Statistica Sinica **24** (2014), 625-652
doi:http://dx.doi.org/10.5705/ss.2012.185

ON CONDITIONALLY HETEROSCEDASTIC AR MODELS WITH THRESHOLDS

Kung-Sik Chan, Dong Li, Shiqing Ling and Howell Tong

University of Iowa, Tsinghua University,
Hong Kong University of Science & Technology
and London School of Economics & Political Science

Abstract: Conditional heteroscedasticity is prevalent in many time series. By viewing conditional heteroscedasticity as the consequence of a dynamic mixture of independent random variables, we develop a simple yet versatile observable mixing function, leading to the conditionally heteroscedastic AR model with thresholds, or a T-CHARM for short. We demonstrate its many attributes and provide comprehensive theoretical underpinnings with efficient computational procedures and algorithms. We compare, via simulation, the performance of T-CHARM with the GARCH model. We report some experiences using data from economics, biology, and geoscience.

Key words and phrases: Compound Poisson process, conditional variance, heavy tail, heteroscedasticity, limiting distribution, quasi-maximum likelihood estimation, random field, score test, T-CHARM, threshold model, volatility.

1. Introduction

We can often model a time series as the sum of a conditional mean function, the drift or trend, and a conditional variance function, the diffusion. See, e.g., Tong (1990). The drift attracted attention from very early days, although the importance of the diffusion did not go entirely unnoticed, with an example in ecological populations as early as Moran (1953); a systematic modelling of the diffusion did not seem to attract serious attention before the 1980s.

For discrete-time cases, our focus, as far as we are aware it is in the econometric and finance literature that the modelling of the conditional variance has been treated seriously, although Tong and Lim (1980) did include conditional heteroscedasticity. Specifically, Engle (1982) proposed the autoregressive conditional heteroscedasticity (ARCH) model, which has attracted much attention. (For a simple introduction see, e.g., Cryer and Chan (2008).)

If we view conditional heteroscedasticity as a consequence of a dynamic mixture of independent random variables, then the ARCH model is just one

form of mixture among many. For the ARCH model, in order to preserve non-negativity of the conditional variance as well as stationarity, constraints on its parameters are necessary, which can sometimes be rather restrictive. To cope with different applications, it has been found necessary to extend the basic ARCH mixing function to ones of increasing complexity, associated with the acronyms GARCH, EGARCH, IGARCH, NGARCH, GARCH-M, QGARCH, GJR-GARCH, TGARCH, fGARCH, and others. These extensions carry with them additional parameter constraints. We would argue that there is no *a priori* reason why the mixing function should be wedded to the form initiated in Engle (1982).

Unobservable/hidden mixing functions have been studied. For example, Schwert (1989) used the two-state regime-switching hidden Markov model to study monthly stock returns. Cai (1994) proposed the switching-AR-Markov-ARCH model and used it to study three-month T-Bill excess return, and Gray (1996) tried to model the conditional distribution of interest rates by a regime-switching model. Some general regime-switching Markov GARCH models were proposed by Hamilton and Susmel (1994) and Haas, Mittnik, and Paolella (2004) However, Schwert (1989) has argued that the extension of the two-state regime-switching hidden Markov model involves more complicated computational procedures and suspected that the benefits from such extensions would not exceed the costs. However, as far as we know, no asymptotic theory is available for the statistical inference of regime-switching Markov GARCH models, perhaps due to its complex nature.

The simple idea of a piecewise constant mixing function was already buried in Moran (1953) and made its debut in Tong and Lim (1980). Later, Tong (1982) discussed how a discontinuous decision process could result in a smooth dynamical system becoming a piecewise smooth one. We pursue the idea systematically in this paper.

Clearly, thresholds are the key to the piecewise-constant approach to modelling the conditional variance function, just as they are to the piecewise-linear approach to modelling the conditional mean function. Recent references of the latter, since its introduction by Tong (1978), are rather extensive. As representatives, we refer to Chan and Li (2007), Chan (2009), Hansen (2011), and Tong (2011). Still, many challenges remain.

Focusing on the conditional variance, sometimes called volatility, this paper studies the threshold model for conditional heteroscedasticity, or T-CHARM for short. A T-CHARM has appealing features: it is always strictly stationary and ergodic *essentially without any restriction on the parameters*; its mixing function is flexible enough that asymmetric heteroscedasticity is not a problem; strong consistency is enjoyed by the quasi-maximum likelihood estimators of the

parameters, and convergence rates are available. In fact, the limiting distribution of the estimated threshold is the same as that of the smallest minimizer of a two-sided compound Poisson process.

Despite an apparent resemblance of T-CHARM to the QGARCH model (Gourieroux and Monfort (1992)) and the Threshold ARCH model (Rabemananjara and Zakoïan (1993); Zakoïan (1994)), in that they all have a non-parametric flavor, the latter two assume *fixed and known* thresholds over which the conditional variance function jumps, while the former infers the *unknown* number and *unknown* locations of the thresholds from the data.

This paper is organized as follows. Section 2 presents the model and its structure. Section 3 studies the estimation of the model parameters. Section 4 gives the limiting distribution of the estimated threshold including a useful approximation. Section 5 considers various statistical tools pertinent to T-CHARM. Section 6 compares the T-CHARM and the GARCH model via simulation. Section 7 illustrates the methodology with sets of data, from finance, the biological and environmental sciences. We conclude the paper with some discussion in Section 8. Proofs of key results are given as *Supplementary Materials*.

2. T-CHARM and Its Probabilistic Structure

A T-CHARM is a threshold autoregressive (TAR) model. In its simplest form

$$X_t = \sigma(X_{t-1})\eta_t, \tag{2.1}$$

where $\{\eta_t\}$ are independent and identically (but not necessarily normally) distributed (i.i.d.) random variables each with zero mean and unit variance, η_t is independent of $\{X_s : s < t\}$, and $\sigma(x)$ is a piecewise constant function of x. Assume that σ takes m distinct values. Specifically, let $\sigma(x) = \sigma_i$ for $x \in R_i$, where σ_i's are distinct positive numbers, and $\{R_i, i = 1, \ldots, m\}$ defines a partition of the real line R, $R = \cup_{i=1}^{m} R_i$ and the R_i's are pairwise disjoint. The R_i's are referred to as regimes. Thus, the conditional variance of X_{t+1} given current and past X's depends only on the regime into which X_t falls. It is clear that $\mathbb{E}(X_t|X_s, s < t) \equiv 0$, so $\{X_t\}$ is a martingale difference sequence of random variables. Consequently, the X's are uncorrelated forming a sequence of white noise. A T-CHARM can be generalized in many ways, just as a TAR model can; see Tong (1978). Thus the argument of the function σ can be replaced by a more general 'state' variable, that can be a function of either an observable or a hidden time series, or both. One could also smooth the piecewise constant function σ, in several ways, e.g. similar to those described in Tong (1978).

We show that the dependence structure of the X's is generally revealed by the autocorrelation structure of some instantaneously nonlinear transformation

of the X's. Thus, for any instantaneous transformation $Y_t = h(X_t)$ with finite second moments and $\mathbb{E}(h(\eta_t)) \neq 0$, the autocorrelation function (ACF) of the transformed process is generally the same as that of some stationary ARMA$(m-1, m-1)$ process. Here the AR and MA orders are generally one less than the number of regimes.

The regime process $\{S_t\}$ is to have $S_t = i$ if and only if $X_t \in R_i, i = 1, \ldots, m$. It is a Markov chain. Let $P = (p_{ij})$ be its 1-step transition probability matrix $p_{ij} = \mathbb{P}(X_{t+1} \in R_j | X_t \in R_i)$. It is well-known that if 1 denotes an m-dimensional vector of unit elements, it is a right eigenvector of P. Consequently, there exists an m-dimensional non-zero left eigenvector u,

$$u^\tau P = u^\tau, \tag{2.2}$$

where τ denotes the transpose of a vector or matrix. If u is non-negative, it can be normalized to sum to 1. Then if the regime process has u as its initial probability distribution, the regime process is stationary. The irreducibility of P is a necessary and sufficient condition for the existence of a unique stationary distribution for the Markov chain, and it is irreducible if and only if $\sum_{t=1}^{\infty} P^t$ is a positive matrix. P is irreducible under very mild conditions, for example, if η_t has a positive probability density function. Henceforth, we assume that P is an irreducible matrix.

We show that the stationary distribution of X_t is a mixture of distributions of the $\sigma_j \eta$, where η has the same distribution as the common distribution of the innovations $\{\eta_t\}$, with u_j as the probability weights, where $u^\tau = (u_1, u_2, \ldots, u_m)$. Conditional on $X_0 = x_0 \in R_i$, $S_{t-1} = j$ with probability P_{ij}^{t-1}, in which case X_t is distributed as $\sigma_j \eta_t$. Because $P_{ij}^{t-1} \to u_j$, the conditional distribution of X_t converges in distribution to the mixture of distributions of $\sigma_j \eta$ with probability u_j. It is readily seen that this limiting mixture distribution is the stationary distribution of $\{X_t\}$. Then the ℓ-step ahead predictive distribution of $X_{t+\ell}$ given X_t in the ith regime is a mixture of distributions of $\sigma_j \eta$ with probability weight P_{ij}^{ℓ}. In practice, if the X's are returns, it is of interest to predict the uncertainty in the ℓ-step-ahead cumulative returns to explore the predictive distribution of $\sum_{k=1}^{\ell} X_{t+k}$. It can be seen that the latter distribution is identical to that of a mixture of distributions corresponding to $\sum_{k=1}^{\ell} \sigma_{j_k} \eta_{t+k}$, with probability weights $P_{i,j_1} \prod_{k=2}^{\ell} P_{j_{k-1},j_k}$, for $j_k \in \{1, \ldots, m\}$, $k = 1, \ldots, \ell$.

It follows from the Cayley-Hamilton Theorem that if $d(x) = \det(xI - P)$ is the characteristic polynomial of P, then $d(P) = 0$. Simply let $d(x) = x^m - \sum_{j=1}^{m} d_j x^{m-j}$, $d(P) = P^m - \sum_{j=1}^{m} d_j P^{m-j}$, where the superscript denotes a matrix power and the zeroth power is I, the identity matrix. Since 1 is the unique eigenvalue of P that is of unit magnitude, $d(x) = (x-1)c(x)$, where $c(x) = x^{m-1} - \sum_{j=1}^{m-1} c_j x^{m-j-1}$ has all its root of magnitude strictly less than 1.

Theorem 1. *Let $\{X_t\}$ be as in Eqn.* (2.1) *and the transition probability matrix P of the associated regime process $\{S_t\}$ be an irreducible $m \times m$ matrix. Let $Y_t = h(X_t)$, where h is a continuous function. Assume that $\{Y_t\}$ admits finite second moments and $\mathbb{E}(h(\eta_t)) \neq 0$. Let $\gamma_k = \gamma_{k,Y}$ be the kth lag auto-covariance of $\{Y_t\}$. Then $\{\gamma_k\}$ satisfies the Yule-Walker equation*

$$\gamma_k = c_1\gamma_{k-1} + \ldots + c_{m-1}\gamma_{k-m+1} \quad \textit{for } k \geq m. \tag{2.3}$$

The fact that $\{\gamma_k\}$ satisfies (2.3) means that the ACF of $\{Y_t\}$ is exactly the same as that of some ARMA$(m-1, m-1)$ process. For, let B be the backshift operator defined by $BY_t = Y_{t-1}$ and observe that because of the Yule-Walker equation, $W_t = c(B)Y_t$ is a process of memory not more than $m-1$ lags, and so $\{W_t\}$ must be an MA$(m-1)$ process by Proposition 3.2.1 of Brockwell and Davis (1991). Thus, in terms of the second order structure, $\{Y_t\}$ is an ARMA$(m-1, m-1)$ process. However, it can be seen from the proof of Theorem 1 that if the vector $\nu^\tau = (\mathbb{E}(h(\sigma_1\eta_t)), \ldots, \mathbb{E}(h(\sigma_m\eta_t)))$ is orthogonal to some eigenvectors of P whose corresponding eigenvalues are less than 1 in magnitude, then the ARMA orders may be lowered; the latter happens only if ν lies in a set of zero Lebesgue measure. This exceptional case occurs, for example, if h is a linear function, in which case $\nu = 0$ and is orthogonal to all eigenvectors of P. Actually, in this case, $\{Y_t\}$ is an ARMA(0, 0) process. However, for a nonlinear transformation h, it is unlikely that ν is orthogonal to any eigenvector of P whose corresponding eigenvalue is less than 1 in magnitude. Thus, we have the generic result that any instantaneous nonlinear transformation of $\{X_t\}$ is an ARMA$(m-1, m-1)$ process. This result forms a basis for tentatively identifying the number of regimes of the T-CHARM. For example, we can consider the square of the X process and tentatively identify its ARMA orders by using existing methodologies such as the Extended ACF (EACF, see Tsay and Tiao (1984)) and others; see, e.g., Cryer and Chan (2008). The number of regimes m can be estimated by adding 1 to the estimated AR order. The identification can be verified by repeating the procedure after taking the absolute value of the original process. This discussion subsumes an earlier result of Gourieroux and Monfort (1992), who derived the ARMA representation for the special case $h(x) = x^2$.

The definition of regimes is likely to be application-specific. Empirically, a simple partition scheme consists of defining $R_j = \{x \in R : r_j < g(x) \leq r_{j+1}\}$, where $-\infty = r_0 < r_1 < r_2 < \cdots < r_{m-1} < r_m = \infty$, and g is some function, for example, the identity function or the absolute value function.

Let $I_t = (I_{2t}, \ldots, I_{mt})^\tau$, where $I_{it} = I\{X_t \in R_i\}$, $i = 2, \ldots, m$, and $I\{\cdot\}$ is the indictor function. Let $a_t = (I\{\sigma_1\eta_t \in R_2\}, \ldots, I\{\sigma_1\eta_t \in R_m\})^\tau$ be an $(m-1)$-dimensional vector and A_t be an $(m-1) \times (m-1)$ matrix whose (i, j)

entry is $I\{\sigma_{j+1}\eta_t \in R_{i+1}\} - I\{\sigma_1\eta_t \in R_{i+1}\}$. Then, we have $I_t = a_t + A_t I_{t-1}$ and after iterating k times, it follows that

$$I_t = \sum_{j=0}^{k} \Big(\prod_{i=0}^{j-1} A_{t-i} \Big) a_{t-j} + \Big(\prod_{i=0}^{k} A_{t-i} \Big) I_{t-k-1},$$

with the convention $\prod_{i=0}^{-1} = I$ and $\prod_{i=0}^{j} A_{t-i} = A_t A_{t-1} \cdots A_{t-j}$. Assume that η_t has a positive density on R and the Lebesgue measure of each R_i is positive. Since the matrices in the sequence $\{A_t\}$ are i.i.d. and $\mathbb{E}|I\{\sigma_j\eta_t \in R_i\} - I\{\sigma_1\eta_t \in R_i\}| < 1$, using the technique in the proof of Theorem 2.1 in Li, Ling, and Tong (2012), the first term converges to $\sum_{j=1}^{\infty} \big(\prod_{i=0}^{j-1} A_{t-i} \big) a_{t-j}$ almost surely (a.s.), while the second term converges to zero a.s. Then the infinite series is strictly stationary and ergodic since it is a function of i.i.d. $\{\eta_t\}$.

Theorem 2. *If the density of η_t, $f(\cdot)$, is positive on R and the Lebesgue measure of each R_i is positive, then* (i)

$$I_t = \sum_{j=0}^{\infty} \Big(\prod_{i=0}^{j-1} A_{t-i} \Big) a_{t-j}$$

and $\sigma(X_t) = \sigma_1 + (\sigma_2 - \sigma_1, \ldots, \sigma_m - \sigma_1) I_{t-1}$; (ii) *model* (2.1) *is strictly stationary and uniformly ergodic.*

The uniform ergodicity of the process follows from Corollary 6.12 of Nummelin (1984) and the fact that $p(x, y)$, the transition probability density function of $\{X_t\}$ (w.r.t. the Lebesgue measure), is bounded below by $K \times f(y)$ for all $x, y \in R$, where $0 < K < \infty$ is the minimum of $\sigma(\cdot)$. We do not impose any condition on $\sigma_i > 0$ in Theorem 2 and the results hold even when $\mathbb{E}|\eta_t| = \infty$. Theorem 2 significantly improves Proposition 1 of Gourieroux and Monfort (1992), who obtained the strict stationarity of the process under similar regularity conditions.

We further study the autocorrelation structure of the volatilities $\{\sigma^2(X_t)\}$ in model (2.1). For $k \geq 0$, simple calculations have

$$\operatorname{cov}(\sigma^2(X_t), \sigma^2(X_{t-k})) = \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_i^2 \sigma_j^2 \delta_{ij}^{(k)},$$

where $\delta_{ij}^{(k)}$ satisfies the iterative equations

$$\delta_{ij}^{(k)} = \sum_{s=1}^{m} \mathbb{P}(\sigma_s \eta_t \in R_i) \delta_{sj}^{(k-1)},$$
$$\delta_{ij}^{(0)} = \mathbb{P}(X_t \in R_i \cap R_j) - \mathbb{P}(X_t \in R_i)\mathbb{P}(X_t \in R_j).$$

For $k = 0$, we get

$$\text{var}(\sigma^2(X_t)) = \sum_{1 \le i < j \le m} (\sigma_j^2 - \sigma_i^2)^2 \mathbb{P}(X_t \in R_j)\mathbb{P}(X_t \in R_i).$$

Here, $\mathbb{P}(X_t \in R_i)$'s can be uniquely determined from

$$\mathbb{P}(X_t \in R_i) = \sum_{j=1}^{m} \mathbb{P}(\sigma_j \eta_t \in R_i)\mathbb{P}(X_t \in R_j) \quad \text{and} \quad \sum_{i=1}^{m} \mathbb{P}(X_t \in R_i) = 1.$$

Thus, it is not hard to obtain the ACF $\{\rho_k\}$ of $\sigma^2(X_t)$ in principle, although the general expression can be complicated. However, for the case $m = 2$, we have simple expressions.

Theorem 3. *Suppose that $m = 2$ and the assumptions in Theorem 2 hold. Then, for $k \ge 0$*

1. $\text{cov}(\sigma^2(X_t), \sigma^2(X_{t-k})) = (\sigma_2^2 - \sigma_1^2)^2 \delta(1-\delta)\{\mathbb{P}(\sigma_2\eta_t \in R_2) - \mathbb{P}(\sigma_1\eta_t \in R_2)\}^k$, *where* $\delta = \mathbb{P}(\sigma_1\eta_t \in R_2)/\{1 - \mathbb{P}(\sigma_2\eta_t \in R_2) + \mathbb{P}(\sigma_1\eta_t \in R_2)\}$;
2. $\rho_k = \{\mathbb{P}(\sigma_2\eta_t \in R_2) - \mathbb{P}(\sigma_1\eta_t \in R_2)\}^k$.

We compare the ACFs of T-CHARM with those of GARCH models. For simplicity we consider GARCH (1,1) model $\sigma_t^2 = \alpha_0 + (\alpha\eta_t^2 + \beta)\sigma_{t-1}^2$, where $\alpha > 0$ and $\beta > 0$. The corresponding ACF $\{\rho_{gk}\}$ is $\rho_{gk} = (\alpha + \beta)^k$, where $\alpha + \beta < 1$. It is well-known that the estimation of ρ_{gk} requires a finite fourth moment condition, $2\alpha^2 + (\alpha + \beta)^2 < 1$, and this condition is rarely satisfied by the estimated parameters. In contrast, we do not restrict the parameters for the ρ_k of T-CHARM. We suggest that model (2.1) can offer a reliable alternative for modelling volatilities. Additionally, it can capture the heavy-tailed property of financial time series because

$$\frac{\mathbb{E}X_t^4}{(\mathbb{E}X_t^2)^2} = (\mathbb{E}\eta_t^4)\frac{\sum_{i=1}^m \sigma_i^4 \mathbb{P}(X_{t-1} \in R_i)}{\{\sum_{i=1}^m \sigma_i^2 \mathbb{P}(X_{t-1} \in R_i)\}^2} \ge \mathbb{E}\eta_t^4,$$

by Jensen's inequality.

The inequality is strict for $m \ge 2$ if the innovations have infinite support and $\{\sigma_i,\, i = 1, \ldots, m\}$ is not a singleton, implying that the T-CHARM(1) generally has a heavier tail than the innovations. For illustration, consider the models

$$y_t = [2I\{y_{t-1} \le 0\} + \sigma I\{y_{t-1} > 0\}]\eta_t, \tag{2.4}$$

$$y_t = [I\{y_{t-1} \le r\} + 2I\{y_{t-1} > r\}]\eta_t, \tag{2.5}$$

with i.i.d. standard normal η_t. The left diagram in Figure 1 plots the theoretical kurtosis of the stationary T-CHARM(1) process as a function of σ for the model

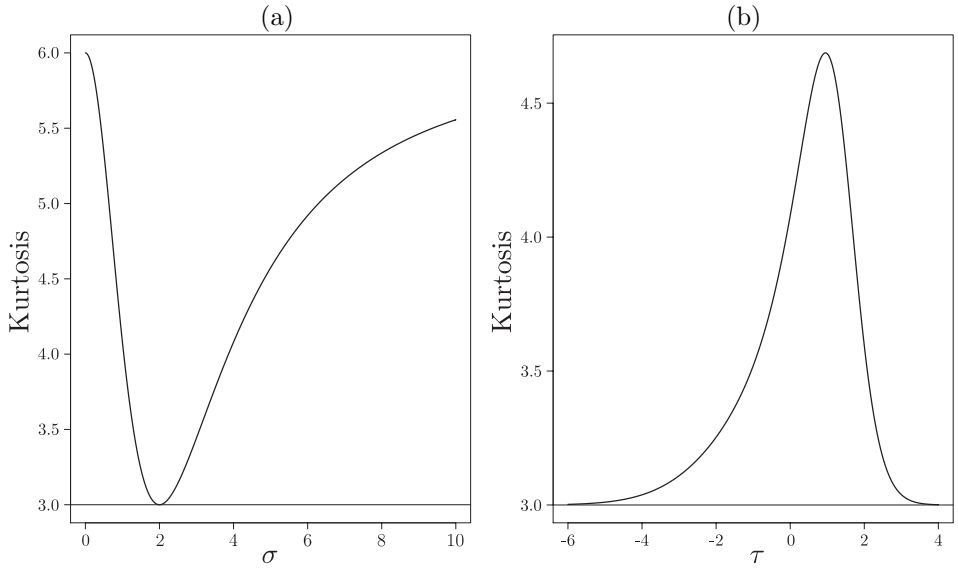

Figure 1. The kurtosis of models (2.4) and (2.5).

at (2.4) while the right figure plots that of the model at (2.5). For both models, the kurtosis exceeds that of the innovations.

In addition to conditional heteroscedasticity, the data (Y_t) may also admit a nonlinear conditional mean structure. Thus, it is of interest to consider a nonlinear autoregressive model with stochastic inputs:

$$Y_t = H(Y_{t-1}, \ldots, Y_{t-p}) + X_t, \tag{2.6}$$

where (X_t) is some stochastic process, e.g., a T-CHARM. The autoregressive function H then accounts for the conditional mean structure while the stochastic input $\{X_t\}$ is generally a white noise that models the conditional heteroscedasticity. In the applications here, we employ linear autoregressive functions with T-CHARM inputs.

It is instructive to examine the case $p = 1$. Backward iteration of (2.6) yields

$$\begin{aligned} Y_t &= X_t + H(Y_{t-1}) \\ &= X_t + H(X_{t-1} + H(X_{t-2} + H(\ldots + X_{t-k} + H(Y_{t-k-1})\ldots))). \end{aligned}$$

Suppose $Y_{t-k-1} = y_0$ and take $Y_{t,k}(y_0) = X_t + H(X_{t-1} + H(X_{t-2} + \ldots + H(X_{t-k} + y_0)\ldots))$. It is plausible that, under some regularity conditions, $Y_{k,t}$ converges to Y_t in some sense, which then provides the latter an expansion in terms of the stochastic inputs. These arguments can be extended to the higher order case by vectorizing (2.6):

$$\mathbf{Y}_t = \mathbf{H}(\mathbf{Y}_{t-1}) + \mathbf{X}_t,$$

where $\mathbf{Y}_t=(Y_t, Y_{t-1}, \ldots, Y_{t-p+1})^\tau$, $\mathbf{H}(u_1, \ldots, u_p)=(H(u_1, \ldots, u_p), u_1, \ldots, u_{p-1})^\tau$ and $\mathbf{X}_t = (X_t, 0, \ldots, 0)^\tau$. Similarly, $\mathbf{Y}_{t,k}(\mathbf{y}_0)$ is defined.

Theorem 4. *Let* $\{Y_t\}$ *satisfy* (2.6). *Suppose that* (i) *there exist* $\mathbf{y}_0$*, positive constants* α, C*, and* $\zeta < 1$ *such that for all positive integer* k*, for all vectors* $\mathbf{y}_0$ *and* $\mathbf{y}$*,*

$$\mathbb{E}\{|\mathbf{Y}_{t,k}(\mathbf{y}_0) - \mathbf{Y}_{t,k}(\mathbf{y})|^\alpha | \mathcal{F}_{t-k-1}\} \le C\zeta^k |\mathbf{y}_0 - \mathbf{y}|^\alpha, \tag{2.7}$$

where $|\cdot|$ *is the Euclidean norm and* $\mathcal{F}_t$ *is the* σ*-algebra generated by* $\{X_{t-\ell}, \ell \ge 0\}$*, and* (ii)

$$D = \sup_t \mathbb{E}|\mathbf{X}_t + H(\mathbf{y}_0) - \mathbf{y}_0|^\alpha < \infty. \tag{2.8}$$

Then there exist random vectors W_t *that are* $\mathcal{F}_t$*-measurable and do not depend on* $\mathbf{y}_0$ *such that, for any* $\mathbf{y}$*,* $\mathbf{Y}_{t,k}(\mathbf{y})$ *converges to* W_t *as* $k \to \infty$ *a.s. and* $\mathbb{E}|\mathbf{Y}_{t,k}(\mathbf{y}) - W_t|^\alpha = O(\zeta^k)$*. Moreover, if* $\{X_t\}$ *is stationary ergodic and* $\mathbf{Y}_0 = W_0$*, then* $\{Y_t, t \ge 0\}$ *is stationary and ergodic.*

See the *Supplementary Materials* for a proof. The preceding theorem generalizes Theorem 2 in Wu and Shao (2004) who considered the problem of iterating i.i.d. random maps; this subsumes the case of i.i.d. $\{X_t\}$ in (2.6). We note that condition (i) holds if there exist a positive constant $\kappa < 1$ and a vector norm $|\cdot|$ such that for any two vectors $u = (u_1, \ldots, u_p)^\tau$ and $v = (v_1, \ldots, v_p)^\tau$, $|\mathbf{H}(u) - \mathbf{H}(v)| \le \kappa |u - v|$. Such a Lipschitz condition holds for any stationary linear autoregressive function with none of its characteristic roots on the unit circle. It also holds for some nonlinear autoregressive function, e.g. exponentially stable, continuous threshold autoregressive functions. For instance, the latter condition holds if $p = 1$ and $H(u) = \phi_0 + \phi_1(u-r)_- + \phi_2(u-r)_+$, where the parameters $\phi_i, i = 0, 1, 2$, and r satisfy $\phi_1 < 1, \phi_2 < 1$ and $\phi_1\phi_2 < 1$, and we use the convention that for any real number x, $x_- = \min(x, 0)$ and $x_+ = \max(x, 0)$. For results with $p > 1$, see Tong (1990). Condition (ii) is a mild moment condition on the stochastic inputs; indeed, it trivially holds for T-CHARM inputs.

3. Quasi-maximum Likelihood Estimation

A simple way to implement (2.1) is to adopt the form $R_i = (r_{i-1}, r_i]$. Here, we consider a slightly more general model in which the threshold variable is allowed to be a functional of the past information. Take

$$\begin{aligned} X_t &= \sigma(W_{t-1})\eta_t, \\ \sigma(W_{t-1}) &= \sum_{i=1}^m \sigma_i I\{r_{i-1} < W_{t-1} \le r_i\}, \end{aligned} \tag{3.1}$$

where W_{t-1} is a known functional of $\{X_{t-1}, \ldots, X_{t-p}\}$, m is the number of regimes, the σ_i's are positive numbers, and $-\infty = r_0 < r_1 < \cdots < r_{m-1} < r_m = \infty$, $\{r_1, \ldots, r_{m-1}\}$ the threshold parameters. We take m to be known. Let $\theta = (\sigma_1^2, \ldots, \sigma_m^2)^\tau$ and $\mathbf{r} = (r_1, \ldots, r_{m-1})^\tau$ with $\theta_0 = (\sigma_{10}^2, \ldots, \sigma_{m0}^2)^\tau$ and $\mathbf{r}_0 = (r_{10}, \ldots, r_{m-1,0})^\tau$ as the true values of θ and $\mathbf{r}$. We do not assume that η_t is normally distributed. To estimate the parameters from data $\{X_1, \ldots, X_n\}$ given the initial values $\{X_{-p}, \ldots, X_0\}$, we adopt the objective function

$$L_n(\theta, \mathbf{r}) = -\frac{1}{2} \sum_{t=1}^{n} \sum_{i=1}^{m} \Big(\log \sigma_i^2 + \frac{X_t^2}{\sigma_i^2} \Big) I_{it} = I_{it}(\mathbf{r}), \tag{3.2}$$

where $I_{it} = I_{it}(\mathbf{r}) = I\{r_{i-1} < W_{t-1} \le r_i\}$. For each $\mathbf{r}$, it is easy to maximize $L_n(\theta, \mathbf{r})$ with respect to θ, say $\hat{\theta}_n(\mathbf{r}) \equiv (\hat{\sigma}_{1n}^2(\mathbf{r}), \ldots, \hat{\sigma}_{mn}^2(\mathbf{r}))^\tau$ with

$$\hat{\sigma}_{in}^2(\mathbf{r}) = \frac{\sum_{t=1}^{n} X_t^2 I_{it}}{\sum_{t=1}^{n} I_{it}}, \quad i = 1, \ldots, m.$$

There are at most finitely many different values of $L_n(\hat{\theta}_n(\mathbf{r}), \mathbf{r})$. By the enumeration approach, we estimate $\mathbf{r}_0$ by

$$\hat{\mathbf{r}}_n = \arg\max_{\mathbf{r}} L_n(\hat{\theta}_n(\mathbf{r}), \mathbf{r}).$$

Using the plug-in method, the estimator of θ_0 is $\hat{\theta}_n = \hat{\theta}_n(\hat{\mathbf{r}}_n)$. We call $(\hat{\theta}_n, \hat{\mathbf{r}}_n)$ the quasi-maximum likelihood estimator (QMLE) of $(\theta_0, \mathbf{r}_0)$.

Generally, $\hat{\mathbf{r}}_n$ takes the form $(W_{(i_1)}, \ldots, W_{(i_{m-1})})^\tau$, where $i_1 < \cdots < i_{m-1}$ and $\{W_{(1)}, \ldots, W_{(n)}\}$ are the order statistics of $\{W_1, \ldots, W_n\}$. If $(W_{(j_1)}, \ldots, W_{(j_{m-1})})^\tau$ is an estimator of $\mathbf{r}_0$, then $L_n(\hat{\theta}_n(\mathbf{r}), \mathbf{r})$ is constant over the $(m-1)$-dimensional cube $\widetilde{R}$, where

$$\widetilde{R} = \{\mathbf{r} = (r_1, \ldots, r_{m-1})^\tau : \ r_i \in [W_{(j_i)}, W_{(j_i+1)}), \ i = 1, \ldots, m-1\}.$$

Thus, there exist infinitely many $\mathbf{r}$ such that $L_n(\cdot)$ can achieve its global maximum and any $\mathbf{r} \in \widetilde{R}$ can be taken as an estimator of $\mathbf{r}_0$. In this case, we choose $(W_{(j_1)}, \ldots, W_{(j_{m-1})})^\tau$ as a representative of $\widetilde{R}$ and denote it as the estimator of $\mathbf{r}_0$. With this, it is not hard to show that $(\hat{\theta}_n, \hat{\mathbf{r}}_n)$ is the QMLE of $(\theta_n, \mathbf{r}_n)$, so

$$(\hat{\theta}_n, \hat{\mathbf{r}}_n) = \arg\max_{\Theta \times \mathcal{R}} L_n(\theta, \mathbf{r}),$$

where $\Theta \times \mathcal{R}$ is the parameter space and $\Theta = R_+^m$ with $R_+ \equiv (0, \infty)$ and $\mathcal{R} = \{\mathbf{r} : -\infty < r_1 < \cdots < r_{m-1} < \infty\}$.

Assumption 1. *The density $f(x)$ of η_t is continuous and positive on R, $\mathbb{E}\eta_t = 0$ and $\mathbb{E}\eta_t^2 = 1$.*

Assumption 2. *The density $f_w(\cdot)$ of W_t is continuous and $f_w(r_{j0}) > 0$ for $j = 1, \ldots, m-1$.*

The proof of the strong consistency of $(\hat{\theta}_n, \hat{\mathbf{r}}_n)$ is similar to that of Theorem 1 in Chan (1993) and is therefore omitted.

Theorem 5. *If* (i) *Assumptions* 1 *and* 2 *hold, and* (ii) $\sigma_{i0}^2 \neq \sigma_{i+1,0}^2$ *for* $i = 1, \ldots, m-1$, *then* $(\hat{\theta}_n, \hat{\mathbf{r}}_n) \to (\theta_0, \mathbf{r}_0)$ *a.s. as* $n \to \infty$.

Condition (ii) in Theorem 5 is required to ensure the identifiability of $\mathbf{r}_0$. By a technique similar to that used in the proof of Proposition 1 in Chan (1993), we establish the convergence rate of $\hat{\mathbf{r}}_n$ and the asymptotic normality of $\hat{\theta}_n$.

Theorem 6. *Under the conditions of Theorem* 5, *if* $\sup_{x \in R}\{(1+|x|)f(x)\} < \infty$ *and* $\kappa_4 \equiv \mathbb{E}\eta_t^4 < \infty$, *then*

(a) $n(\hat{\mathbf{r}}_n - \mathbf{r}_0) = O_p(1)$;

(b) $\sqrt{n} \sup_{|\mathbf{r}-\mathbf{r}_0| \leq B/n} |\hat{\sigma}_{in}^2(\mathbf{r}) - \hat{\sigma}_{in}^2(\mathbf{r}_0)| = o_p(1)$ *for any fixed* $B \in (0, \infty)$.

Furthermore,

$$\sqrt{n}(\hat{\sigma}_{in}^2(\mathbf{r}_0) - \sigma_{i0}^2) \Longrightarrow \mathcal{N}\Big(0,\ \frac{(\kappa_4 - 1)\sigma_{i0}^4}{F_w(r_{i0}) - F_w(r_{i-1,0})}\Big),\ i = 1, \ldots, m,$$

and all the normalized estimators are asymptotically independent, where $F_w(x)$ is the cumulative distribution function of W_t, and henceforth the symbol $\Longrightarrow$ indicates weak convergence.

A proof is given in the *Supplementary Materials*. These results are similar to those of Chan (1993). See also Li and Ling (2012).

4. Inference for the Threshold Parameter $\mathbf{r}_0$

To study the limiting distribution of $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$, we consider the profile log-likelihood process

$$\widetilde{L}_n(\mathbf{s}) = -2\Big\{L_n\big(\hat{\theta}_n(\mathbf{r}_0 + \frac{\mathbf{s}}{n}),\ \mathbf{r}_0 + \frac{\mathbf{s}}{n}\big) - L_n\big(\hat{\theta}_n(\mathbf{r}_0),\ \mathbf{r}_0\big)\Big\}, \tag{4.1}$$

where $\mathbf{s} = (s_1, \ldots, s_{m-1})^\tau \in R^{m-1}$.

Let $\mathbb{D}(R^{m-1})$ denote the function space consisting of uniform limits of sequences of simple functions defined on R^{m-1} that is equipped with the Skorokhod metric (see Seijo and Sen (2011); Li and Ling (2012)). By Theorem 6 and Taylor's

expansion, $\widetilde{L}_n(\mathbf{s})$ can be approximated in $\mathbb{D}(R^{m-1})$ by

$$\begin{aligned}\wp_n(\mathbf{s}) &= L_n(\theta_0, \mathbf{r}_0 + \frac{\mathbf{s}}{n}) - L_n(\theta_0, \mathbf{r}_0) \\ &= \sum_{i=1}^{m-1}\sum_{t=1}^{n}\Big[\xi_t^{(i+1,i)} I\big\{r_{i0} + \frac{s_i}{n} < W_{t-1} \le r_{i0}\big\} I\{s_i < 0\} \\ &\quad + \xi_t^{(i,i+1)} I\big\{r_{i0} < W_{t-1} \le r_{i0} + \frac{s_i}{n}\big\} I\{s_i \ge 0\}\Big],\end{aligned}$$

where

$$\xi_t^{(i,j)} = \log\frac{\sigma_{i0}^2}{\sigma_{j0}^2} + \Big(\frac{\sigma_{j0}^2}{\sigma_{i0}^2} - 1\Big)\eta_t^2, \quad i, j = 1, \ldots, m. \tag{4.2}$$

We define $m-1$ independent one-dimensional two-sided compound Poisson processes $\{\mathcal{P}_j(z), z \in R\}$ as

$$\mathcal{P}_j(z) = I\{z < 0\} \sum_{k=1}^{N_1^{(j)}(|z|)} U_k^{(j+1,j)} + I\{z \ge 0\} \sum_{k=1}^{N_2^{(j)}(z)} V_k^{(j,j+1)}, \tag{4.3}$$

for $j = 1, \ldots, m-1$, where $\{N_1^{(j)}(z), z \ge 0\}$ and $\{N_2^{(j)}(z), z \ge 0\}$ are independent Poisson processes with $N_1^{(j)}(0) = N_2^{(j)}(0) = 0$ a.s. and with the same jump rate $f_w(r_{j0})$. Here $\{U_k^{(i,j)}\}_{k=1}^{\infty}$ and $\{V_k^{(i,j)}\}_{k=1}^{\infty}$ are mutually independent copies of $\xi_1^{(i,j)}$. We work with the left continuous version of $N_1^{(j)}(z)$ and the right continuous version of $N_2^{(j)}(z)$.

Consider the spatial compound Poisson process

$$\wp(\mathbf{s}) = \sum_{j=1}^{m-1} \mathcal{P}_j(s_j), \quad \mathbf{s} = (s_1, \ldots, s_{m-1})^\tau \in R^{m-1}. \tag{4.4}$$

Clearly, $\wp(\mathbf{s}) \to \infty$ a.s. as $|\mathbf{s}| \to \infty$ since $\mathbb{E}U_t^{(i+1,i)} > 0$ and $\mathbb{E}V_t^{(i,i+1)} > 0$. Therefore, there exists a unique random $(m-1)$-dimensional cube $[\mathbf{M}_-, \mathbf{M}_+) \equiv [M_-^{(1)}, M_+^{(1)}) \times \cdots \times [M_-^{(m-1)}, M_+^{(m-1)})$ at which the process $\{\wp(\mathbf{s}), \mathbf{s} \in R^{m-1}\}$ attains its global minimum a.s.,

$$[\mathbf{M}_-, \mathbf{M}_+) = \arg\min_{\mathbf{s} \in R^{m-1}} \wp(\mathbf{s}).$$

From (4.4), the minimization is equivalent to

$$[M_-^{(j)}, M_+^{(j)}) = \arg\min_{z \in R} \mathcal{P}_j(z), \quad j = 1, \ldots, m-1.$$

Note that the processes $\{\mathcal{P}_j(z)\}$ are independent, and so are $\{M_-^{(j)}\}$, $j = 1, \ldots, m-1$. Modifying slightly the proof of Theorem 3.3 in Li and Ling (2012), we can prove the following

Theorem 7. *If the conditions in Theorem* 6 *hold, then* $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$ *converges weakly to* $\mathbf{M}_-$ *and its components are asymptotically independent as* $n \to \infty$. *Furthermore,* $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$ *is asymptotically independent of* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *which is asymptotically normal.*

We describe how to implement $M_-^{(j)}$ or $\mathbf{M}_-$. From (4.2) and (4.3), we know that the density of $M_-^{(j)}$ is determined by the jump rate and the jump distributions. We can simulate $M_-^{(j)}$ by simulating the compound Poisson process $\mathcal{P}_j(z)$ in (4.3) on the interval $[-T, T]$ for any given $T > 0$ large enough since the expectations of the jumps $U_k^{(j+1,j)}$ and $V_k^{(j,j+1)}$ are positive. Modifying Algorithm 6.2 of Cont and Tankov (2004, p.174) for a one-sided compound Poisson process, we have an algorithm for a two-sided compound Poisson process.

Algorithm

Step 1. Sample $N_1^{(j)}$ and $N_2^{(j)}$ from Poisson distributions with the same parameter $f_w(r_{j0})T$ as the total number of jumps on the intervals $[-T, 0]$ and $[0, T]$, respectively.

Step 2. Sample independent jump time sequences $\{U_1, \ldots, U_{N_1^{(j)}}\}$ and $\{V_1, \ldots, V_{N_2^{(j)}}\}$, where $\{U_i\} \overset{i.i.d.}{\sim} U[-T, 0]$ and $\{V_i\} \overset{i.i.d.}{\sim} U[0, T]$. Here $U[a, b]$ denotes the uniform distribution on the interval $[a, b]$.

Step 3. Sample mutually independent sequences $\{\eta_1, \ldots, \eta_{N_1^{(j)}}\}$ and $\{\eta_1, \ldots, \eta_{N_2^{(j)}}\}$ from $f(x)$. Use them to produce mutually independent jump-size sequences $\{Y_1, \ldots, Y_{N_1^{(j)}}\}$ and $\{Z_1, \ldots, Z_{N_2^{(j)}}\}$ by (4.2), respectively.

For $z \in [-T, T]$, the trajectory of (4.3) is given by

$$\mathcal{P}_j(z) = I\{z < 0\} \sum_{i=1}^{N_1^{(j)}} I\{U_i > z\} Y_i + I\{z \geq 0\} \sum_{j=1}^{N_2^{(j)}} I\{V_j < z\} Z_j. \tag{4.5}$$

Then we take the smallest minimizer of $\mathcal{P}_j(z)$ in (4.5) on $[-T, T]$ as an observation of $M_-^{(j)}$. By repeating the algorithm, we can get a sequence of observations of $M_-^{(j)}$, from which we can infer the distribution of $n(\hat{r}_{jn} - r_{j0})$.

In practice, however, since only one sample $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ is available, we can use it to estimate θ_0 and $f_w(r_{j0})$ by $\hat{\theta}_n$ and $\widehat{f}_w(\hat{r}_{jn})$, where $\widehat{f}_w(\cdot)$ is a kernel density estimator of $f_w(\cdot)$. Then we can calculate the residuals $\{\hat{\eta}_1, \ldots, \hat{\eta}_n\}$ and use them to construct a kernel density estimator $\widehat{f}(\cdot)$ of $f(\cdot)$.

When θ_0, $f_w(r_{j0})$, and $f(x)$ are all unknown, we substitute their consistent estimators $\hat{\theta}_n$, $\widehat{f}_w(\hat{r}_{jn})$, and $\widehat{f}(x)$ in the Algorithm and denote the corresponding compound Poisson processes as $\{\widehat{\mathcal{P}}_j(z)\}$. Then we can get an approximation $\widehat{M}_-^{(j)}$

of $M_-^{(j)}$. By Theorem 16 in Pollard (1984, p.134), $\widehat{\mathcal{P}}_j(z) \Longrightarrow \mathcal{P}_j(z)$ conditionally on $\mathcal{X}_n$ in $\mathbb{D}(R)$, in probability. By Theorem 3.1 (on the continuity of the smallest argmax functional) in Seijo and Sen (2011), $\widehat{M}_-^{(j)} \Longrightarrow M_-$ conditionally on $\mathcal{X}_n$, in probability.

Theorem 8. *If the conditions in Theorem* 6 *hold then, in probability,*

$$\lim_{n\to\infty} |\mathbb{P}(\widehat{M}_-^{(j)} \le x|\mathcal{X}_n) - \mathbb{P}(M_-^{(j)} \le x)| = 0$$

at each x *for which* $\mathbb{P}(M_-^{(j)} = x) = 0$. *That is,* $\widehat{M}_-^{(j)}|_{\mathcal{X}_n} \Longrightarrow M_-^{(j)}$, *in probability.*

To illustrate the efficacy of the Algorithm, we consider the simple example

$$X_t = [2I\{X_{t-1} \le 0\} + 0.5I\{X_{t-1} > 0\}]\eta_t, \tag{4.6}$$

where η_t is i.i.d. standard normal and the sample size is 400. In Figure 8 in the *Supplementary Materials*, (a) gives the density of $n(\hat{r}_n - r_0)$, obtained by 10,000 replications; (b) shows the density of M_- when θ_0, $f_w(r_0)$, and $f(x)$ are all known. When a sample $\mathcal{X} = \{x_1, \ldots, x_{400}\}$ is given and fixed, (c) and (d) display the density of $\widehat{M}_-$. Here, 1,000 replications were used for (c) and 10,000 replications for (d). Comparing (c) with (d) there, we find that the more the number of the replications, the more precise the density of $\widehat{M}_-$.

5. Test of T-CHARM (m) Against T-CHARM ($m+1$)

We denote a T-CHARM with m regimes by T-CHARM (m). In applications one needs to determine the number of regimes. This is a non-standard inferential problem and it is not obvious how to count the number of *independently* adjusted parameters. Here we consider testing T-CHARM(m) against T-CHARM($m+1$). Our experience suggests that m is unlikely to be much larger than 2 or 3 in applications. Specifically, under the null H_0, the T-CHARM is as in (3.1). Under the alternative H_1, there is an additional threshold, denoted by r, which lies in the k-regime $(r_{k-1}, r_k]$, such that $\sigma(W_{t-1})$ in the T-CHARM($m+1$) can be written as

$$\sigma(W_{t-1}) = \sum_{\substack{i=1\\ i\neq k}}^{m} \sigma_i^2 I_{it} + \sigma_{1k}^2 I\{r_{k-1} < W_{t-1} \le r\} + \sigma_{2k}^2 I\{r < W_{t-1} \le r_k\}.$$

Under H_1, the log quasi-likelihood function (ignoring a constant) is

$$L_n(\tilde{\theta}, \mathbf{r}, \sigma_{1k}, \sigma_{2k}, r) = -\frac{1}{2}\sum_{t=1}^{n}\Bigg\{\sum_{\substack{i=1\\ i\neq k}}^{m}\Big(\log\sigma_i^2 + \frac{X_t^2}{\sigma_i^2}\Big)I_{it} + \Big(\log\sigma_{1k}^2 + \frac{X_t^2}{\sigma_{1k}^2}\Big)I_{1kt} + \Big(\log\sigma_{2k}^2 + \frac{X_t^2}{\sigma_{2k}^2}\Big)I_{2kt}\Bigg\},$$

where $\tilde{\theta} = (\sigma_1^2, \cdots \sigma_{k-1}^2, \sigma_{k+1}^2, \ldots, \sigma_m^2)^\tau$, $I_{1kt} = I_{1kt}(r_{k-1}, r) = I\{r_{k-1} < W_{t-1} \le r\}$ and $I_{2kt} = I_{2kt}(r, r_k) = I\{r < W_{t-1} \le r_k\}$. For each r, the profile log-likelihood ratio test statistic is

$$\begin{aligned} L_{kn}(r) &= L_n(\hat{\tilde{\theta}}, \hat{\mathbf{r}}, \hat{\sigma}_{1k}(r), \hat{\sigma}_{2k}(r), r) - L_n(\hat{\theta}, \hat{\mathbf{r}}) \\ &= -\frac{1}{2}\sum_{t=1}^{n}\Big[\Big\{\log\hat{\sigma}_k^2(\hat{\mathbf{r}}) + \frac{X_t^2}{\hat{\sigma}_k^2(\hat{\mathbf{r}})}\Big\}I_{kt}(\hat{\mathbf{r}}) + \Big\{\log\hat{\sigma}_{1k}^2(r) + \frac{X_t^2}{\hat{\sigma}_{1k}^2(r)}\Big\}I_{1kt}(\hat{r}_{k-1}, r) \\ &\quad + \Big\{\log\hat{\sigma}_{2k}^2(r) + \frac{X_t^2}{\hat{\sigma}_{2k}^2(r)}\Big\}I_{2kt}(r, \hat{r}_k)\Big], \end{aligned}$$

where

$$\hat{\sigma}_{1k}^2(r) = \frac{\sum_{t=1}^n X_t^2 I_{1kt}}{\sum_{t=1}^n I_{1kt}} \quad \text{and} \quad \hat{\sigma}_{2k}^2(r) = \frac{\sum_{t=1}^n X_t^2 I_{2kt}}{\sum_{t=1}^n I_{2kt}}.$$

Routine analysis then yields

$$\begin{aligned} 2L_{kn}(r) &= A_k(r)\Big\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\eta_t^2 - 1)I_{1kt} - \frac{F_w(r) - F_w(r_k)}{F_w(r_{k+1}) - F_w(r_k)}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\eta_t^2 - 1)I_{kt}\Big\}^2 \\ &\quad + o_p(1) \\ &= A_k(r)\{D_{kn}(r)\}^2 + o_p(1), \end{aligned}$$

where

$$A_k(r) = \frac{F_w(r_k) - F_w(r_{k-1})}{\{F_w(r) - F_w(r_{k-1})\}\{F_w(r_k) - F_w(r)\}},$$

and $o_p(1)$ holds uniformly in $r \in [r_{i-1}, r_i]$. By Theorem 3.1 in Ling and Tong (2011), $\{D_{kn}(r) : r \in [r_{i-1}, r_i]\}$ converges weakly to a centered Gaussian process $G_k(r) : r \in [r_{i-1}, r_i]\}$ with covariance kernel $K(r, s) = \mathrm{cov}(G_k(r), G_k(s))$ given by

$$\begin{aligned} (\kappa_4 - 1)\Big[&\min\{F_w(r) - F_w(r_{k-1}), F_w(s) - F_w(r_{k-1})\} \\ &- \frac{\{F_w(r) - F_w(r_{k-1})\}\{F_w(s) - F_w(r_{k-1})\}}{F_w(r_k) - F_w(r_{k-1})}\Big]. \end{aligned}$$

Thus, $\{F_w(r_k) - F_w(r_{k-1})\}^{-1/2} D_{kn}(r)$ converges weakly to a scalar multiple of the standard Brownian bridge $B_k(s) - sB_k(1)$, where $s = (F_w(r) - F_w(r_k))/(F_w(r_{k+1}) - F_w(r_k)) \in [0, 1]$. Since $\max_{r \in (r_{k-1}, r_k]} L_{kn}(r) = \infty$, we have to restrict the range of r to $[c_{1,k}, c_{2,k}]$ so that the corresponding s lies in $[a_{1,k}, a_{2,k}]$, a proper subset of $(0, 1)$, say $(0.05, 0.95)$, in each regime. In this way, we obtain a useful limiting distribution, similar to those in Bai and Perron (1998) for testing multiple change-point problems and Chan (1990) for testing a threshold AR model. Since we have m regimes under the null hypothesis, the LR test can be applied regime by regime, with an adjustment for multiple testing via the Bonferroni inequality.

Specifically, the LR test statistic for an additional threshold in the k-th regime equals

$$T_{k,n} = 2(\kappa_4 - 1)^{-1} \sup_{r \in [c_{1,k},\ c_{2,k}]} L_{kn}(r),$$

where $\kappa_4 \equiv \mathbb{E}\eta_t^4$ assumed finite.

Theorem 9. *Under H_0, if the density function of η_t is bounded and positive, then it follows that*

$$T_{k,n} \Longrightarrow \sup_{s \in [a_{1,k}, a_{2,k}]} \frac{(B_k(s) - sB_k(1))^2}{s(1-s)},$$

where the $B_k(s)$, $k = 1, \ldots, m$, are independent standard Brownian motions.

Since the test is applied regime by regime, consider the case $m = 1$ and the range of s be $[a, 1-a]$ for some $0 < a < 1/2$. Asymptotically, the null distribution of the square root of likelihood ratio test (after some further normalization) is equivalent to the distribution of the maximum of the normalized absolute Brownian bridge over the interval $[a, 1-a]$. Dirkse (1975) showed that the latter is the distribution of the maximum of the absolute value of a stationary Ornstein-Uhlenbeck process over the interval $[-\alpha, \alpha]$, where $\alpha = 0.5 \log(1/a - 1)$, and that

$$\mathbb{P}\Big(\sup_{-\alpha \le t \le \alpha} |U(t)| > c \Big) \sim \sqrt{\frac{2}{\pi}} \exp\Big(-\frac{c^2}{2} \Big) \Big(\alpha c - \frac{\alpha}{c} + \frac{2}{c} \Big) \tag{5.1}$$

for c large. (Here, we replace the typographical error of $1/c$ in Dirkse (1975) by $2/c$, thanks to Professor D. O. Siegmund.) Thus, we can compute the approximate p-value of the LRT using

$$p_0(\tilde{c}) = \sqrt{\frac{2}{\pi}} \exp\Big(-\frac{\tilde{c}^2}{2} \Big) \Big(\alpha \tilde{c} - \frac{\alpha}{\tilde{c}} + \frac{2}{\tilde{c}} \Big), \tag{5.2}$$

where $\tilde{c}$ is the square root of the (normalized) LRT statistic. Thus the p-value p_0 is asymptotically uniformly distributed over $[0, 1]$ in the sense that $\mathbb{P}(p_0 < p | H_0) \sim p$ for $p \to 0$.

We consider two modified LRTs that make use of the location of the quasi-likelihood estimator of the threshold. The first is based on the observation that, were the threshold value under the alternative known to be r_0, we could use a more powerful likelihood ratio test with the threshold fixed at r_0, whose asymptotic null distribution is χ_1^2. In practice, the threshold is unknown and we consider the likelihood ratio test with a wide range of possible threshold values, which reduces the power of the test. For the case of a Brownian bridge, it is known that its global maximum value is independent of the location where the maximum is attained, with the latter having a uniform distribution. This independence

property also holds, at least asymptotically, for the Ornstein-Uhlenbeck process restricted to a fixed finite interval, say $[h_1, h_2]$, and that the marginal distribution of the location of the global maximum has a uniform distribution over $[h_1, h_2]$; we sketch a proof in the *Supplementary Materials*. Hence, we can modify the calibration of the LRT as follows. Consider the LRT implemented with the threshold searched over the $a \times 100$ to $(1-a) \times 100$ percentiles of the threshold variable, and that the square root of the LRT attains its maximum value, say, $\tilde{c}$ at the β percentile. We then compute the p-value by (5.2) with α there replaced by $\beta_M = \log(1/\min(\beta, 1-\beta) - 1)$. If p_1 is the p-value so computed,

$$p_1(\tilde{c}, \beta) = \sqrt{\frac{2}{\pi}} \exp\Big(-\frac{\tilde{c}^2}{2}\Big)\Big(\beta_M \tilde{c} - \frac{\beta_M}{\tilde{c}} + \frac{2}{\tilde{c}}\Big).$$

We claim that p_1 is asymptotically uniformly distributed over [0, 1], so it provides a valid calibration of the LRT under the null hypothesis. To see this, recall that under our conjecture β is asymptotically independent of $\tilde{c}$, and $\log((1-\beta)/\beta)/2$ is uniformly distributed over $[-\alpha, \alpha]$. Thus, β_M is uniformly distributed over $[0, 2\alpha]$. The independence of $\tilde{c}$ and β then implies that $\mathbb{E}(p_1|\tilde{c}) = \sqrt{2/\pi}\exp(-\tilde{c}^2/2)(\alpha\tilde{c} - \alpha/\tilde{c} + 2/\tilde{c})$, and hence p_1 shares the same asymptotic behavior of p_0. For alternatives with the threshold parameter close to the 50th percentile, β is close to one half, rendering the calculation of p_1 to one based on a narrow interval around the 50th percentile, which then increases the power of the LRT in discerning threshold structure. On the other hand, if the threshold parameter is close to the data extremes, then p_1 is now calibrated on a very wide interval, thereby decreasing the power of detecting the alternative. A third approach is to compute the p-value as if the search was over the union of the intervals from $a \times 100$ to $\min(\beta, 1-\beta) \times 100$ percentiles and from $\max(\beta, 1-\beta) \times 100$ to $(1-a) \times 100$ percentiles. Effectively, this approach computes the p-value as

$$p_2(\tilde{c}, \beta) = \sqrt{\frac{2}{\pi}} \exp\Big(-\frac{\tilde{c}^2}{2}\Big)\Big(\beta_T \tilde{c} - \frac{\beta_T}{\tilde{c}} + \frac{2}{\tilde{c}}\Big),$$

where $\beta_T = \log(\min(\beta, 1-\beta)/(1 - \min(\beta, 1-\beta))) - \log(a/(1-a))$. We can similarly show that $\mathbb{E}(p_2|\tilde{c}) = p_0$.

We used simulation to examine the empirical properties of the three approaches for calibrating the p-value of the LRT. We simulated the threshold martingale difference process

$$X_t = \sigma(X_{t-1})\eta_t,$$

where $\sigma(X_{t-1}) = \{1 + \gamma I(X_{t-1} > r_0)\}\sigma^2$ and $\{\eta_t\}$ are independent standard normal random variables. The noise variance ratio ranged from 0.5 to 2, with

Table 1. Empirical size of the nominal 5% LRT with p-values computed by the proposed methods.

sample size	p_0	p_1	p_2
100	0.075	0.089	0.096
200	0.057	0.070	0.077
500	0.048	0.061	0.067
1,000	0.052	0.064	0.068

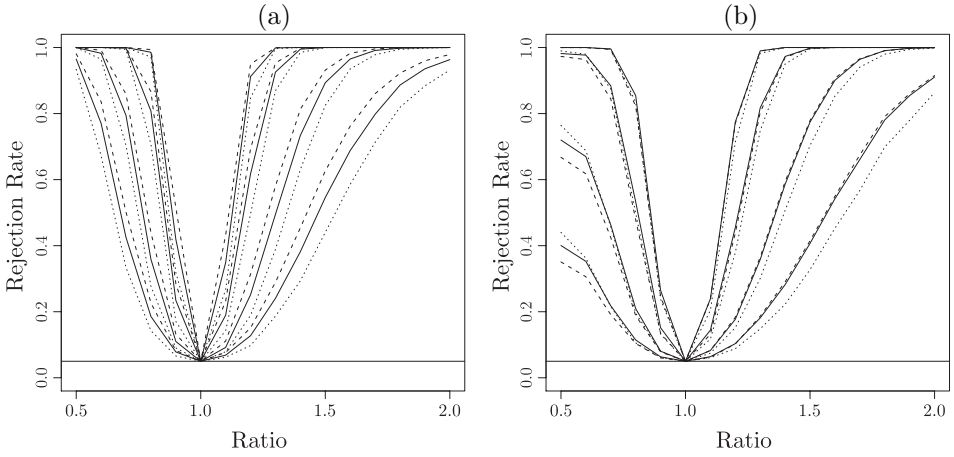

Figure 2. Empirical rejection rate of the LRT; all approaches for calculating the p-values are corrected to have exactly 5% empirical size by appropriately shifting the p-values. The threshold is 0 (-0.8) for the left (right) subfigure. Solid lines are the power curves for the LRT with the p-value computed by p_0; dashed lines are those for p_1 and dotted lines those for p_2. The four lines of each type, from bottom to top, correspond to sample size $n = 100$, 200, 500 and 1,000, respectively.

increment 0.1. We tried $r_0 = -0.8$ and 0. The threshold value -0.8 corresponds to the 6.6 percentile with $\gamma = 0.5$ but increases to the 30.7 percentile at $\gamma = 2$. On the other hand, the threshold value 0 ranges within the 49–50 percentiles as the noise variance ratio increases from 0.5 to 2. We considered sample sizes 100, 200, 400, and 1,000. All experiments were replicated 10,000 times. Table 1 displays the empirical sizes of the LRT with the p-values computed by the three proposed methods. For sample size 100, all result in higher rejection rates than the nominal 5% level, especially for p_1 and p_2. However, for sample sizes at least 200, the empirical sizes of the tests are increasingly closer to the nominal 5%. Figure 2 displays the empirical power curves of the LRT corresponding to the three methods of calibrating the p-values; in comparing the powers of the methods, we have corrected for slight size differences by shifting the p-values so that they all have 5% empirical size. All methods enjoy good power for

detecting threshold martingale differences but, as expected, p_1 has the highest power when the threshold is close to the median of the threshold variable, whereas p_2 dominates the other two approaches when the threshold is close to the extremes of the data.

6. Simulation Studies

Given the popularity of the GARCH model, it is pertinent to compare T-CHARM with GARCH on their empirical performances. We do this via a simulation study using three data generating mechanisms: a GARCH(1,1) model, a T-CHARM(2) model, and a stochastic volatility model.

The performance of a fitted model in predicting the 1-step ahead conditional variance can be measured by the sum of the absolute value of its (relative) volatility prediction errors,

$$d_{\text{model}} = \sum_{t=1}^{n} \left| \frac{X_t^2 - \hat{\sigma}_{t|t-1}^2}{\hat{\sigma}_{t|t-1}^2} \right|. \tag{6.1}$$

We use

$$\text{RATIO} = \frac{d_{\text{GARCH}}}{d_{\text{TCHARM}}}$$

for comparing the two models. All numerical results in this section are based on 1,000 replications, with sample size $n = 400$.

Data were generated from the GARCH(1, 1) model

$$\begin{aligned} X_t &= \sigma_t \eta_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned}$$

where $\{\eta_t\}$ is standard Gaussian white noise. Letting $\alpha_0 = 0.1$, $\beta_1 = 0.5$, and varying α_1 from 0 to 1, Figure 3(a) exhibits the empirical performance of the (fitted) GARCH (1, 1) relative to the (fitted) T-CHARM(m), for $m = 1, 2$. It shows that GARCH(1, 1) is inferior to T-CHARM over the range $0 \le \alpha_1 \le 0.2$, otherwise GARCH(1, 1) is superior.

Next, we generated data from the T-CHARM(2) model

$$y_t = [2I\{y_{t-1} \le 0\} + \sigma I\{y_{t-1} > 0\}]\eta_t$$

with σ varying from 1 to 3. Figure 3(b) displays the ratio of the absolute predictive volatility error of GARCH(p, p) to that of T-CHARM(2), as a function of σ, for $p = 1, 2$. Here GARCH(2, 2) had similar or slightly better performance than GARCH(1, 1), and both are consistently outperformed by T-CHARM, although GARCH improves its performance as σ approaches 2. (RATIO ≈ 1.005 at $\sigma = 2$.)

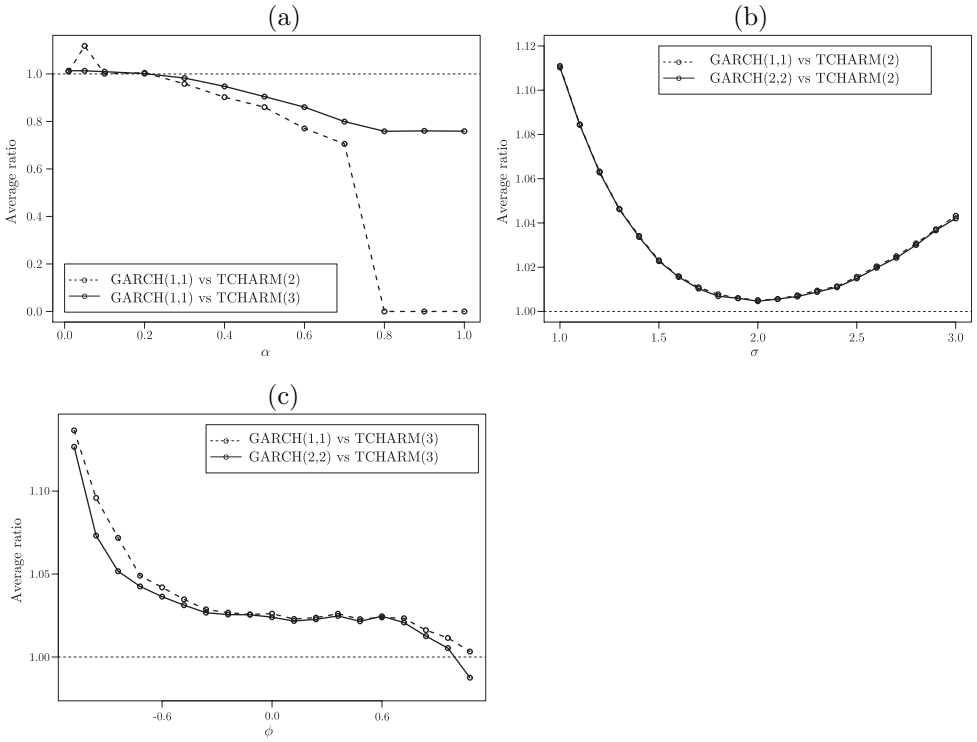

Figure 3. The ratio of the absolute volatility prediction error d_{model} of the GARCH model relative to the TCHARM.

Finally, we generated data from the stochastic volatility model

$$X_t = \eta_t \exp\left(\frac{h_t}{2}\right), \quad h_t = \phi h_{t-1} + e_t,$$

where both $\{\eta_t\}$ and $\{e_t\}$ are independent, standard Gaussian white noise, and the parameter ϕ varies from -0.9 to 0.9 with increment 0.1. Figure 3(c) shows, as a function of ϕ, the performance of GARCH(p, p) relative to T-CHARM(2) with threshold variable X_{t-1}^2. GARCH(2, 2) only slightly improves GARCH(1, 1) in terms of approximating the underlying stochastic volatility model. Except for ϕ close to 0.9, T-CHARM(2) consistently outperforms both GARCH(1, 1) and GARCH(2, 2).

The simulation study had GARCH (T-CHARM) performing better than the other model when GARCH (T-CHARM) was the true data mechanism, as expected. T-CHARM(3), however, provided better approximation than GARCH in the stochastic volatility model examined in our simulation study. Increasing the GARCH order only slightly enhanced its approximation capability, while approximation by T-CHARM was significantly improved by increasing the number of regimes.

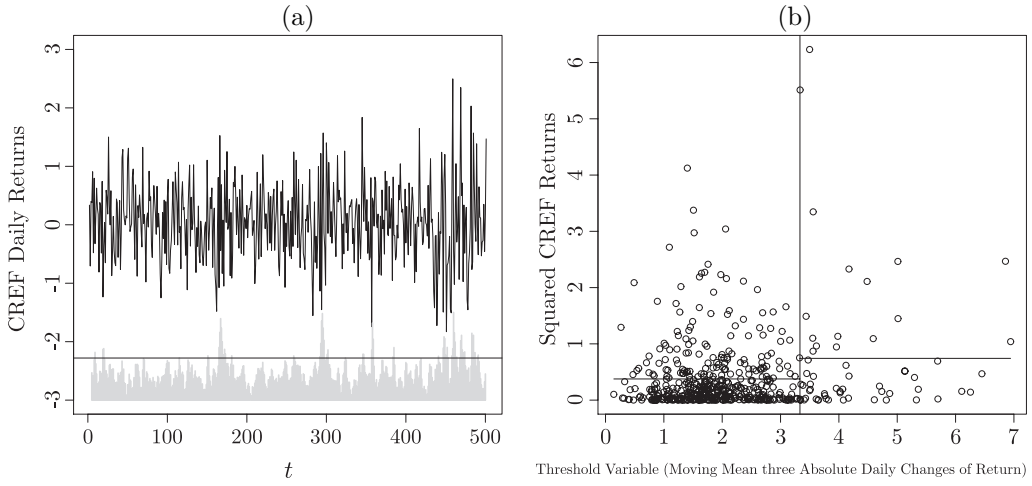

Figure 4. Left subfigure: CREF daily returns, black solid line; the threshold variables are plotted as vertical gray lines on the bottom with the heights of the lines proportional to the W's. The horizontal line indicates the estimated threshold. Right subfigure: Scatter diagram of the squared CREF daily returns versus the threshold variable. Vertical line separates the the two regimes and the horizontal lines indicate the estimated variances of the two regimes.

7. T-CHARM of Some Time Series

We illustrate the application of the T-CHARM with some time series.

The first example is a financial time series – the daily values of a unit of the CREF stock fund over the period from August 26, 2004 to August 15, 2006. The CREF stock fund comprises several thousand stocks. Since stocks are traded only on so-called trading days, the data do not change over the non-trading days. For simplicity, we analyze returns, namely the first differences of the logarithmic transformed daily values, as if they were equally spaced. The returns, denoted $\{X_t\}$, have been studied by Cryer and Chan (2008), who fitted a GARCH(1, 1) model in order to capture the conditional variance structure of the data. See Figure 4(a). Changes in the conditional variance of the innovations may be signified by substantial fluctuations in past returns. This suggests the potential of using a more complex threshold variable, for example a function of finitely many past returns, than some lag of X. We therefore consider threshold variables of the form $W_{t-1} = \sum_{j=1}^{k} |X_{t-j} - X_{t-j-1}|$, (The choice of $k = 3$ will be justified as being sufficient for the CREF example.) Specifically, we consider

$$X_t = \sum_{i=1}^{m} \sigma_i I\{r_{i-1} < W_{t-1} \le r_i\}\eta_t. \tag{7.1}$$

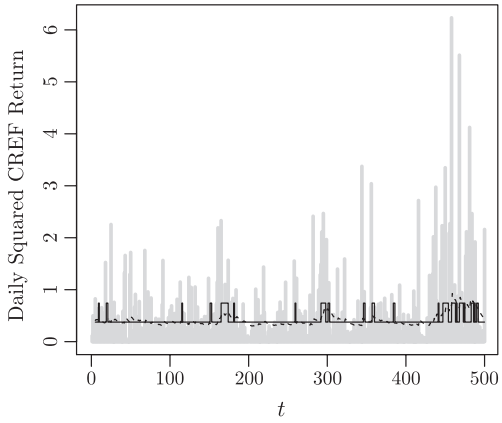

Figure 5. Fitted conditional variance process from T-CHARM (solid line) and that of the GARCH(1, 1) model (dashed line), with the squared CREF returns plotted as gray background.

We fit a two-regime T-CHARM with the threshold searched between the 5 and 95 percentiles of the threshold variable W_t where $k = 3$; by quasi-likelihood estimation, $\hat{\sigma}_1^2 = 0.3765(0.0272)$, $\hat{\sigma}_2^2 = 0.7420(0.147)$, and $\hat{r} = 3.333$, where the standard errors are enclosed in parentheses; see Figure 4(b). Based on the method of Section 4, and using the empirical standardized residual distribution, we obtain (2.256, 4.024) as a 95% confidence interval of the threshold parameter, which is asymmetric about the threshold estimate. Model diagnostics suggest the normality of the residuals. Assuming normality, the 95% confidence interval of the threshold parameter is (2.321, 4.144), which is close to the preceding confidence interval. Confidence intervals for $\sigma_i^2, i = 1, 2$, can be readily constructed by making use of Theorem 6, on the logarithmic scale with the aid of the delta method, then followed by back-transformation in order to ensure positivity. The presence of a threshold is justified by the LR test whose p-value is $p_0 = 0.018$. The other approaches for calculating the conditional p-values yield $p_1 = 0.025$ and $p_2 = 0.012$. The threshold is approximately the 88th percentile of the threshold variable with the number of data falling in the two regimes being 438 and 58. No further thresholds are needed based on LR tests for the presence of further thresholds in each of the two regimes. The choice of $k = 3$ is justified by treating k as a parameter and estimating it by profile quasi log-likelihood; upon fitting the model with k ranging from 1 to 5 and identical effective sample size yields the profile likelihood -25.54, -29.32, -25.00, -28.01 and -26.29, respectively, and maximized at $k = 3$. Model diagnostics (Figure 9 in the *Supplementary Materials*) show that the model provides a good fit to the data.

Figure 5 shows the conditional variance processes from the fitted T-CHARM

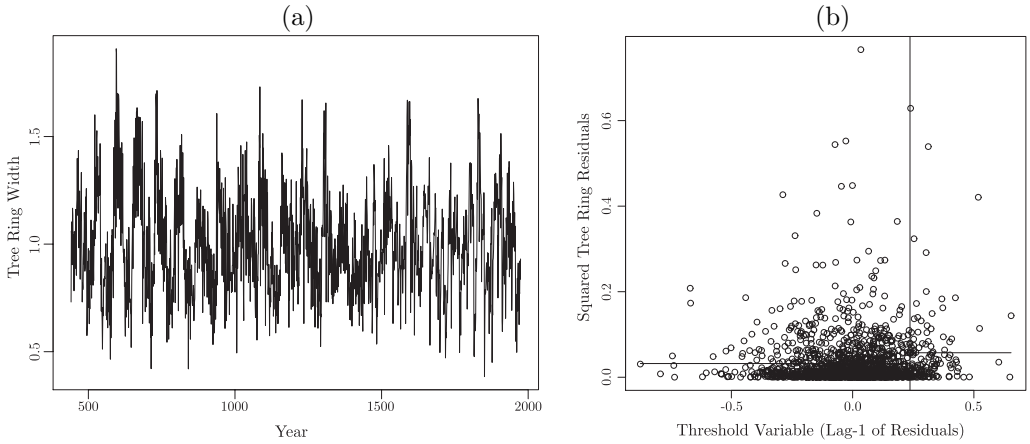

Figure 6. Left subfigure: Time plot of the annual tree ring width. Right subfigure: Scatter diagram of the squared tree ring residuals versus the lag 1 of the residuals, threshold variable. Vertical line separates the two regimes and the horizontal lines indicate the estimated variances of the two regimes.

and the GARCH(1, 1) model reported by Cryer and Chan (2008). They appear to complement each other on a global scale. On a finer scale, it appears that the T-CHARM captures some of the troughs during periods of high conditional variance whereas the GARCH(1, 1) model tends to smooth them. On the other hand, as three-parameter models, both models have log-likelihoods of similar order of magnitude, bearing in mind their approximate nature. See the *Supplementary Materials* for further discussion on the out-of-sample predictive performance of the fitted T-CHARM model versus the GARCH(1, 1) model.

The next two examples lie outside economics and finance. Each time series has a non-trivial conditional mean structure that can be modelled by an ARIMA model with possibly additive outliers, but the errors are conditionally heteroscedastic white noise that we model by some T-CHARM. The parameters in the mean function are generally distinct from those parametrizing the conditional variance function. Consequently, the mean structure may be estimated first, by making use of the white-noise nature of the errors and ignoring the conditional variance structure. Then the parameters of the conditional variance function can be subsequently estimated through a quasi-likelihood with the ARIMA residuals treated as if they were the true innovations. It can be readily checked that, under some mild regularity conditions, the mean parameters and the variance parameters are asymptotically independent of each other, so that the their standard errors can be obtained separately from each of the two steps.

The second example is a long time series of annual tree ring width (Figure 6(a)), with the measurements taken from a tree in a location at high altitude in

Argentina. The time series spans the years 441 to 1974, and it was contributed by J. Boninsegna to the NOAA Paleoclimatology database

`http://www.ncdc.noaa.gov/paleo/metadata/noaa-tree-2782.html`.

An IMA(1, 1) model was initially identified and fitted to the data with the MA coefficient given by -0.6110 with standard error 0.0216. The residuals of the fitted IMA(1, 1) model appeared to be white noise in the sense that the residual ACF was only marginally significant at lag 11 and five higher lags out of the 100 lags examined. This observation was corroborated by the Ljung-Box test based on the first k lags of the residual ACF with k ranging from 3 to 100. On the other hand, the absolute residuals appear to be correlated; see Figure 12 in the *Supplementary Materials*. We fit a two-regime T-CHARM to the residuals to account for the conditional heteroscedasticity, with lag 1 of the IMA(1, 1) error as the threshold variable. The parameter estimates were $\hat{\sigma}_1^2 = 0.03205(0.00153)$, $\hat{\sigma}_2^2 = 0.05729(0.00893)$, and $\hat{r} = 0.2366$ (95% confidence interval: (0.1619, 0.3049)), approximately the 91th percentile; see Figure 6(b).

The threshold structure was supported by the LR test for T-CHARM with p-value $p_0 = 0.005$. The other two methods of calculating the conditional p-values yielded $p_1 = 0.008$ and $p_2 = 0.002$. The first regime contained 1402 observations while the second 131 observations. No further thresholds were needed by reference to the LR test for the presence of further thresholds in each of the two regimes. The fitted T-CHARM successfully captured the conditional heteroscedasticity in the data as there were no residual ARCH effects in the standardized residuals from the fitted T-CHARM, by reference to the McLeod-Li test up to 100 lags. The fitted T-CHARM suggests that during fast-growing years, tree growth is much more variable, with a variance that almost doubles that during non-fast-growing years. What caused the observed variations is unclear, but it may be related to the fact that over non-fast-growing regime the tree ring width is bounded whereas this is not so in the fast-growing regime.

Next, we tried a GARCH model. The sample EACF of the absolute residuals of the IMA(1, 1) model tentatively suggested GARCH(1, 2) and GARCH(2, 2) models, but only the ARCH(1) coefficient was found to be significant, at 5% level. Eventually we chose the GARCH(1, 1) model, whose conditional variance is $h_t = \beta h_{t-1} + \alpha_0 + \alpha_1 X_{t-1}^2$ where X_t stands for the IMA(1, 1) errors, as this model passed the McLeod-Li test but the simpler ARCH(1) model did not. The GARCH estimates, with their standard errors enclosed in parentheses, were $\hat{\alpha}_0 = 0.0284(0.00656)$, $\hat{\alpha}_1 = 0.0987(0.0288)$, and $\hat{\beta} = 0.0747(0.198)$. While the T-CHARM shed some light on the tree-growing process, it is unclear to us as to how to interpret the fitted GARCH(1, 1) model. Finally, both fitted models involved three parameters with comparable quasi-likelihoods, again bearing in mind their approximate nature.

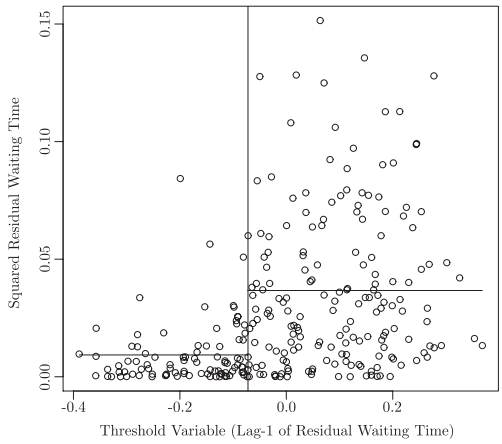

Figure 7. Squared AR(1) residual waiting time plotted against the lag-1 of the log waiting time; vertical line separates the two regimes and horizontal lines indicate the estimated variances of the two regimes.

The third example concerns the time series of waiting times between the starts of consecutive eruptions of the Old Faithful geyser. The data have been extensively studied in the literature. We used the version of the data collected from August 1–15, 1985; see Azzalini and Bowman (1990), Härdle (1991), the recent review by Zucchini and MacDonald (2009, Chap. 10), and the references therein. Figure 13 in the *Supplementary Materials* plots the scatter diagrams of the log waiting time against its lag k, for $k = 1, \ldots, 6$. The plot highlights the main features of the data: the waiting time is strongly associated with lag 1 values but much less so with values at higher lags; the presence of conditional heteroscedasticity; a number of outliers. The last two features seem to be largely ignored in the literature.

Preliminary statistical analysis suggested the possibility of an AR(1) model plus additive outliers for the mean structure, which we then fit to the data yielding the AR(1) coefficient estimate -0.571 (0.0497) and mean 4.248 (0.0065), and adjusted for five outliers at epochs 22, 37, 172, 237 and 266. The residuals appeared to be white, based on the Ljung-Box test up to lag 50. However, the residuals were highly conditionally heteroscedastic: if the previous waiting time is under-predicted, then the subsequent waiting time is subject to much larger uncertainty; see Figure 7. We then fit a two-regime T-CHARM with the lag 1 of the AR(1) errors as the threshold variable, giving the parameter estimates $\hat{\sigma}_1^2 = 0.0093(0.0011), \hat{\sigma}_2^2 = 0.037(0.0031)$ (almost four times larger than the variance of the lower regime), and $\hat{r} = -0.072$ (95% confidence interval: $(-0.088, -0.058)$), which is about the 34th percentile. The existence of the threshold was supported by the LR test with p-value $< 10^{-5}$, for all three methods of calculation, and

there is no need for more thresholds, based on the LR test. The standardized residuals from the T-CHARM were no longer conditionally heteroscedastic, based on the McLeod-Li test up to lag 50. However, the standardized residuals appear to be somewhat heavy tailed based on the quantile-quantile normal score plot (not shown). It seems that the GARCH formulation is unsuited for the AR(1) residuals; we tried to fit a GARCH(1, 1) model to the AR(1) errors as suggested by the sample EACF of the absolute residuals, but the quasi-likelihood estimation failed to converge.

8. Discussion

We have presented a viable alternative to the ARCH-type models. This provides a simple and versatile approach to conditional heteroscedasticity, based on an observable mixture of independent random variables. The model requires minimal conditions for statistical inference and often offers interpretable results. Computation is quite straightforward.

Asymptotic results can provide a basis for calculating the approximate confidence intervals of threshold parameters. Preliminary investigation of the two-regime case suggests that tabulation of the relevant quantiles, based on Monte Carlo, is precise if the variances of the two regimes are quite different. An interesting research problem is to augment this approach with the asymptotic framework for the case of nearly equal variances (Yao (1987)), with the ultimate goal of providing confidence statements for the threshold parameters; see also Hansen (1997, 2000, 2011).

See Chan et al. (2012) for a discussion of a spatial generalization of the proposed threshold approach.

Acknowledgements

We express our sincere thanks to the Co-editor and two anonymous referees for their constructive suggestions. The research was partially supported by the US National Science Foundation (DMS-1021292) and the US National Institute of Health (NIH R01 HL112986-01A1) (to Chan), Hong Kong RGC Grants numbered HKUST 603413, HKUST 641912 and FSGRF12SC12 (to Ling), the Distinguished Visiting Professorship at the University of Hong Kong, the Saw Swee Hock Professorship and the Institute of Mathematical Science, the last two at the National University of Singapore (to Tong).

References

Azzalini, A. and Bowman, A. W. (1990). A look at some data on the Old Faithful geyser. *Appl. Statist.* **39**, 357-365.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 47-78.

Brockwell, P. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer-Verlag, New York.

Cai, J. (1994). Markov model of unconditional variance in ARCH. *J. Bus. Econom. Statist.* **12**, 309-316.

Chan, K. S. (1990). Testing for threshold autoregression. *Ann. Statist.* **18**, 1886-1894.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21**, 520-533.

Chan, K. S. (2009). *Exploration of a Nonlinear World.* World Scientific, Singapore.

Chan, K. S., Li, D., Ling, S. and Tong, H. (2012). *On conditionally heteroscedastic AR models with thresholds.* Technical Report Serial Number 484, Department of Statistics and Actuarial Science, The University of Hong Kong.

Chan, K. S. and Li, W. K. (2007). Threshold models and new developments in time series. *Statist. Sinica* **17**, 1-287.

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes.* Chapman & Hall.

Cryer, J. D. and Chan, K. S. (2008). *Time Series Analysis: With Applications in R.* Springer-Verlag, New York.

Dirkse, J. P. (1975). An absorption probability for the Ornstein-Uhlenbeck process. *J. Appl. Probab.* **12**, 595-599.

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

Gourieroux, C. and Monfort, A. (1992). Qualitative threshold ARCH models. *J. Econometrics* **52**, 159-199.

Gray, S. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *J. Finan. Econom.* **42**, 27-62.

Haas, M., Mittnik, S. and Paolella, M. (2004). A new approach to Markov-switching GARCH models. *J. Finan. Econom.* **2**, 493-530.

Hamilton, J. and Susmel, R. (1994). Autoregressive conditional heteroscedasticity and changes in regime. *J. Econometrics* **64**, 307-333.

Hansen, B. E. (1997). Inference in TAR models. *Stud. Nonlinear Dyn. Econom.* **2**, 1-14.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* **68**, 575-603.

Hansen, B. E. (2011). Threshold autoregression in economics. *Stat. Interface* **4**, 123-128.

Härdle, W. (1991). *Smoothing Techniques: With Implementation in S.* Springer, New York.

Li, D. and Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *J. Econometrics* **167**, 240-253.

Li, D., Ling, S. and Tong, H. (2012). On moving-average models with feedback. *Bernoulli* **18**, 735-745.

Ling, S. and Tong, H. (2011). Score based goodness-of-fit tests for time series. *Statist. Sinica* **21**, 1807-1829.

Moran, P. A. P. (1953). The statistical analysis of the Canadian lynx cycle. *Aust. J. Zoo.* **1**, 163-173.

Nummelin, E. (1984). *General Irreducible Markov Chains and Non-Negative Operators.* Cambridge University Press, Cambridge.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

Rabemananjara, R. and Zakoïan, J. M. (1993). Threshold ARCH models and asymmetries in volatility. *J. Appl. Econometrics* **8**, 31-49.

Schwert, G. W. (1989) Business cycles, financial crises and Stock volatility. *Carnegie-Rochester Conference Series on Public Policy* **31**, 83-126.

Seijo, E. and Sen, B. (2011). A continuous mapping theorem for the smallest argmax functional. *Electron. J. Stat.* **5**, 421-439.

Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing* (Edited by C. H. Chen). Sijthoff and Noordhoff, Amsterdam.

Tong, H. (1982). Discontinuous decision processes and threshold autoregressive time series modelling. *Biometrika* **69**, 274-276.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press, Oxford.

Tong, H. (2011). Threshold models in time series analysis - 30 years on. *Stat. Interface* **4**, 107-118.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Soc. Ser. B* **42**, 245-292.

Tsay, R. S. and Tiao, G. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA Models. *J. Amer. Statist. Assoc.* **79**, 84-96.

Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41**, 425-436.

Yao, Y. C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Ann. Statist.* **15**, 1321-1328.

Zakoïan, J. M. (1994). Threshold heteroscedastic models. *J. Econom. Dynam. Control* **18**, 931-955.

Zucchini, W. and MacDonald, I. (2009). *Hidden Markov Models for Time Series: An Introduction Using R.* Chapman & Hall/CRC, London.

Department of Statistics & Actuarial Science, University of Iowa, IA, 52242, USA.

E-mail: kung-sik-chan@uiowa.edu

Mathematical Sciences Center, Tsinghua University, Beijing, China.

E-mail: li20072010@gmail.com

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong.

E-mail: maling@ust.hk

Department of Statistics, London School of Economics and Political Science, London WC2A 2AE, United Kingdom.

E-mail: howell.tong@gmail.com

(Received June 2012; accepted March 2013)