4] Brandt, A., (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Probab.,* 18, 211220.

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

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

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

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

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

[10] 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.

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

[12] Gonzalo, J., Martinez, O., (2006). Large shocks vs. small shocks. Or does size matter? May be so. *Journal of Econometrics* 127(1-2), 311-347.

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

[14] Gonzalo, J., Pitarakis, J., (1998). On the Exact Moments of Asymptotic Distributions in an Unstable Ar(1) with Dependent Errors. *International Economic Review Vol.* 39, 1, 71-88.

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

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

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

[18] Hansen, Bruce E., (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics,* vol. 93(2), 45-368.

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

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

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

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

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

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

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

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

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

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

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

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

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