Maximum Likelihood Estimation of STAR and STAR-GARCH Models: A Monte Carlo Analysis*

Felix Chan and Michael McAleer

Department of Economics
University of Western Australia

Revised: November 2001

Abstract

Theoretical and practical interest in non-linear time series models, particularly regime switching models, has increased rapidly in recent years. Given the substantial research activity in analysing time-varying volatility through Generalised Autoregressive Conditional Heteroscedasticity (GARCH) processes, it would seem to be important to investigate regime switching models with GARCH errors. Two of the most popular specifications in this class are the Smooth Transition Autoregressive - GARCH (STAR-GARCH) and STAR - Smooth Transition GARCH (STAR-STGARCH) models. However, there is very little known regarding the theoretical or finite sample properties of the estimators of such models. This paper investigates the finite sample properties of maximum likelihood estimation (MLE) of STAR and STAR-GARCH models through numerical simulation. The sensitivity of MLE of STAR models to different variances in the unconditional shocks, the effects of fixing the threshold value and the transition rate of STAR models, and misspecification of the conditional mean and the transition function of STAR-GARCH models, are examined.

*The authors are grateful for helpful comments and suggestions from participants at the International Conference on Modelling and Forecasting Financial Volatility, Perth, Australia, September 2001. The first author wishes to acknowledge an Australian Postgraduate Award and an Individual Research Grant at the University of Western Australia. The second author wishes to acknowledge the financial support of the Australian Research Council.

1 Introduction

Theoretical and practical interest in non-linear time series models has increased rapidly in recent years. In particular, regime switching models have become popular in the class of non-linear models. Given the substantial research activity in analysing time-varying volatility through Generalised Autoregressive Conditional Heteroscedasticity (GARCH) processes (see Engle (1982) and Bollerslev (1986)), it is also of interest to investigate regime switching models with GARCH errors. Two of the most popular specifications in this class are the Smooth Transition Autoregressive - GARCH (STAR-GARCH) and STAR - Smooth Transition GARCH (STAR-STGARCH) models.

Although STAR-GARCH and STAR-STGARCH have recently been used in forecasting (Franses, Neele and van Dijk (1998), Lundbergh and Teräsvirta (1999, 2000)), the structural and statistical properties of these models have not yet been established. Furthermore, any diagnostic tests for these models *assume* consistency and asymptotic normality, but these statistical properties cannot be examined theoretically because the regularity conditions are as yet unknown. Consequently, inferences based on these assumptions may not be valid. Moreover, information criteria such as the Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC) may not be useful for gauging the adequacy of these models as the properties of the log-likelihood functions are also presently unknown.

The lack of knowledge of the statistical properties of these models can cause difficulties in selecting the most efficient optimization algorithm. As noted by Lundbergh and Teräsvirta (1999) and van Dijk, Teräsvirta and Franses (2000), the convergence of the Quasi-Maximum Likelihood Estimator (QMLE) is sensitive to the initial values. In fact, it is unclear as to whether different algorithms would produce the same estimates even if the initial values were sufficiently close to the optimum values.

The purpose of this paper is to investigate the finite sample properties of MLE relating to estimation and misspecification for STAR and STAR-GARCH models. A lack of statistical and structural properties has made the use of existing diagnostic tests problematic for this class of models. Moreover, it is intended that the numerical simulation results will provide motivation for further theoretical analysis relating to these models.

This paper presents the results of four experiments. Experiment 1 investigates the sensitivity of MLE of the STAR model to different variances in the unconditional shocks. It is shown that the sensitivity of the estimates to different variances depends on the transition function of the underlying process. The performance of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is also discussed.

Experiment 2 examines the robustness of the MLE by fixing certain parameters in STAR models, namely the threshold value and the transition rate. The results show that, in the case of Logistic STAR (LSTAR), the accuracy of the estimates improves when both the transition rate and the threshold value are fixed. However, fixing only one parameter does not necessarily improve estimation. In the case of Exponential STAR (ESTAR), fixing only one parameter is often sufficient to improve the accuracy of the remaining estimates. This raises an issue regarding the robustness of MLE for LSTAR, as well as the performance of some optimisation algorithms.

Experiment 3 investigates the effects of MLE for AR-GARCH when the data follow a STAR-GARCH process. For the LSTAR-GARCH and ESTAR-GARCH models, misspecification in the conditional mean does not cause substantial bias for the MLE of the GARCH components in finite samples. The magnitude of the bias, however, depends on the transition function of the underlying process.

Finally Experiment 4 analyses the effects of misspecifying the transition function, specifically, the effects on the MLE of the LSTAR-GARCH model when the data follow an ESTAR-GARCH process, and vice-versa. The results show that the misspecification of the transition function often leads to parameter estimates of the lagged dependent variable being greater than one in absolute value for at least one regime. However, misspecifying the transition function does not seem to affect the estimates for the GARCH components.

The plan of the paper is as follows. Section 2 presents some recent theoretical developments for the GARCH, STAR, STAR-GARCH and STAR-STGARCH models. Section 3 provides the results of the simulation experiments. An empirical illustration with STAR-GARCH to analyse Standard and Poor's 500 (S&P) Index returns is reported in Section 4. Concluding remarks are given in Section 5.

2 Volatility Models

This section discusses some recent theoretical developments in modelling GARCH, STAR, STAR-GARCH and STAR-STGARCH. Model definitions, regularity conditions, and statistical properties, where available, will be discussed, with an emphasis on the importance of deriving the structural and statistical properties of the models.

2.1 ARCH/GARCH

Consider the ARMA(r, s) model:

$$y_t = \sum_{i=1}^{r} \phi_i y_{t-i} + \sum_{i=1}^{s} \theta_i \varepsilon_{t-i} + \varepsilon_t, \tag{2.1}$$

in which ε_t follows an Autoregressive Conditional Heteroscedasticity, ARCH(p), process if

$$\varepsilon_t = \eta_t \sqrt{h_t} \tag{2.2}$$

where

$$\eta_t \sim i.i.d.(0,1) \tag{2.3}$$

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \tag{2.4}$$

and $\omega > 0$, $\alpha_i \geq 0$ for all $i = 1,..,p$, are sufficient conditions for $h_t > 0$ for all $t = 1,..,T$. This model was proposed by Engle (1982) to relax the traditional assumption of a constant one-period forecast variance. Engle showed that ARCH(p) has a constant unconditional variance but a time-varying conditional variance.

Engle (1982) showed that ε_t is second-order stationary (that is, possesses finite second moment) if and only if all the roots of the characteristic polynomial

$$(1 - \sum_{i=1}^{p} \alpha_i z^i) = 0$$

lie outside the unit circle. It was assumed that the process ε_t starts infinitely far in the past, with finite $2m$th moment, an assumption which is not possible to check in practice. Engle also derived the regularity condition for the existence of the moments for ARCH(1), specifically, the $2m$th moment exists if and only if

$$\alpha_1^m \prod_{j=1}^{m} (2j-1) < 1.$$

Milhøj (1985) avoided Engle's assumption and showed that ε_t is second-order stationary if and only if

$$\sum_{i=1}^{p} \alpha_i < 1, \tag{2.5}$$

and also derived the regularity condition for the existence of moments without the restrictive assumption. This result is identical to Engle's in the case of ARCH(1) with normal η_t, but cannot be given an explicit form in the case of ARCH(p) and $m > 2$.

It is interesting to note that equation (2.5) is not a necessary condition for the strict stationarity of the ARCH(p) model. The necessary and sufficient condition for the strict stationarity of ARCH(p) was derived by Bougerol and Picard (1992).

Engle (1982) suggested two possible methods for estimating the parameters in equations (2.1) and (2.4) namely the Least Squares Estimator (LSE) and the Maximum Likelihood Estimator (MLE). The LSE is given as

$$\hat{\delta} = \Big(\sum_{t=2}^{T} \tilde{\varepsilon}_{t-1} \tilde{\varepsilon'}_{t-1} \Big)^{-1} \sum_{t=2}^{T} \tilde{\varepsilon}_{t-1} \varepsilon_t \tag{2.6}$$

where $\hat{\delta} = (\omega, \alpha_1, ..., \alpha_p)'$ and $\tilde{\varepsilon}_t = (1, \varepsilon_t^2, ..., \varepsilon_{t-p+1}^2)'$. Weiss (1986) and Pantula (1989) showed that $\hat{\delta}$ is consistent and asymptotic normal if

$$E(\varepsilon_t^8) < \infty,$$

which is a rather strong condition.

The conditional log-likelihood function of (2.1) given observations ε_t, $t = 1, .., T$, can be written as:

$$l(\delta) = \frac{-1}{2T} \sum_{t=1}^{T} \left(\ln h_t + \frac{\varepsilon_t^2}{h_t} \right), \tag{2.7}$$

and the MLE is given as

$$\hat{\delta} = \text{argmax}_{\delta \in \Theta} l(\delta),$$

assuming that $\delta \in \Theta$, a compact subset of $\mathbb{R}^{p+1}$. Engle (1982) showed that the information matrix of this function is block-diagonal, so that the parameters in the conditional mean and the conditional variance can be estimated separately without loss of asymptotic efficiency. The residuals from the estimated conditional mean equation can be used to estimate the conditional variances.

Moreover, the MLE is referred to as the Quasi-MLE (QMLE) when η_t is not normal. Weiss (1986) and Pantula (1989) showed that the QMLE is consistent and asymptotic normal if

$$E(\varepsilon_t^4) < \infty.$$

This result was extended by Ling and McAleer (1999b), who showed that the QMLE is consistent and asymptotic normal if

$$E(\varepsilon_t^2) < \infty.$$

The Berndt, Hall, Hall, Hausman (1974) algorithm (BHHH) is often used to determine $\hat{\delta}$ but, as suggested by Mak, Wong and Li (1997), this algorithm has convergence problems if the initial values are not sufficiently close to the final solutions. In such cases, a Newton-Raphson procedure is recommended.

Bollerslev (1986) extended ARCH(p) by including the lagged values of the conditional variance to yield the GARCH(p, q) model, namely

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i}, \tag{2.8}$$

where $\omega > 0$, $\alpha_i \geq 0$ for all $i = 1, .., p$, $\beta_i \geq 0$ for all $i = 1, .., q$, are sufficient conditions for $h_t > 0$ for all $t = 1, .., T$. If $\beta_i = 0$ for all i, then GARCH(p, q) reduces to ARCH(p). All the structural and statistical properties of GARCH hold for ARCH, in general, except for one case which is discussed below.

The necessary and sufficient condition for the second-order stationarity of (2.8) was established by Bollerslev (1986) as

$$\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i < 1.$$

Nelson (1990) derived the necessary and sufficient condition for the strict stationarity and ergodicity of GARCH(1,1) as

$$E(\ln(\alpha_1 \eta_t^2 + \beta_1)) < 0.$$

This log-moment condition allows $\alpha_1 + \beta_1$ to be slightly larger than 1, in which case the variance is not finite (i.e. $E(\varepsilon_t^2) = \infty$). Note that this condition holds for GARCH(1,1) but not for ARCH(1), that is, $E(\ln(\alpha_1 \eta_t^2)) < 0$ does not ensure strict stationarity and ergodicity for ARCH(1) because the condition is derived under the assumption that $\beta_1 \neq 0$. Moreover, this condition is not easy to apply in practice as it is the mean of an unknown random variable with unknown parameters.

For GARCH(p,q), the necessary and sufficient condition for strict stationarity and ergodicity was established by Bougeral and Picard (1992) and Nelson (1990). The necessary and sufficient condition for the existence of the $2m$th moment of the GARCH(1,1) model was provided by Bollerslev (1986), who also provided the necessary and sufficient condition for the fourth-order moments of the GARCH(1,2) and GARCH(2,1) models. He and Teräsvirta (1999a) obtained the moment conditions of a family of GARCH(1,1) models using a similar method as in Bollerslev (1986). Ling and McAleer (1999c) derived the sufficient condition for the existence of the stationary solution of this family of GARCH(1,1) models, showed that He and Terasvirta's (1999a) condition was necessary but not sufficient, and provided the sufficient condition. He and Teräsvirta (1999b) also examined the fourth moment structure of the general GARCH(p,q) process. In the case of GARCH(1,1), the fourth moment condition under normality is given by

$$(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 < 1. \tag{2.9}$$

Ling (1999) obtained a sufficient condition for the existence of the $2m$th moment for the GARCH(p,q) model, based on Theorem 2.1 in Ling and Li (1997) and Theorem 2 in Tweedie (1988). The sufficient condition is given as

$$\rho[E(A_t^{\otimes m})] < 1,$$

where $\rho(A) = \max\{\text{eigenvalues of a matrix A}\}$, and A_t is given by:

$$A_t = \left(\begin{array}{ccc|ccc} \alpha_1 \eta_t & \cdots & \alpha_p \eta_t & \beta_1 \eta_t & \cdots & \beta_q \eta_t \\ & I_{(p-1)\times(p-1)} & O_{(p-1)\times 1} & & O_{(p-1)\times q} & \\ \hline \alpha_1 & \cdots & \alpha_p & \beta_1 & \cdots & \beta_q \\ & O_{(q-1)\times p} & & & I_{(q-1)\times(q-1)} & O_{(q-1)\times 1} \end{array} \right).$$

Unlike Bollerslev (1986) and He and Teräsvirta (1999a,b), Ling's method does not assume that the GARCH(p, q) process starts infinitely far in the past with finite $2m$th moment, and has a far simpler form than Milhøj's (1985) result. This condition is also necessary for the existence of the $2m$th moment, as demonstrated by Ling and McAleer (1999a). Thus, the moment structure of a general GARCH(p, q) has been fully established. As an extension of the GARCH(p, q) process, Ling and McAleer (1999a) derived the necessary and sufficient moment conditions of the asymmetric power GARCH(p, q) model of Ding et al. (1993).

The parameters of the GARCH(p, q) model are often estimated by MLE, or by QMLE when the normality of η_t is not assumed, with the log-likelihood function of GARCH(p,q) being identical to that of (2.7), except that h_t in this case follows (2.8) instead of (2.4). For GARCH(1,1), Lee and Hansen (1994) and Lumsdaine (1996) showed that the QMLE is consistent and asymptotic normal if

$$E\big(\ln(\alpha_1\eta_t^2 + \beta_1)\big) < \infty, \qquad \beta_1 \neq 0.$$

Ling and Li (1997) showed that the *local* QMLE for GARCH(p, q) is consistent and asymptotic normal if

$$E(\varepsilon_t^4) < \infty.$$

For the *global* QMLE, Ling and McAleer (1999a) showed that

$$E(\varepsilon_t^2) < \infty$$

is sufficient for consistency, and

$$E(\varepsilon_t^6) < \infty$$

is sufficient for asymptotic normality.

The QMLE is often more efficient than LSE for ARMA-GARCH(p, q) models. This result was first observed by Engle (1982) through a simple fixed design regression model with an ARCH(1) process. Pantula (1989) also showed that the MLE is more efficient than LSE for an AR model with ARCH(1) errors. The QMLE is efficient only if η_t is normal. When η_t is not normal, adaptive estimation is useful to obtain efficient estimators. Some useful references include Bickel (1982), Robinson (1988), Stoker (1991), and Ling and McAleer (1999d). As the adaptive estimation method is not yet available in most econometric software packages due to its computational complexity, this explains in part the popularity of MLE.

It is important to note that the choice of lag length in the conditional variance equation has not been thoroughly investigated in the literature. Engle (1982) proposed an LM test for ARCH effects, and used the test to decide the appropriate lag length. Bollerslev (1986) used a similar test to decide the lag length of GARCH in an empirical example, but admitted that his choice was arbitrary. Some researchers choose the lag length for their models based on model adequacy, using criteria such as AIC and SBC, while others choose their models based on in-sample forecast performance.

A distinct characteristic of GARCH-type models is their ability to capture volatility clustering. If the shock from the previous period is high (low), the large (small) value of ε_{t-1}^2 will then influence h_t. The GARCH model can also be fitted to leptokurtic financial data, and can be adapted for conditional Student t-distributed (GARCH-t) errors. GARCH also offers computational advantages over its extended versions. For example, the log-likelihood function of GARCH is relatively simple. However, there are several deficiencies in the linear GARCH model, as observed by Nelson (1991). First, it is an empirical regularity that the impact of a large negative shock is greater than a large positive shock, but a small positive shock has a larger impact than a small negative shock. This type of asymmetric behaviour cannot be captured by the symmetric GARCH model, as the conditional variance is a function only of past *squared* errors and past conditional variances. The most popular asymmetric GARCH models in the literature today are Glosten, Jagannathan and Runkle's (1992) GJR-GARCH and Nelson's (1991) Exponential GARCH (EGARCH). It is important to note that the structural and statistical properties of these two asymmetric GARCH models have not yet been fully developed, particularly for EGARCH.

Moreover, it is important to impose a restriction on all of the parameters in GARCH models to ensure that the conditional variances are positive. These restrictions can create difficulties in estimating even the simple GARCH model, especially when the data exhibit extreme observations and outliers.

2.2 STAR, STAR-GARCH and STAR-STGARCH

Non-linear time series models have become very popular in recent years. As regime switching models are particularly popular in the class of non-linear models, it is of interest to investigate regime switching models with GARCH errors. Regime switching models will be discussed here, with an emphasis on Smooth Transition Autoregressive (STAR) models.

Tong (1978) and Tong and Lim (1980) proposed the Threshold Autoregressive (TAR) model. The TAR model assumes that the regimes switch from one to another, as determined by the *threshold variables*, s_t, relative to the *threshold value*, c. Consider the two regime case:

$$y_t = (\phi_{11} + \sum_{i=2}^{r} \phi_{1i} y_{t-i+1})(1 - I(s_t - c)) + (\phi_{21} + \sum_{i=2}^{r} \phi_{2i} y_{t-i+1}) I(s_t - c) + \varepsilon_t, \tag{2.10}$$

where

$$I(s_t - c) = \begin{cases} 0, & s_t < c \\ 1, & s_t \geq c. \end{cases}$$

Model (2.10) can also be written as:

$$y_t = \begin{cases} \phi_{11} + \sum_{i=2}^{r} \phi_{1i} y_{t-i+1} + \varepsilon_t, & s_t < c \\ \phi_{21} + \sum_{i=2}^{r} \phi_{2i} y_{t-i+1} + \varepsilon_t, & s_t \geq c. \end{cases}$$

The threshold variable, s_t, is usually (but not always) defined as a linear combination of the lagged values of y_t, that is,

$$s_t = \sum_{i=1}^{k} \pi_i y_{t-i},$$

which is often referred to as a Self Exciting TAR (SETAR). van Dijk, Teräsvirta and Franses (2000) relaxed this definition of threshold variables to include non-linear combinations of the lagged values of y_t and of other exogenous variables.

Equation (2.10) is similar to a standard model of structural change, apart from the definition of threshold variable and threshold value, which assumes that the regimes switch from one to another instantly. To allow for a smooth transition, Terašvirta (1994) proposed the Smooth Transition Autoregressive (STAR) model:

$$y_t = (\phi_{11} + \sum_{i=2}^{r} \phi_{1i} y_{t-i+1})(1 - G(s_t; \gamma, c)) + (\phi_{21} + \sum_{i=2}^{r} \phi_{2i} y_{t-i+1}) G(s_t; \gamma, c) + \varepsilon_t, \tag{2.11}$$

in which $G(s_t; \gamma, c)$ is the transition function, assumed to be at least twice differentiable, ranging from 0 to 1, and γ is the transition rate.

Although two regimes will suffice for many empirical cases, it is straightforward to extend (2.10) to more than two regimes. Denoting $\phi_i = (\phi_{i1}, ...\phi_{ir})'$ and $x_t = (y_{t-1}, ..., y_{t-r})'$, equation (2.10) can be rewritten as

$$y_t = (\phi_1' x_t)(1 - I(s_t - c)) + (\phi_2' x_t) I(s_t - c) + \varepsilon_t.$$

An m-regime (Multiple Regime) model can be written as

$$y_t = \sum_{i=1}^{m} \phi_i' x_t (I(s_t - c_{i-1}) - I(s_t - c_i)) + \varepsilon_t, \tag{2.12}$$

where $c_1 < c_2 < .... < c_m$, and

$$I(s_t - c_i) = \begin{cases} 0, & s_t < c_i \quad \text{or} \quad i = m \\ 1, & s_t \geq c_i \quad \text{or} \quad i = 0. \end{cases}$$

For any $s_t \in [c_{i-1}, c_i)$, $y_t = \phi_i' x_t + \varepsilon_t$ for all $i = 1, ..., m$. Therefore, the regime is determined by the threshold variable, s_t, relative to the threshold value, c_i. In order to incorporate the idea of smooth transition in equation (2.10), replace the function $I(s_t - c_i)$ in (2.12) with the transition function $G(s_t; \gamma, c_i)$ for all i, yielding

$$y_t = \sum_{i=1}^{m} \phi_i' x_t (G_{i-1}(s_t; \gamma_{i-1}, c_{i-1}) - G_i(s_t; \gamma_i, c_i)) + \varepsilon_t, \tag{2.13}$$

where $G_i(s_t; \gamma_i, c_i)$ is assumed to be at least twice differentiable, ranging from 0 to 1, $G_0 = 1$ and $G_m = 0$. Equation (2.13) is known as the Multiple Regime Smooth Transition Autoregressive (MRSTAR) model.

An extension of the basic model permits the parameter vector ϕ_i to change over time, which is known as the Time Varying STAR (TV-STAR) model (van Dijk, Lundbergh and Teräsvirta (2000)).

Several transition functions are available, with the most popular being the first-order logistic function:

$$G(s_t; \gamma, c) = \frac{1}{1 + \exp(-\gamma(s_t - c))},$$

with the following properties:

$$\begin{aligned}
\lim_{s_t \to -\infty} G(s_t; \gamma, c) &\to 0, \\
\lim_{s_t \to \infty} G(s_t; \gamma, c) &\to 1, \\
G(s_t; 0, c) &= \frac{1}{2} \\
\lim_{\gamma \to -\infty} G(s_t; \gamma, c) &\to 0, \\
\lim_{\gamma \to \infty} G(s_t; \gamma, c) &\to 1.
\end{aligned}$$

A STAR model with a logistic transition function is the Logistic STAR (LSTAR). Although the logistic function is used frequently, other choices include the Exponential STAR (ESTAR) model given by:

$$G(s_t; \gamma, c) = 1 - \exp(-\gamma(s_t - c)^2), \qquad \gamma > 0,$$

and the n^{th}-order LSTAR:

$$G(s_t; \gamma, c) = (1 + \exp(-\gamma \prod_{i=1}^{n} (s_t - c_i)))^{-1}, \qquad \gamma > 0, \quad c_i < c_j \quad \forall \quad i < j.$$

In order to use this model effectively, it is important to choose an appropriate transition function and threshold variable. There exist many LM-type tests to determine the appropriate choice of $G(s_t; \gamma, c)$ and s_t (a comprehensive survey of the modelling strategy in the STAR framework is given in van Dijk, Teräsvirta and Franses (2000)).

Generally, the modelling cycle starts with a test of parameter constancy, such as testing whether STAR is more appropriate than a simple linear AR model. Assuming that LSTAR with two regimes is the preferred model, a test of parameter constancy is given by

$$H_{A0} : \phi_{11} = \phi_{21}, \quad \phi_{12} = \phi_{22}.$$

Parameters within the transition function, γ and c, are not involved in the null hypothesis, yielding unidentified nuisance parameters. Consider the null hypothesis of linearity as a test of

$$H_{B0} : \gamma = 0,$$

in which

$$G(s_t; 0, c) = \frac{1}{2}.$$

so that the STAR model can be written as

$$y_t = \frac{\phi_{11} + \phi_{21}}{2} + \frac{\phi_{12} + \phi_{22}}{2} y_{t-1} + \varepsilon_t,$$

which is linear, regardless of the truth of H_{A0}. Thus, it is important to include parameters in the transition function for purposes of testing. This problem can be avoided by expressing the transition function by its Taylor expansion around $\gamma = 0$, which is a simple but important technique for hypothesis testing in STAR-type models.

When the transition function and the threshold variable have been determined, the parameters in STAR can be estimated by Non-linear Least Squares (NLS). If

$$y_t = F(x_t; \phi) + \varepsilon_t,$$

the NLS estimator is given by

$$\hat{\phi} = argmin_\phi \sum_{t=1}^{T} (y_t - F(x_t; \phi))^2 = argmin_\phi \sum_{t=1}^{T} \varepsilon_t^2.$$

If ε_t is normal, NLS is equivalent to MLE, otherwise NLS can be interpreted as QMLE. Wooldridge (1994) and Pötscher and Prucha (1997) demonstrated that NLS is consistent and asymptotic normal under certain regularity conditions.

STAR models, especially LSTAR models, have been successfully applied in a number of areas. Teräsvirta and Anderson (1992) and Teräsvirta, Tjøstheim and Granger (1994) characterised the different dynamics of industrial production indexes for various OECD countries during expansions and recessions using LSTAR models. Moreover, Lundbergh and Teräsvirta (2000) examined the forecast performances of the LSTAR model for unemployment rates in Denmark and Australia, arguing that many unemployment rates exhibit asymmetries in that the rate of increase is often higher than the rate of decrease. Their results showed that the STAR model is superior to its AR counterpart.

A STAR-GARCH model allows ε_t in equation (2.13) to follow a GARCH process, as defined in (2.8). This extension has not yet been investigated thoroughly. Lundbergh and Teräsvirta (1999) give a comprehensive exposition of this model, but do not provide any statistical properties or regularity conditions for stationarity on the existence of its moments. These important properties have not yet been established. However, as the information matrix of STAR-GARCH is block diagonal, the parameters in the conditional mean and conditional variance equations can be estimated separately, as in the case of ARMA-GARCH. Therefore, the general GARCH properties discussed previously are also expected to hold for this model.

A further extension of the STAR-GARCH model is to incorporate the concept of regime switching in the GARCH components, resulting in the STAR-Smooth Transition GARCH (STAR-STGARCH) model. Let $\theta_i = (\theta_{i0}, ..., \theta_{i(p+q)})'$, $\Gamma_t = (1, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, ...\varepsilon_{t-p}^2, h_{t-1}, ..., h_{t-q})'$, and H_i be at least twice differentiable for all $i > 0$, with $H_0 = 1$ and $H_m = 0$. Denote a new threshold variable as

$$r_t = \sum_{i=1}^{k} \zeta_i \varepsilon_t,$$

with threshold values $d_i \in \mathbb{R}$ for all i. Then the STAR-STGARCH model is the same as equation (2.13), with

$$\varepsilon_t = \eta_t \sqrt{h_t},$$

where

$$h_t = \sum_{i=1}^{m} (\theta_i' \Gamma_t)(H_{i-1}(r_t; \xi_{i-1}, d_{i-1}) - H_i(r_t; \xi_i, d_i)). \tag{2.14}$$

The choice of H_i is similar to that of G_i, but is not restricted to equal G_i in the general case.

STAR-STGARCH is novel and has a number of distinct characteristics. First, it is nonlinear, both in the conditional mean and in the conditional variance. The GARCH component is useful for capturing volatility clustering, while the threshold variables and threshold values are useful if the data exhibit regime switching behaviour for varying y_t and ε_t. STAR-STGARCH also exhibits asymmetries, as it can be represented by setting the threshold value to 0. Consider a simple two-regime case, with $r_t = \varepsilon_t$, $d = 0$ and

$$H(\varepsilon_t; \xi, 0) = \frac{1}{1 + \exp(-\xi(\varepsilon_t - 0))},$$

so that equation (2.14) can be rewritten as

$$h_t = (\theta_1' \Gamma_t)(1 - H(\varepsilon_t; \xi, 0)) + (\theta_2' \Gamma_t) H(\varepsilon_t; \xi, 0).$$

Thus, in the extreme cases where $\varepsilon_t \to -\infty$ and $\varepsilon_t \to \infty$,

$$h_t = (\theta_1' \Gamma_t),$$
$$h_t = (\theta_2' \Gamma_t),$$

respectively. Therefore, the first regime is associated with $\varepsilon_t < 0$ and the second regime with $\varepsilon_t \geq 0$.

Although this model is potentially useful for empirical data that exhibit non-linearity and threshold behaviour, there are as yet no structural or statistical results of the MLE. Furthermore, as asymmetric behaviour is permitted, the information matrix for this model is no longer block diagonal and the two-stage estimation method is no longer valid. It is also important to note that, as observed in van Dijk, Teräsvirta and Franses (2000), the MLE for both STAR-GARCH and STAR-STGARCH is extremely sensitive to initial values. The choice of algorithm in approximating the optimal solution is also crucial in terms of convergence. These problems may be resolved by investigating the structural and statistical properties of the model. This paper focuses on STAR and STAR-GARCH models, with STAR-STGARCH model given completeness.

There are several LM-type specification tests to analyse the most appropriate model. Lundbergh and Teräsvirta (1999) and van Dijk, Teräsvirta and Franses (2000) provide a list of such tests. However, these tests are based on the assumption that the model is stationary and ergodic, an assumption which cannot be checked, in general, as no regularity conditions are available. Therefore, any empirical examples using STAR-GARCH and STAR-STGARCH remain questionable in terms of their reliability and stability.

Evaluating forecast performance is also problematic. As noted by van Dijk, Teräsvirta and Franses (2000), even though non-linear time series models often capture certain characteristics of the data better than their linear counterparts, the forecast performance of the former is not always superior, and is sometimes worse. Clements and Hendry (1998) and Diebold and Nason (1990) discuss various reasons for this phenomenon.

3 Simulation Results

The remainder of the paper investigates four different issues regarding the finite sample properties of MLE for STAR and STAR-GARCH models using numerical simulations. Unless otherwise stated, all STAR and STAR-GARCH models considered have the following stationary specifications, with the true parameter values given in Table 1 (see Franses and van Dijk (2000, pp.73-74)):

$$y_t = (\phi_{11} + \phi_{12} y_{t-1})(1 - F(y_{t-1}; \gamma, c)) + (\phi_{21} + \phi_{22} y_{t-1}) F(y_{t-1}; \gamma, c) + \varepsilon_t, \tag{3.1}$$

where $\varepsilon_t \sim NID(0, \sigma^2)$. In the case of STAR-GARCH, the specification is given by:

$$\varepsilon_t = \eta_t \sqrt{h_t}, \tag{3.2}$$

$$\eta_t \sim NID(0, 1), \tag{3.3}$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{3.4}$$

Parameters	ϕ_{11}	ϕ_{12}	ϕ_{21}	ϕ_{22}	γ	c	ω	α	β
Values	-0.3	-0.5	0.1	0.5	1	0	0.01	0.2	0.75

Table 1: Parameter Values

for LSTAR:

$$F(y_{t-1}; 1, 0) = \frac{1}{1 + exp(-y_{t-1})},$$

and for ESTAR:

$$F(y_{t-1}; 1, 0) = \frac{1}{1 - exp(-(y_{t-1})^2)}.$$

Furthermore, each experiment is based on 1000 replications, each with 3000 data points. The programs used in this paper have been written in Ox version 3 (this version by J.A. Doornik and M. Ooms can be obtained free of charge from http://hicks.nuff.ox.ac.uk/Users/Doornik/). The primary reason for using Ox instead of other existing programs written in GAUSS, such as those written by S. Lundbergh (http://ideas.uqam.ca/ideas/data/Softwares/bocbocodeG111201.html), is the gain in computational time (see http://www.scientificweb.de/ncrunch/ncrunch.pdf).

3.1 Experiment 1: Different STAR Variances

The purpose of this experiment is to examine the sensitivity of MLE for STAR models to different variances in the unconditional shocks. This is of interest because it is often difficult to obtain robust estimates of non-linear models. Specifically, estimates for non-linear models are often highly sensitive to the data and to the model specification. Such sensitivity is often related to the distinct structures of the log-likelihood functions, which affects the performance of conventional optimisation algorithms, such as Newton or Quasi-Newton (for example, BFGS) - type methods. The performance of different algorithms in optimising the likelihood functions for STAR-type models have not previously been investigated thoroughly.

Intuitively, if MLE is robust, then the means of the estimates should be relatively stable, regardless of the variance of the unconditional shocks. In order to test this hypothesis, data generated by LSTAR and ESTAR processes are simulated with 5 different variances, namely 0.1^2, 0.25^2, 0.5^2, 1 and 1.25^2. In each case, 1000 replications were generated and estimated using MLE. The data for the first set of 1000 replications were based on a variance of 0.1^2, the second set of 1000 replications on a variance of 0.25^2, and so on, through to the fifth set of 1000 replications based on a variance of 1.25^2.

As

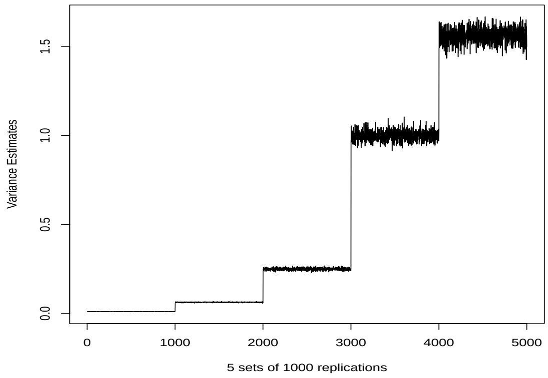

Figure 1: $\hat{\sigma^2}$ estimates for LSTAR

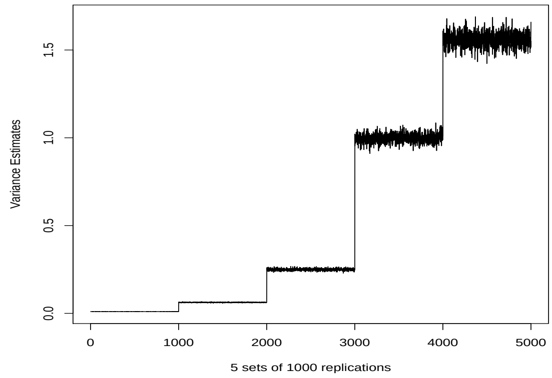

Figure 2: $\hat{\sigma^2}$ estimates for ESTAR

shown in Figures 1 and 2, the variances were estimated quite accurately in all cases. However, the standard deviations of the estimates increased with the true variance. The standard deviations of the $\hat{\sigma^2}$ estimates are given in Table 2, and show similar features for both LSTAR and ESTAR.

σ^2	LSTAR	ESTAR
0.1^2	0.000269	0.000254
0.25^2	0.00156	0.00162
0.5^2	0.00649	0.00646
1	0.0269	0.0253
1.25^2	0.0399	0.0404

Table 2: Standard Deviations of $\hat{\sigma^2}$ for different variances

Moreover, the variability of the estimates for the threshold value, $\hat{c}$, increases as the variance of the shocks increases, as shown in Figure 3. However, the same does not hold for the other estimates. Interestingly, the standard deviations of the parameter estimates for $\sigma = 0.25$, 0.5, 1 and 1.25 are much greater than for $\sigma = 0.1$, but do not increase monotonically.

One possible explanation is that the performance of the BFGS algorithm was poor for $\sigma > 0.1$, thereby resulting in meaningless estimates which distorted the standard deviations. This interpretation is supported by the fact that there appear to be two distinct groups for each parameter estimate. The first group, which consists of 98% to 99% of the replications, has means which are close to the true values. However, the second group consists of parameter estimates which are much greater in absolute value than those in the first group. Furthermore, the estimates in the second group are often clustered in only one direction (either positively or negatively), which also affects the means of the parameter estimates. These cases may reflect the lack of robustness of the BFGS algorithm.

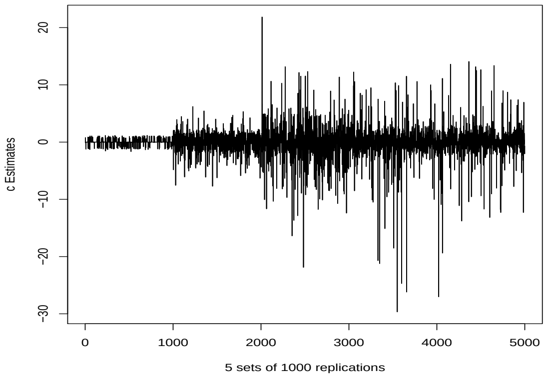

Figure 3: Estimates of $\hat{c}$ for LSTAR and different variances

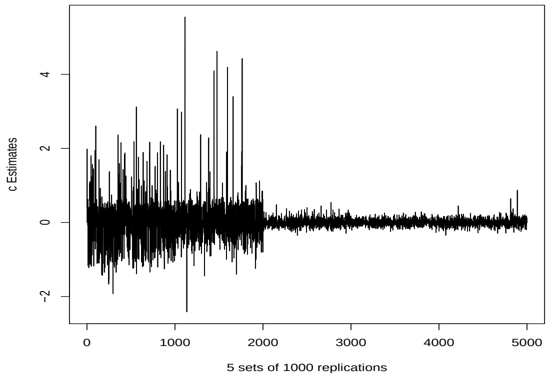

Figure 4: Estimates of $\hat{c}$ for ESTAR and different variances

The results are quite surprising for the exponential case. As shown in Figure 4, the variability of the threshold estimates decreases as the variance of the shocks increases! Furthermore, this result holds for $\hat{\phi_{11}}$, $\hat{\phi_{12}}$, $\hat{\phi_{21}}$, $\hat{\phi_{22}}$, but not for $\hat{\gamma}$. Although the variability of the estimates changes as the variance changes, the means of the estimates remain close to the true values. Overall, the estimates of $\hat{\gamma}$ seem to be insensitive to changes in the variance.

3.2 Experiment 2: Fixing the Threshold Value and Transition Rate

Given the previous findings, it is worth investigating whether the accuracy of the estimates in the two regimes can be improved by fixing some parameters, namely the threshold value, c, and the transition rate, γ. This is the purpose of the second experiment. In this case, both the LSTAR and ESTAR processes, as defined in equation (3.1) with $\sigma^2 = 0.25^2$ (see Franses and van Dijk (2000, pp.73-74)), were simulated and estimated under 4 different scenarios, as follows:

1. Estimating all the parameters;

2. Fixing the transition rate, γ, at the true value;

3. Fixing the threshold value, c, at the true value;

4. Fixing both γ and c, at their true values.

Not surprisingly, fixing both the transition rate and the threshold value at their true values improves the accuracy of the parameter estimates in both regimes. However, fixing only the transition rate or the threshold value can worsen the estimates for the logistic case, (see, for example, Figures 5 to 8). The $\hat{\phi_{12}}$ estimates can exceed 5 when either γ or c is fixed, even though $\phi_{12} = -0.5$. However, the $\hat{\phi_{12}}$ estimates seem to be within a reasonable range when neither parameter is fixed, or when both γ and c are fixed.

Interestingly, fixing only the transition rate or the threshold variable improves the accuracy of the other estimates for the exponential case. These results may have the following implications:

1. As LSTAR is a more complicated model than ESTAR, it is more difficult for the algorithm to converge to the true values.

2. BFGS is not a robust algorithm for estimating the LSTAR model.

Point 2 is closely related to point 1, as BFGS is a Quasi-Newton algorithm, which uses numerical first and second derivatives. Therefore, if the derivatives are not approximated accurately, it is likely that it will produce incorrect estimates. A sensible strategy for investigating this issue is to compare the estimates provided by different algorithms. This is an issue which does not seem to have been considered in the literature.

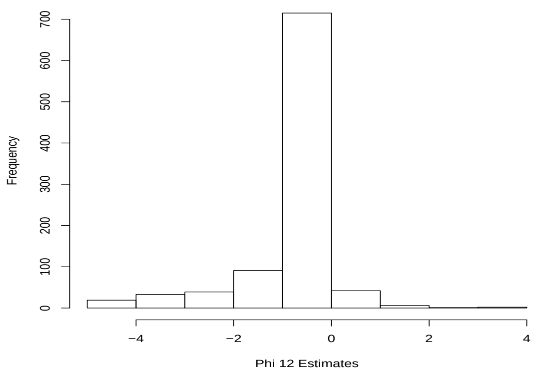

Figure 5: Estimates of $\hat{\phi_{12}}$ for LSTAR with no fixed parameters

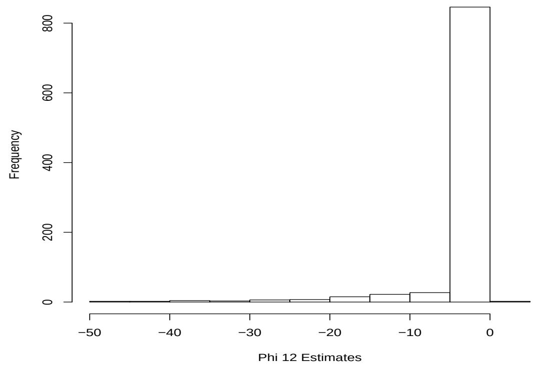

Figure 6: Estimates of $\hat{\phi_{12}}$ for LSTAR with fixed γ

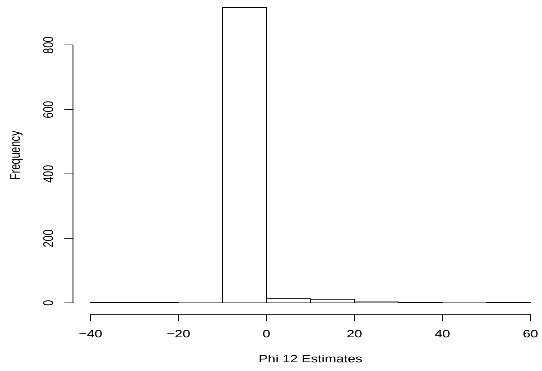

Figure 7: Estimates of $\hat{\phi_{12}}$ for LSTAR with fixed c

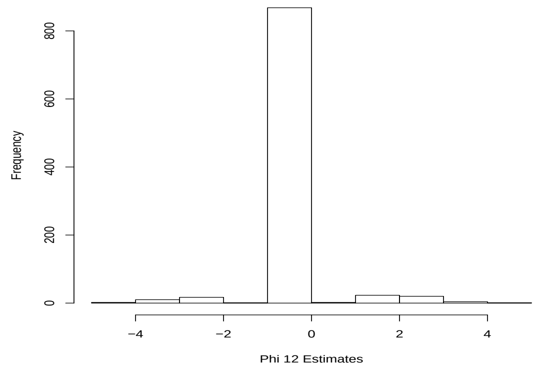

Figure 8: Estimates of $\hat{\phi_{12}}$ for LSTAR models with fixed γ and c

3.3 Experiment 3: Misspecification of the Conditional Mean

The third experiment is concerned with the problem of misspecification, namely the effects on the MLE of misspecifying a STAR-GARCH model as an AR-GARCH model are investigated. Furthermore, it is also of interest whether the misspecified conditional mean affects the MLE of the conditional variance. As the information matrix for a STAR-GARCH model is block diagonal (see Lundbergh and Teräsvirta (1999)), the effects of misspecifying the conditional mean on the MLE for the GARCH components should not be substantial, especially the standard errors.

This experiment is conducted by replicating 1000 series from an LSTAR-GARCH process, as defined in equations (3.1) and (3.4). AR-GARCH models are then estimated for each replication, where the specification of the AR-GARCH process is given as follows:

$$\begin{aligned} y_t &= \phi_1 + \phi_2 y_{t-1} + \varepsilon_t, \\ \varepsilon &= \eta_t \sqrt{h_t}, \\ h_t &= \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \end{aligned}$$

The same experiment is repeated for the exponential case.

A simple two-regime STAR model is given by:

$$y_t = (\phi_{11} + \phi_{12} y_{t-1})(1 - F(s_t; \gamma, c)) + (\phi_{21} + \phi_{22} y_{t-1}) F(s_t; \gamma, c) + \varepsilon_t,$$

which can be rewritten as

$$y_t = (\phi_{11} + (\phi_{21} - \phi_{11}) F(s_t; \gamma, c)) + (\phi_{12} + (\phi_{22} - \phi_{12}) F(s_t; \gamma, c)).y_{t-1} + \varepsilon_t.$$

The last expression shows that the effects of misspecifying a STAR-GARCH model as an AR-GARCH will depend on the choice of transition function. Indeed, one would expect that $\hat{\phi}_j$ of the AR-GARCH model is the mean of $(\phi_{1j} + (\phi_{2j} - \phi_{1j})F(s_t; \gamma, c))$, $j = 1, 2$. Although this seems to hold for the logistic case, it does not hold for the exponential case.

3.3.1 Logistic Conditional Mean

Figures 9 and 10 show the histograms of the $\hat{\phi}_1$ and $\hat{\phi}_2$ estimates, the means of which are -0.0780 and 0.0699, respectively.

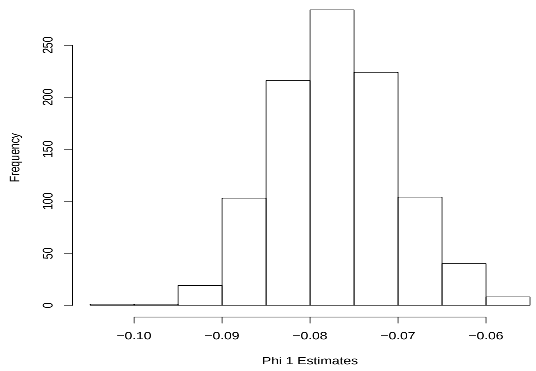

Figure 9: Estimates of $\hat{\phi}_1$ for AR-GARCH

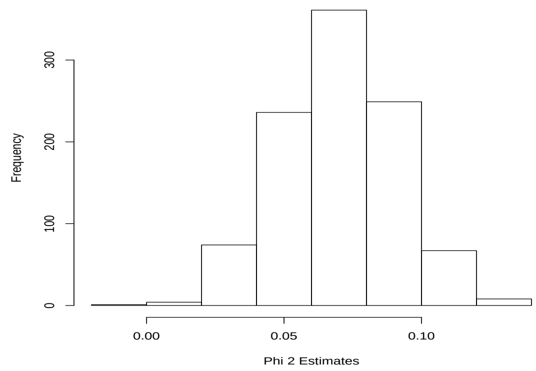

Figure 10: Estimates of $\hat{\phi}_2$ for AR-GARCH

The $\hat{\alpha}$ and $\hat{\beta}$ estimates are shown in Figures 11 and 13, with the means for $\hat{\alpha}$ and $\hat{\beta}$ being 0.214 and 0.731, respectively. These figures suggest that the estimates of the conditional variance are relatively close to the true values of 0.2 and 0.75, respectively. However, estimating LSTAR-GARCH models using these data yield $\hat{\alpha}$ and $\hat{\beta}$ estimates for the conditional variance, as shown in Figures 12 and 14. It is worth noting that the estimates for $\hat{\alpha}$ and $\hat{\beta}$ from estimating an LSTAR-GARCH model are 0.2011 and 0.7465, respectively, which are very close to the true values of the parameters.

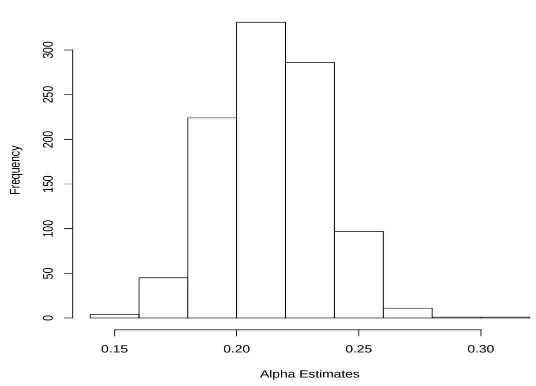

Figure 11: Estimates of $\hat{\alpha}$ for AR-GARCH

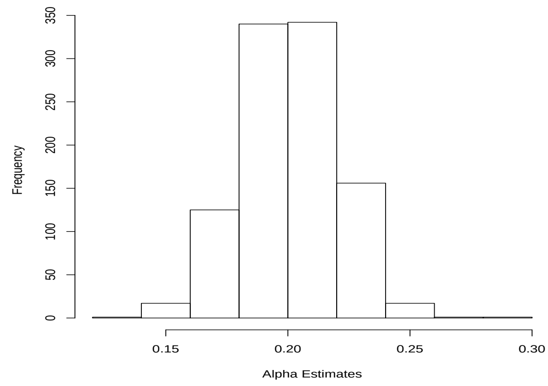

Figure 12: Estimates of $\hat{\alpha}$ for LSTAR-GARCH

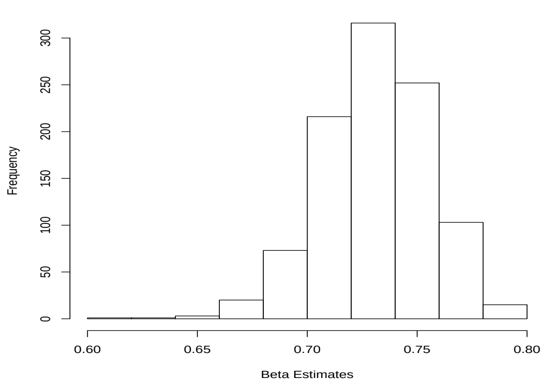

Figure 13: Estimates of $\hat{\beta}$ for AR-GARCH

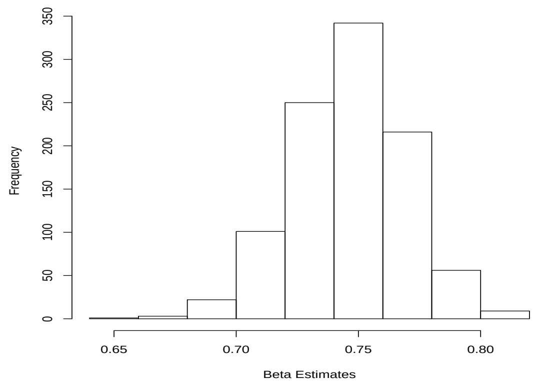

Figure 14: Estimates of $\hat{\beta}$ for LSTAR-GARCH

These results suggest that misspecification in the conditional mean leads to an upward (downward) bias in the $\hat{\alpha}$ ($\hat{\beta}$) estimates in finite samples. However, the bias does not seem to be substantial. Whether or not this bias is likely to cause greater problems in practice than other empirical issues, such as the choice of initial values in estimating a GARCH model (see Brooks, Burke and Persand (2001)) or the effects of outliers and extreme observations (see Verhoeven and McAleer (1999)), is an area for future research.

3.3.2 Exponential Conditional Mean

Figures 15 and 16 show the histograms of the $\hat{\phi}_1$ and $\hat{\phi}_2$ estimates, the means of which are -0.253 and -0.239, respectively.

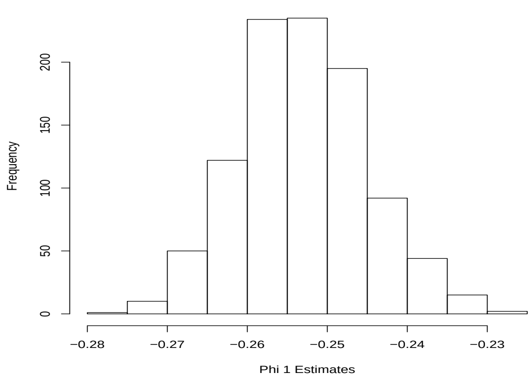

Figure 15: Estimates of $\hat{\phi}_1$ for AR-GARCH

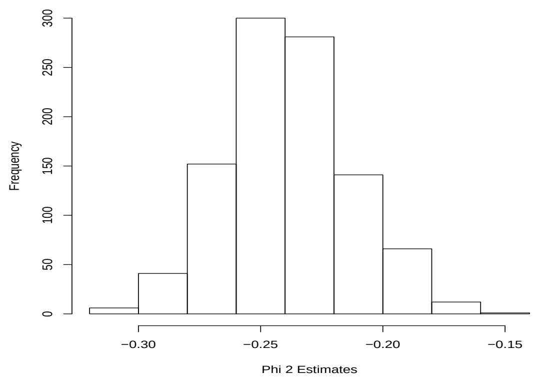

Figure 16: Estimates of $\hat{\phi}_2$ for AR-GARCH

Again, misspecification in the conditional mean leads to an upward (downward) bias in the $\hat{\alpha}$ ($\hat{\beta}$) estimates, as shown in Figures 17 to 20. Furthermore, the bias arising from this case seems to be more substantial than from the LSTAR-GARCH model. The means of the $\hat{\alpha}$ and $\hat{\beta}$ estimates for the AR-GARCH model are 0.236 and 0.703, respectively, while the means of the $\hat{\alpha}$ and $\hat{\beta}$ estimates from estimating ESTAR-GARCH model are 0.202 and 0.7462, respectively, which are very close to the true values of the parameters.

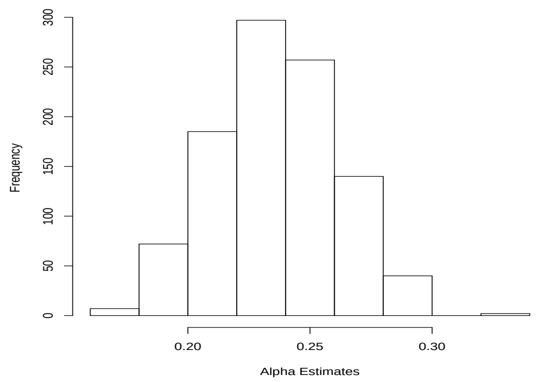

Figure 17: Estimates of $\hat{\alpha}$ for AR-GARCH

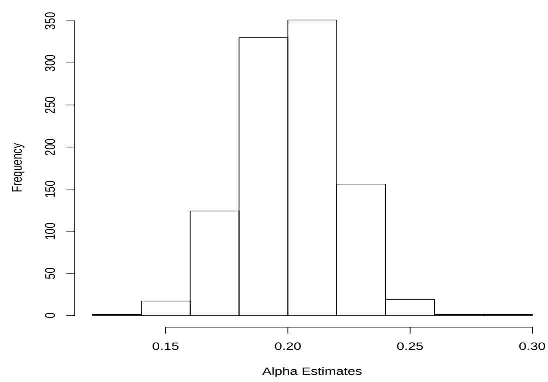

Figure 18: Estimates of $\hat{\alpha}$ for LSTAR-GARCH

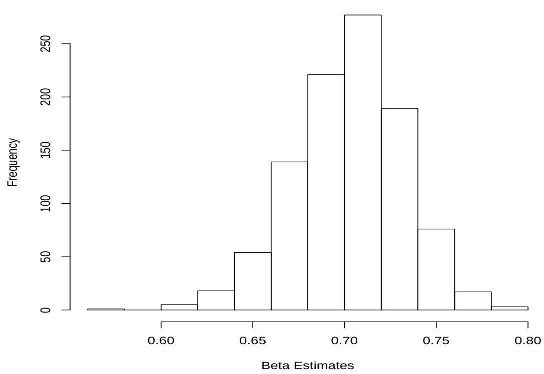

Figure 19: Estimates of $\hat{\beta}$ for AR-GARCH

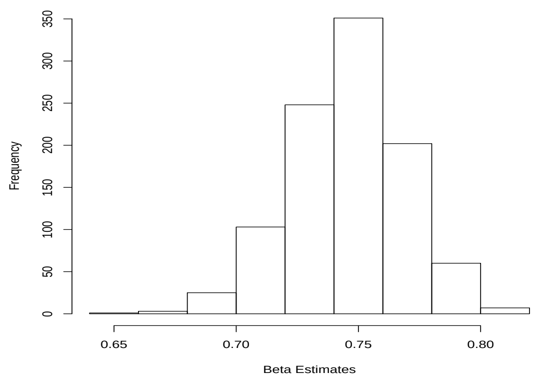

Figure 20: Estimates of $\hat{\beta}$ for LSTAR-GARCH

3.4 Experiment 4: Misspecification of the Transition Function

The fourth experiment investigates the effects on the parameter estimates of STAR-GARCH when the transition function is misspecified. In turn, the experiment involves estimating an LSTAR-GARCH (respectively, ESTAR-GARCH) model using data generated by an ESTAR-GARCH (respectively, LSTAR-GARCH) process, as defined in equations (3.1) and (3.4). The effects of misspecifying the transition function on the MLE for STAR-GARCH can be examined with the results given in Table 3 (the true processes are given in parentheses).

	LSTAR-GARCH (ESTAR-GARCH)	ESTAR-GARCH (LSTAR-GARCH)
$\hat{\phi_{11}}$	3.175	-0.202
$\hat{\phi_{12}}$	15.163	0.075
$\hat{\phi_{21}}$	-3.299	17.123
$\hat{\phi_{22}}$	-14.090	-2.240
$\hat{\gamma}$	1.888	0.280
$\hat{c}$	2.361	0.098
$\hat{\omega}$	0.011	0.010
$\hat{\alpha}$	0.215	0.201
$\hat{\beta}$	0.728	0.746

Table 3: Parameter Estimates when the Transition Function is Misspecified

The second column contains the means of all the parameter estimates for LSTAR-GARCH using data generated by an ESTAR-GARCH process, and the third column contains the means of all the parameter estimates for ESTAR-GARCH using data generated by an LSTAR-GARCH process. An interesting finding is that the parameter estimates in the GARCH components are reasonably close to the true values. However, these estimates also suggest that misspecifying the transition function when the true function is exponential causes a greater bias in the estimates of the GARCH components, which agrees with the findings from Experiment 3 (that is, the bias is greater when the true process is ESTAR-GARCH).

A second interesting result is that the threshold value, c, has increased in both cases, though the bias seems to be worse when the data follow an ESTAR-GARCH process. Furthermore, the effects on the $\hat{\gamma}$ estimates depend on the transition function. When the data follow an ESTAR-GARCH process, the MLE of γ for LSTAR-GARCH seems to be biased upwards. However, the MLE of γ for ESTAR-GARCH seems to be biased downwards when the data follow an LSTAR-GARCH process.

Moreover, the parameter estimates for each regime exhibit interesting patterns. In the case of LSTAR-GARCH, the intercept parameters, $\hat{\phi_{11}}$ and $\hat{\phi_{21}}$, are similar in absolute value. The same holds for the slope coefficients, $\hat{\phi_{12}}$ and $\hat{\phi_{21}}$. However, with a high threshold value of 2.361, it would seem that the first regime dominates the second most of the time. It is worth noting that the slope coefficients in this case both exceed 1 in absolute value, indicating two non-stationary processes, even though both regimes in the true underlying model are stationary.

For ESTAR-GARCH, the parameter estimates in the first regime are close to zero, but those in the second regime exceed 1 in absolute value. These patterns should help practitioners to verify the specification of their models.

4 An Empirical Illustration

This section investigates the sensitivity of the QMLE for STAR-GARCH models using S&P 500 Index returns from 1/1/1986 to 11/4/2000, giving 3726 data points. The plot of S&P returns is given in Figure 21. Both LSTAR-GARCH and ESTAR-GARCH are estimated using rolling windows of size 3226. The impact of each observation on the estimates can be observed by examining their dynamic paths.

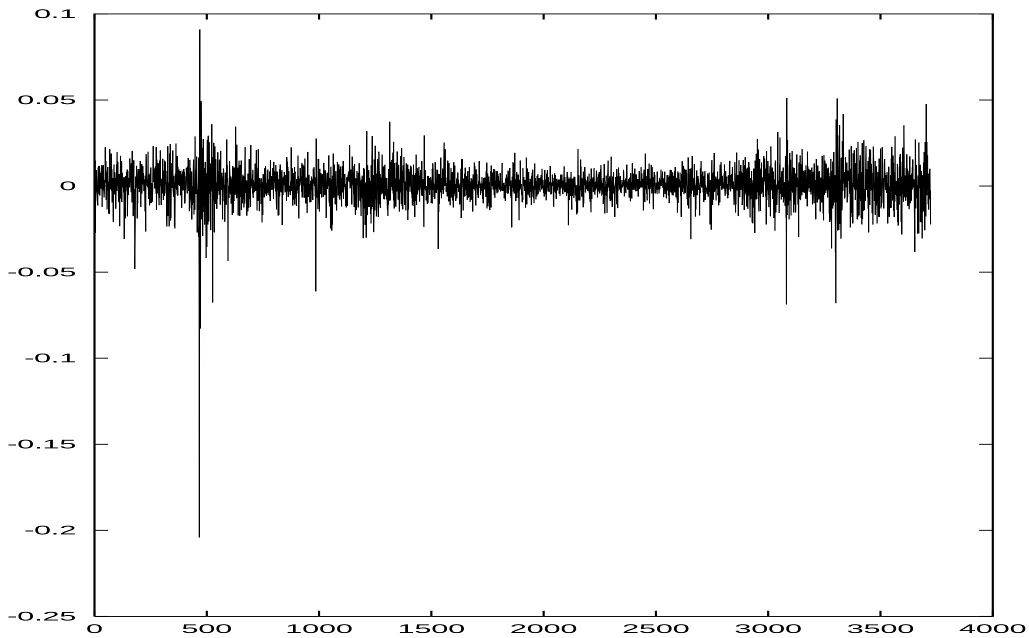

Figure 21: Returns of S&P 500 for 1/1/1986 - 11/4/2000

Figures 22 and 23 show the dynamic paths of the threshold estimates, $\hat{c}$, for LSTAR-GARCH and ESTAR-GARCH, with means 0.0358 and 0.234, respectively. The most interesting finding from these graphs is the clustering of the threshold estimates, that is, a series of positive estimates is followed by a series of negative estimates. Another interesting issue is the variability of the estimates. It appears that the threshold estimates are highly sensitive to the data, as well as to the specification of the transition function. These results agree with the findings from previous experiments. Furthermore, the threshold estimates also seem to be sensitive to the starting values of the series. This observation is supported by the fact that the threshold values exceed 5 for both models when the sample begins at data point 467, which corresponds to the stock market crash of the 13 October 1987. However, the estimates remain close to zero after this outlier has left the rolling samples.

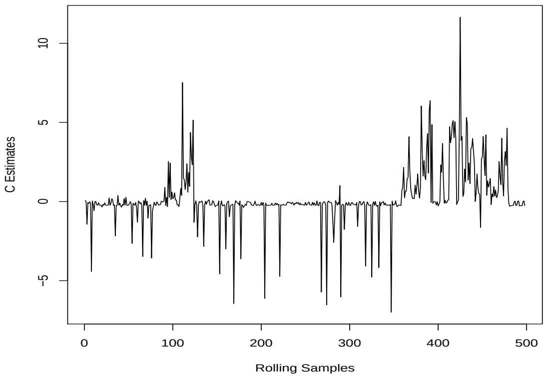

Figure 22: Dynamic Path of the Threshold Values: Logistic

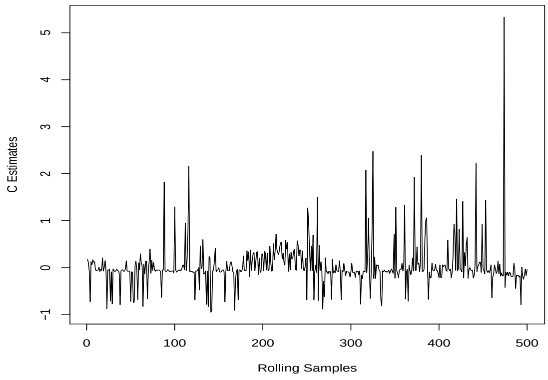

Figure 23: Dynamic Path of the Threshold Values: Exponential

Table 4 presents the means of all the parameter estimates for the three models. These estimates suggest that there may be a misspecification problem with the transition function. In the LSTAR-GARCH case, the parameter estimates in each regime represent a non-stationary process, corresponding to the findings from Experiment 4. Moreover, in the ESTAR-GARCH case, only the second regime represents a non-stationary process, which also corresponds to the findings in Experiment 4. Furthermore, the transition rate and the threshold value are both greater for LSTAR-GARCH than for ESTAR-GARCH, which also corresponds to the findings in Experiment 4. This result suggest that neither the logistic nor the exponential function is appropriate for describing the switching behaviour of this data set. It is important to note, however, that the findings from Experiment 4 are necessary but not sufficient for misspecification. Other misspecification problems could also lead to similar results as those reported here.

It is important to note that, for LSTAR-GARCH, the number of rolling samples in which no observation is greater or smaller than the corresponding estimated threshold values are 139 and 189, respectively. In the case of ESTAR-GARCH, the number of rolling samples in which no observation is greater or smaller than the corresponding threshold values are 80 and 129, respectively. These samples are problematic as it appears that a lack of information for one regime may lead to unreliable parameter estimates of the model. This is another indication of model misspecification.

	LSTAR-GARCH	ESTAR-GARCH	AR-GARCH
$\hat{\phi_{11}}$	-16.8512	-0.07131	0.001261
$\hat{\phi_{12}}$	-18.795	-0.21041	0.018557
$\hat{\phi_{21}}$	17.42053	13.9072	-
$\hat{\phi_{22}}$	11.73428	-28.7238	-
$\hat{\gamma}$	0.23295	0.035591	-
$\hat{c}$	11.55064	1.791424	-
$\hat{\omega}$	6.69E-06	1.37E-06	7.38E-06
$\hat{\alpha}$	0.10311	0.082924	0.110836
$\hat{\beta}$	0.854071	0.906356	0.842498

Table 4: Means of Parameter Estimates for Three Empirical Models

Figures 24 to 29 display the dynamic paths of the GARCH components for the AR-GARCH, LSTAR-GARCH and ESTAR-GARCH models. As shown in these figures, ESTAR-GARCH seems to be extremely sensitive to the outlier, as supported by the dramatic decrease (increase) in $\hat{\alpha}$ ($\hat{\beta}$). These results suggest that the finite sample bias for estimating ESTAR-GARCH is even more substantial in the presence of outliers and extreme observations. This is a new result and would benefit from further investigation, possibly via the use of different trimming algorithms, as suggested in Verhoeven and McAleer (1999).

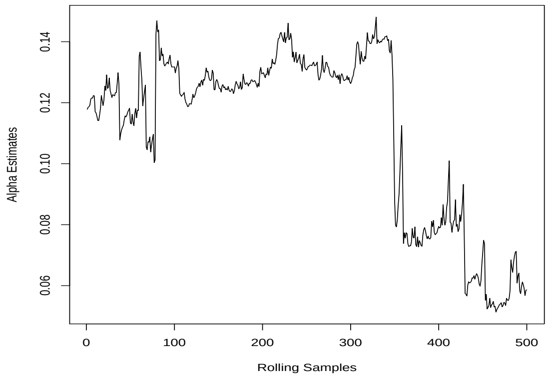

Figure 24: AR-GARCH $\hat{\alpha}$

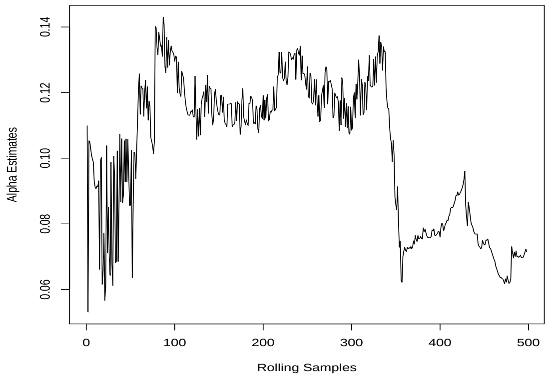

Figure 25: LSTAR-GARCH $\hat{\alpha}$

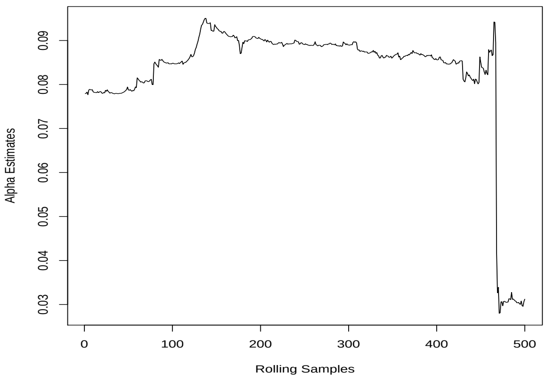

Figure 26: ESTAR-GARCH $\hat{\alpha}$

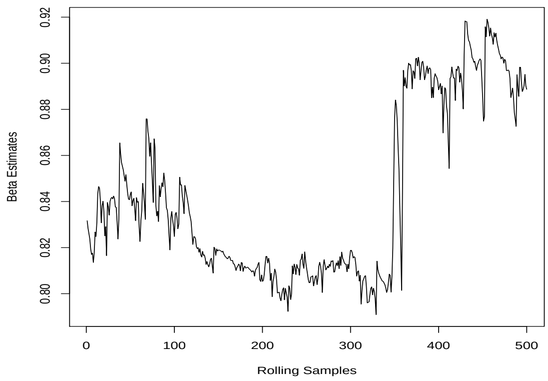

Figure 27: AR-GARCH $\hat{\beta}$

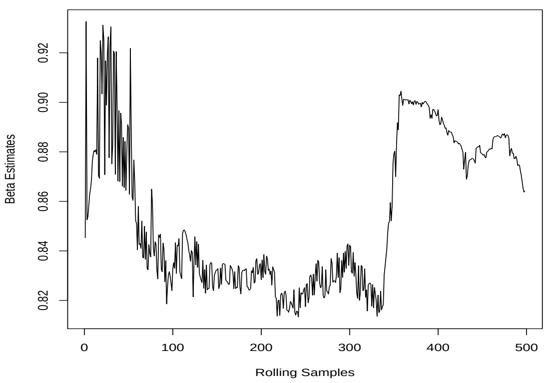

Figure 28: LSTAR-GARCH $\hat{\beta}$

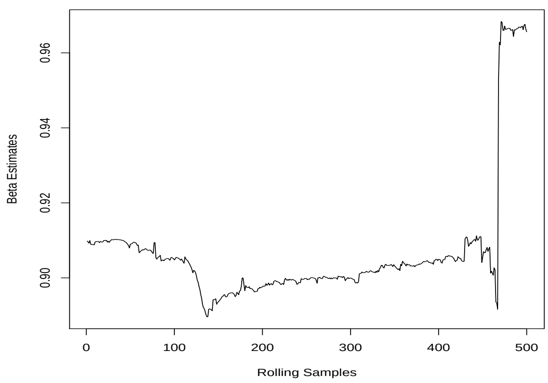

Figure 29: ESTAR-GARCH $\hat{\beta}$

5 Concluding Remarks

The paper presented a number of simulation results regarding the finite sample properties of MLE for STAR and STAR-GARCH models. Sensitivity of the estimates to different variances was investigated for both LSTAR and ESTAR models. These results showed that the variability of the threshold value, c, was generally directly related to the magnitude of the variance of the unconditional shocks for LSTAR, but inversely related to the variance of the unconditional shocks for ESTAR. This is an interesting finding which requires some structural and statistical properties before a satisfactory explanation can emerge. Such theoretical properties are not yet available in the literature.

Furthermore, the paper has also shown that fixing either the threshold value or the transition rate does not necessarily improve the accuracy of the estimates. In fact, for LSTAR, it may actually worsen the estimates. However, fixing the threshold value or the transition rate improves the accuracy of the other estimates for the ESTAR model, which implies that BFGS may not be a robust algorithm for LSTAR.

Misspecification of the conditional mean was also examined. Specifically, AR-GARCH models were estimated using data generated by LSTAR-GARCH and ESTAR-GARCH processes. The results showed that the estimates for the GARCH components were not biased substantially. However, the bias in finite samples seemed to be worse if the underlying process was ESTAR-GARCH rather than LSTAR-GARCH, which suggests that the magnitude of the finite sample bias depends on the functional form of the transition function.

The last experiment showed that misspecifying the transition function does not lead to substantial bias in the estimates of the GARCH components, but it could yield seemingly meaningless estimates of the conditional mean. In misspecifying an exponential function as a logistic, the coefficient estimates for the lagged dependent variables were both greater than 1 in absolute value, indicating two non-stationary processes. Furthermore, the threshold values, as well as the transition rates, were biased upwards. However, in the case of misspecifying a logistic function as an exponential, only the coefficient estimates of the intercept and lagged dependent variable in the second regime were greater than 1 in absolute value, with the estimates in the first regime seemingly biased towards 0. The estimates of the threshold value have also been biased upwards, but the transition rate seemed to be biased downwards, contrary to pevious findings.

Overall, the results presented here reflect the complex nature of STAR-type models. The results suggest that structural and statistical results are required before such models can be applied with confidence in practice. Proper diagnostic tests would be crucial in evaluating these models and their forecasts, but these tests cannot be established without an understanding of the theoretical properties of the models.

The performance of different optimisation algorithms is also an important issue. Numerical results presented here should assist in ensuring that an appropriate algorithm can be chosen to optimise the log-likelihood function, and thereby obtain robust estimates of the parameters for the various models.

REFERENCES

E.K. Berndt, B.H. Hall, R.E. Hall, and J.A. Hausman. Estimation and Inference in Non-linear Structural Models. *Annals of Economic and Social Measurement*, 3:653–665, 1974.

P.J. Bickel. On Adaptive Estimation. *Annals of Statistics*, 10:647–671, 1982.

T. Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

P. Bougerol and N.M. Picard. Stationarity of GARCH Processes and of Some Non-negative Time Series. *Journal of Econometrics*, 52:115–127, 1992.

C. Brooks, S.P. Burke, and G. Persand. Benchmarks and the Accuracy of GARCH Model Estimation. *International Journal of Forecasting*, 17:45–56, 2001.

C.H. Chen (editor). *Pattern Recognition and Signal Processing.* Amsterdam: Sijhoff and Noordoff, 1978.

M.P. Clements and D.F. Hendry. *Forecasting Economic Time Series.* Cambridge: Cambridge University Press, 1998.

F.X. Diebold and J.A. Nason. Non-parametric Exchange Rate Prediction. *Journal of International Economics*, 28:315–332, 1990.

D. van Dijk, S. Lundbergh, and T. Teräsvirta. Time Varying Smooth Transition Autoregressive Models. To appear in *Journal of Business and Economics Statistics.* 2000.

D. van Dijk, T. Teräsvirta, and P.H. Franses. Smooth Transition Autoregressive Models - A Survey of Recent Developments. To appear in *Econometric Reviews.* 2000.

Z. Ding, C.W.J. Granger, and R.F. Engle. A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1:83–106, 1993.

R.F. Engle. Autoregressive Conditional Heteroscedasticity, with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50:987–1007, 1982.

R.F. Engle and D.L. McFadden (editor). *Aspects of Modelling Nonlinear Time Series.* Amsterdam: Elsevier Science, 1994.

P.H. Franses and D. van Dijk. *Non-linear Time Series Models in Empirical Finance.* Cambridge University Press, 2000.

P.H. Franses, J. Neele, and D. van Dijk. Forecasting Volatility with Switching Persistence GARCH Models. *Econometric Institute Report, EI-9819*, 1998.

L. Glosten, R. Jagannathan, and D. Runkle. On the Relation Between the Expected Value and Volatiity of Nominal Excess Return on Stocks. *Journal of Finance*, 46:1779–1801, 1992.

C. He and T. Terašvirta. Properties of Moments of a Family of GARCH Processes. *Journal of Econometrics*, 92:173–192, 1999a.

C. He and T. Teräsvirta. Fourth Moment Structure of the GARCH(p, q) Process. *Econometric Theory*, 15:824–846, 1999b.

S.W. Lee and B.E. Hansen. Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator. *Econometric Theory*, 10:29–52, 1994.

S. Ling. On the Probabilitistic Properties of a Double Threshold ARMA Conditional Heteroscedasticity Model. *Journal of Applied Probability*, 36:1–18, 1999.

S. Ling and W.K. Li. On Fractional Integrated Autoregressive Moving Average Time Series Models with Conditional Heteroscedasticity. *Journal of the American Statistical Association*, 92:1184–1194, 1997.

S. Ling and M. McAleer. Necessary and Sufficient Moment Conditions for the GARCH(r, s) and Asymmetric Power GARCH(r, s) Models. To appear in *Econometric Theory.* 1999a.

S. Ling and M. McAleer. Asymptotic Theory for a Vector ARMA-GARCH Model. To appear in *Econometric Theory.* 1999b.

S. Ling and M. McAleer. Stationarity and the Existence of Moments of a Family of GARCH Processes. To appear in *Journal of Econometrics.* 1999c.

S. Ling and M. McAleer. On Adaptive Estimation in Nonstationary Models with GARCH Errors. To appear in *Annals of Statistics.* 1999d.

R.L. Lumsdaine. Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models. *Econometrica*, 64:575–596, 1996.

S. Lundbergh and T. Teräsvirta. Modelling Economic High Frequency Time Series with STAR-STGARCH Models. *SSE/EFI Working Paper Series in Economics and Finance*, No. 291, 1999.

S. Lundbergh and T. Teräsvirta. Forecasting with Smooth Transition Autoregressive Models. *SSE/EFI Working Paper Series in Economics and Finance*, No. 390, 2000.

T.K. Mak, H. Wong, and W.K. Li. Estimation of Nonlinear Time Series with Conditional Heteroscedasticity Variances by Iteratively Weighted Least Squares. *Computational Statistics & Data Analysis*, 24:169–178, 1997.

A. Milhöj. The Moment Structure of ARCH Processes. *Scandinavian Journal of Statistics*, 12:281–292, 1985.

D.B. Nelson. Stationarity and Persistence in the GARCH(1,1) Model. *Econometric Theory*, 6:318–334, 1990.

D.B. Nelson. Conditional Heteroscedasticity in Asset Returns: A New Approach. *Econometrica*, 59:347–370, 1991.

S.G. Pantula. Estimation of Autoregressive Models with ARCH Errors. *Sankhya B*, 50:119–138, 1989.

B.M. Pötscher and I.V. Prucha. *Dynamic Nonlinear Econometric Models - Asymptotic Theory.* Springer-Verlag, 1997.

P.M. Robinson. Semiparametric Econometrics: A Survey. *Journal of Applied Econometrics*, 3:35–51, 1988.

T.M. Stoker. *Lectures on Semiparametric Econometrics.* CORE, Louvain-la-Neuve, Belgium, 1991.

T. Terašvirta. Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association*, 89:208–218, 1994.

T. Teräsvirta and H.M. Anderson. Characterizing Nonlinearities in Business Cycle Using Smooth Transition Autoregressive Models. *Journal of Applied Econometrics*, 7:119–136, 1992.

T. Teräsvirta, D. Tjøstheim, and C.W.J. Granger. Aspects of Modelling Nonlinear Time Series. In Engle and McFadden (1994), pages 2917-2957.

H. Tong. On a Threshold Model. In Chen (1978), pages 101-141.

H. Tong and K.S. Lim. Threshold Autoregressive, Limit Cycles and Data. *Journal of the Royal Statistical Society B*, 42:245–292, 1980.

R.L. Tweedie. Invariant Measure for Markov Chains with No Irreducibility Assumptions. *Journal of Applied Probability*, 25A:275–285, 1988.

P. Verhoeven and M. McAleer. Modelling Outliers and Extreme Observations for ARMA-GARCH Processes. Submitted. 1999.

A.A. Weiss. Asymptotic Theory for ARCH Models: Estimation and Testing. *Econometric Theory*, 2:107–131, 1986.

J.M. Wooldridge. Estimation and Inference for Dependent Processes. In Engle and McFadden (1994), pages 2639-2738.