Discussion Paper No. 539

**ESTIMATING SMOOTH TRANSITION
AUTOREGRESSIVE MODELS
WITH GARCH ERRORS IN THE PRESENCE
OF EXTREME OBSERVATIONS
AND OUTLIERS**

Felix Chan
and
Michael McAleer

May 2001

Estimating Smooth Transition Autoregressive Models with GARCH Errors in the Presence of Extreme Observations and Outliers *

Felix Chan and Michael McAleer

Department of Economics
University of Western Australia

April 2001

Abstract

This paper investigates several empirical issues regarding quasi-maximum likelihood estimation of Smooth Transition Autoregressive (STAR) models with GARCH errors, specifically STAR-GARCH and STAR-STGARCH. Convergence, the choice of different algorithms for maximising the likelihood function, and the sensitivity of the estimates to outliers and extreme observations, are examined using daily data for S&P 500, Heng Seng and Nikkei 225 for the period January 1986 to April 2000.

1   Introduction

Interest in non-linear time series models has increased rapidly in recent years. In particular, regime switching models have been rather popular in the class of non-linear models. Given the substantial research activity in analysing time-varying volatility through GARCH process (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 (STAR) - Generalised Autoregressive Conditional Heteroscedasticity (STAR-GARCH) and STAR - Smooth Transition GARCH (STAR-STGARCH) models.

Although STAR-GARCH and STAR-STGARCH are popular and have been used widely in forecasting (Franses, Neele and van Dijk (1998), Lundbergh and Teräsvirta (1999, 2000)), the statistical and structural properties of these models have not yet been fully established. Furthermore, the existing diagnostic tests for these models assume consistency and asymptotic normality, but these assumed statistical properties cannot be examined in detail 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 (2000a), 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 optimised values.

This paper provides empirical evidence to show that different algorithms produce substantially different estimates for the same model. Consequently, the interpretation of the model can differ according to the choice of algorithm.

Moreover, forecast performances may also be affected. This is contrary to the common belief that different algorithms will produce similar estimates (Lundbergh and Teräsvirta (2000)). This paper also shows that the convergence of the QMLE for the STAR-STGARCH model depends on the choice of transition functions.

The second part of the paper examines the effects of extreme observations and outliers on the QMLE for the STAR-GARCH and STAR-STGARCH models. This is of interest because STAR-type models were not designed to accommodate extreme observations and outliers. However, these models are often used to model financial time series, which frequently exhibit excessive kurtosis. Therefore, it is important to investigate the robustness of the QMLE for STAR-type models in the presence of extreme observations and outliers in order to determine how best to accommodate such data.

This paper also provides empirical evidence to show that the effects of extreme observations and outliers on the QMLE for the STAR component in a STAR-GARCH model depend on the choice of transition functions. The effects of such data on the QMLE for the GARCH component are similar to those for ARMA-GARCH, as reported in Verhoeven and McAleer (1999). This result has not previously been investigated. Moreover, empirical evidence also suggests that the QMLE for STAR-STGARCH models is sensitive to the presence of extreme observations and outliers.

The plan of the paper is as follows. Section 2 gives a brief outline of recent developments for the GARCH, STAR, STAR-GARCH and STAR-STGARCH models. Section 3 discusses the data. Section 4 presents a detailed discussion on various optimisation algorithms and their effects on the QMLE for STAR-GARCH models. Section 5 investigates the effects of extreme observations and outliers on the estimates for both STAR-GARCH and STAR-STGARCH models. Section 6 gives some concluding remarks.

2   Models

This section provides a brief discussion of recent developments in modeling GARCH, STAR, STAR-GARCH and STAR-STGARCH. Model definitions,

characterisitics and statistical properties 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:

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

in which ε_t is said to follow an Autoregressive Conditional Heteroscedasticity, ARCH(p), process if

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

where

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

$$h_t = \omega_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2. \tag{2.4}$$

This model was proposed by Engle (1982) to relax the traditional assumption of a constant one-period forecast variance. Engle showed that this model has a constant unconditional variance but a non-constant variance conditional on the past.

Engle (1982) showed that ε_t is second-order stationary (that is, the second moment of ε_t is finite) 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. This assumption is clearly not possible to check in practice. However, Engle (1982) also derived the regularity condition for the existence of the moments for ARCH(1), specifically the $2m$th moment exists if and only if

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

3

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

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

He also derived the regularity condition for the existence of moments without the restrictive assumption. Milhøj's 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

$$\hat{\delta} = \frac{\sum_{t=2}^{n} \tilde{\varepsilon}_{t-1} \tilde{\varepsilon}_t}{\sum_{t=2}^{n} \tilde{\varepsilon}_{t-1} \tilde{\varepsilon}'_{t-1}} \tag{2.6}$$

where $\hat{\delta} = (\omega_0, \alpha_1, ..., \alpha_p)$ 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:

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

so that the MLE is given as

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

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, which implies the

4

parameters in the conditional mean and the conditional variance can be estimated separately without loss of asymptotic efficiency. The estimated errors given by the estimated conditional mean equation can be used to estimate the equation of the conditional variance. However, it is important to note that the restrictions $\omega_0 > 0, \alpha_i > 0$ for all $i = 1, .., p$, are required to ensure that the conditional variance is positive.

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 should be used.

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

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

If $\beta_i = 0$, for all i, then GARCH(p, q) reduces to an ARCH(p) model. All the mathematical and statistical properties of GARCH hold for ARCH, in general, except for one case which will be 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 strict stationarity and ergodicity for GARCH(1,1) as

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

This 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, and also involves 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) examined the fourth moment structure of the general GARCH(p, q) process. In the case of GARCH(1,1), the fourth moment condition 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

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

6

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

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

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 completely 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 in a GARCH(p, q) model are often estimated by MLE, or by QMLE when the normality of η_t is not assumed. The log-likelihood function of GARCH(p,q) is identical to that of (2.7), except for the definition of h_t. Naturally, h_t in this case follows the definition of (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

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

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

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

7

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) and Stoker (1991). This estimation method is not yet available in most econometric software packages due to its computational complexity, which explains in part the popularity of MLE in this literature.

It is important to note that the choice of lag length in the conditional variance equation has not been well 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 the 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. The GARCH model also offers computational advantages over extended versions thereof. 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.

Moreover, it is important to impose a restriction on all of the parameters in GARCH models to ensure the positivity of the conditional variances. These restrictions can create difficulties in estimating GARCH, 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. Regime switching models are very popular in the class of non-linear models, so it is of interest to investigate regime switching models with GARCH errors. Regime switching models will be discussed here, with 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:

$$y_t = (\phi_{10} + \sum_{i=1}^{r} \phi_{1i} y_{t-i})(1 - I(s_t - c)) + (\phi_{20} + \sum_{i=1}^{r} \phi_{2i} y_{t-i})) I(s_t - c) + \varepsilon_t \quad (2.10)$$

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:

$$y_t = \begin{cases} \phi_{10} + \sum_{i=1}^{r} \phi_{1i} y_{t-i} & s_t < c \\ \phi_{20} + \sum_{i=1}^{r} \phi_{2i} y_{t-i} & s_t \geq c. \end{cases}$$

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

$$s_t = \sum_{i=0}^{k} \pi_i y_{t-i}$$

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

9

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, Teräsvirta (1994) proposed the Smooth Transition Autoregressive (STAR) model:

$$y_t = (\phi_{10} + \sum_{i=1}^{r} \phi_{1i} y_{t-i})(1 - G(s_t; \gamma, c)) + (\phi_{20} + \sum_{i=1}^{r} \phi_{2i} y_{t-i}))G(s_t; \gamma, c) + \varepsilon_t$$

(2.11)

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

Although two regimes will suffice in 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 TAR (or Multiple Regime TAR, MRTAR) model can be written as

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

(2.12)

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

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

For any $s_t \in [c_{i-1}, c_i)$, $y_t = \phi_i x_t$ for all $i = 1...m$. Therefore, the regime is determined by the threshold variable, s_t, relative to the threshold value, c_i. 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

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

(2.13)

where $G_i(s_t; \gamma_i, c_i)$ is assumed to be a twice differentiable function 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, Teräsvirta and Franses (2000b)).

There are many choices of transition function, 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:

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

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:

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

and the nth-order LSTAR:

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

In order to use this model effectively, it is important to choose the 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 under the STAR framework is given in van Dijk, Teräsvirta and Franses (2000a)).

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 with 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 Square (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 the 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 natural extension has not yet been investigated widely. Lundbergh and Teräsvirta (1999) give a comprehensive exposition of this model, but do not provide any statistical properties or regularity conditions for the existence of its moments and for stationarity. These important properties have not yet been established. However, as the information matrix of the log-likelihood function 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 described earlier 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 component, 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 a twice differentiable function 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

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

13

where

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

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

STAR-STGARCH is novel and has a number of distinct characteristics. First, it is non-linear, not only in the conditional mean, but also 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 data that exhibit non-linearity and threshold behaviour, there are as yet no results as to the moment structure or statistical properties 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 (2000a), 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 understanding the statistical properties and moment structure of the model.

There are several LM-type specification tests to analyse the most appropriate model. Lundbergh and Teräsvirta (1999) and van Dijk, Teräsvirta and Frances (2000a) 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 as yet available. Therefore, any empirical examples using STAR-GARCH and STAR-STGARCH remain questionable in terms of their reliability and stability.

There is as yet no theoretical result regarding the stationarity of the STAR, STAR-GARCH or STAR-STGARCH models. Consider a two regime STAR model, as defined in (2.13), which can be rewritten as

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

Does the condition

$$
\phi_{12}(1 - G(s_t; \gamma, c)) + \phi_{22}G(s_t; \gamma, c) < 1
$$

ensure stationarity? This is not at all obvious as the transition function $G(s_t; \gamma, c)$ is a function of the endogenous variable y_t. More research in establishing the statistical properties and regularity conditions of these models is required before these models can be used with confidence.

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

3 Data

Following Lundbergh and Teräsvirta (1999), all regimes for each model estimated below are assumed to follow an AR(1) process, ε_t is assumed to be GARCH(1,1), s_t is set equal to y_{t-1}, and r_t is set equal to ε_{t-1}.

All models are estimated using three different stock indices, namely Standard and Poor's 500 Composite Index (S&P), Heng Seng Index and Nikkei 225 Index. The data were obtained through the DataStream database service and the sample is from 1/1/1986 to 11/4/2000, giving a total of 3725 data points for each index.

Of primary concern are stock returns, R_t, which are calculated as

$$R_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

where Y_t denotes the index at time t.

The returns for S&P 500, Heng Seng and Nikkei 225 are given in the following figures:
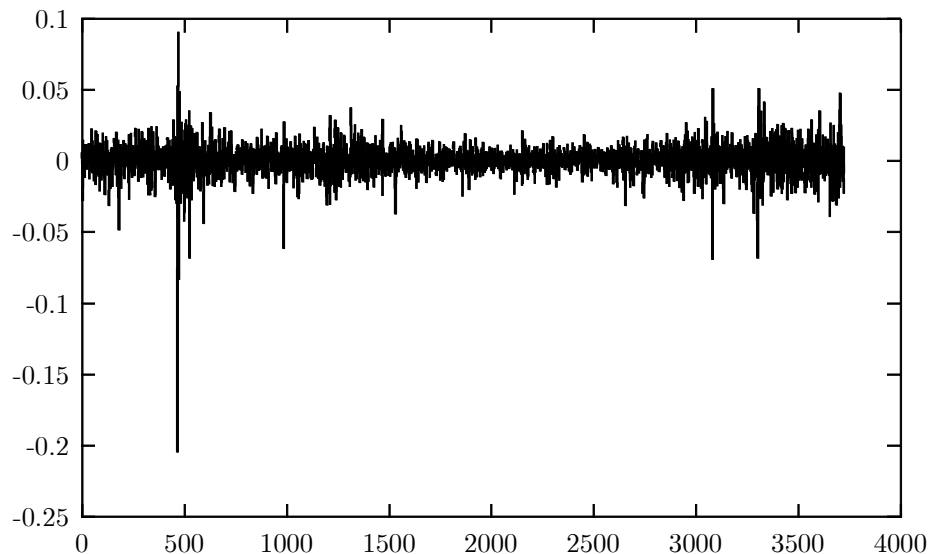
Figure 1: S&P Returns
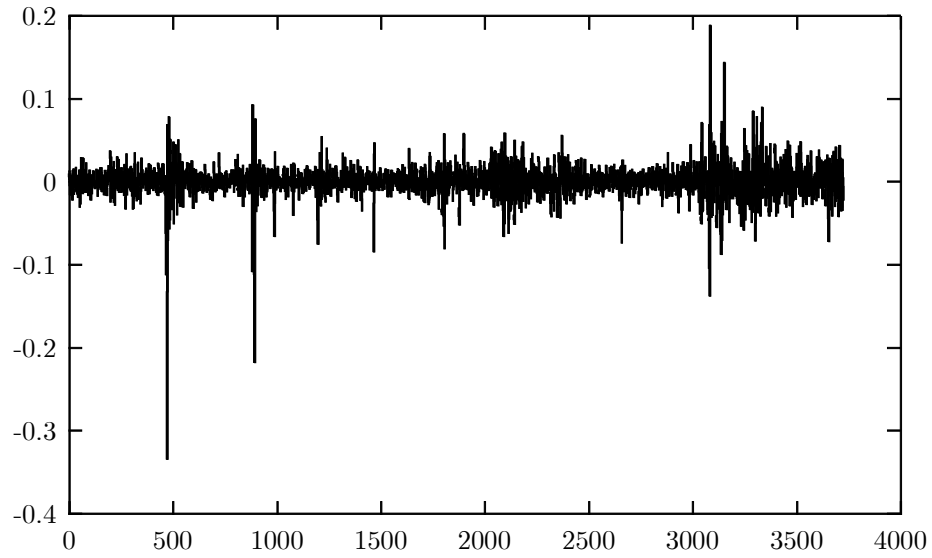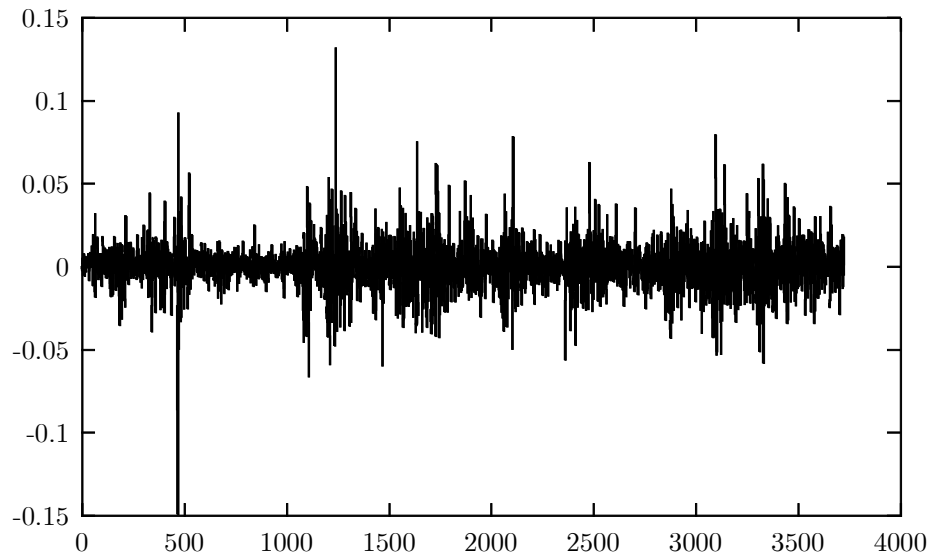
Figure 2: Heng Seng Returns

Figure 3: Nikkei 225 Returns

S&P appears to be less volatile than either Nikkei 225 or Heng Seng, especially during the early to late 1990's (observations 1500 to 3000) before the Asian economic and financial crises. S&P also has fewer extreme observations and outliers than Nikkei and Heng Seng.

Nikkei seems to have more positive shocks than S&P and Heng Seng. Although Nikkei seems more volatile, the volatility is relatively low compared with Heng Seng, which seems to be the most volatile. There are some obvious outliers and extreme observations for all three indices. Heng Seng also appears to have the highest number of outliers.

An obvious similarity among the three indices is the enormous decrease in returns at observation 474, which corresponds with the share market crash in October 1987. This is also the most significant outlier in the three indices. The second largest decrease in returns is observation 894 for Heng Seng, which corresponds with the Tianenman Square incident in Beijing on 4 June 1989.

4   Optimisation Algorithms

Estimation for STAR-type models is problematic as their novelty means that existing econometric software packages do not yet have appropriate algorithms programmed. STAR can be estimated by Non-linear Least Squares (NLS) (see van Dijk, Teräsvirta and Franses (2000a)). STAR-GARCH can be estimated by a two-stage procedure, which involves estimating STAR by NLS, then using the residuals to estimate GARCH by QMLE. However, this procedure is not appropriate for STAR-STGARCH, for which the information matrix is not block diagonal, so the estimates have to be obtained simultaneously. It is worth noting that NLS is equivalent to MLE under the assumption of normality. If this assumption does not hold, then NLS is equivalent to QMLE.

In practice, estimation of STAR and STAR-GARCH models can be problematic. An attempt was made in EViews to estimate LSTAR by NLS with $s_t = Y_{t-1}$ for S&P 500, Heng Seng and Nikkei, but was unsuccessful because of computational problems in calculating the near singular Hessian matrix.

A GAUSS version 3 program was used to estimate various STAR-type models by optimising the respective likelihood functions. Several attempts were made to estimate LSTAR and ESTAR for S&P, Heng Seng and Nikkei, for which the estimates of the variance did not converge. This suggests three possibilities: (i) the variance is not constant, so that STAR-GARCH should be used; (ii) the use of alternative optimisation algorithms (see for example, Luenberger (1989)); and (iii) the use of alternative initial values.

In order to investigate the effects of different algorithms on the estimates of STAR-GARCH models, two were chosen to maximise the log-likelihood function of a LSTAR-GARCH model, as defined in equations (2.13) and (2.8), using S&P, Heng Seng and Nikkei data. The two algorithms used are the Newton method and Broyden-Fletcher-Goldfarb-Shanno method (BFGS), a quasi-Newton method, with the same initial values, yielding the estimates in Table 1 (one asterisk denotes significance at the 5% level and two asterisks denote significance at 1%).

Table 1: LSTAR-GARCH Estimates

	S&P		Heng Seng		Nikkei	
	Newton	BFGS	Newton	BFGS	Newton	BFGS
$\hat{\phi_{11}}$	1.7833**	-3.0150**	-0.2238**	0.0142	-0.2796**	-1.5166**
$\hat{\phi_{12}}$	1.1204**	-4.1005**	2.0523**	-4.9601**	-0.1930**	0.1396**
$\hat{\phi_{21}}$	-0.2862**	3.0088**	0.2260**	0.0007	0.0013**	1.5131**
$\hat{\phi_{22}}$	-1.3476**	3.8090**	-1.8170**	0.3180**	0.0126	-0.2892**
$\hat{\gamma}$	2.4515**	1.4818**	4.3080**	3.9604**	1.6278**	1.0020**
$\hat{c}$	0.8634**	-0.0402**	-0.0162**	-0.8311**	-2.0104**	-0.0586**
$\hat{\omega}$	1.17e-6	1.17e-6**	7.08e-6**	7.10e-6**	3.2e-6**	3.21e-6**
$\hat{\alpha}$	0.0753**	0.0754**	0.1488**	0.1486**	0.1475**	0.1465**
$\hat{\beta}$	0.9166**	0.9166**	0.8394**	0.8394**	0.8519**	0.8529**

As shown in Table 1, the two algorithms produce different estimates for the conditional mean but not of the conditional variance. In fact, a similar set of estimates for the conditional variance can be obtained by estimating a simple AR(1)-GARCH(1,1) model using the Berndt-Hall-Hall-Hausman (BHHH) algorithm. This suggests that only the estimates of the conditional mean are sensitive to the choice of algorithm, which is contrary to the findings in Lundbergh and Teräsvirta (2000).

In particular, the high threshold values for Heng Seng imply that the first regime dominates the second. However, since BFGS gives substantially different estimates from Newton, it is difficult to determine which set of estimates will produce better forecasts. Moreover, the estimates from the two algorithms are highly significant, which makes interpretation problematic. As robustness to the choice of algorithm is not in evidence here, this stresses the importance of establishing regularity conditions for consistency and asymptotic normality.

A possible explanation of these differences is that the Hessian matrix of the log-likelihood function of the LSTAR-GARCH model is not accurately approximated by BFGS, in which case the covariance matrix will also be unreliable. Another possibility is that there exists more than one optimum

for the log-likelihood function, or that it is flat. This is supported by the similar mean likelihood scores of the two algorithms, as shown in Table 2.

Table 2: Maximum Likelihood Score

Index	Newton	BFGS
S&P	4.24018	4.24018
Heng Seng	3.75686	3.75664
Nikkei	3.92471	3.92487

As the structural and statistical properties of the STAR-GARCH models have not yet been established, it is difficult to provide a clear and unambiguous explanation for this result.

5   Extreme Observations and Outliers

The technical definitions of extreme observations and outliers are somewhat arbitrary. Extreme observations are often referred to as being 2 to 3 standard deviations from the mean. Outliers are often defined as being more than 3 standard deviations from the mean. The difference between extreme observations and outliers is that outliers can also be defined as observations that were not generated from the same population as the other observations in the sample.

Stock returns often contain more extreme observations and outliers as compared with a normal distribution. Consequently, the distribution seems to have fatter tails, or excessive kurtosis, than a normal distribution.

In order to examine the effects of extreme observations and outliers on the estimates for LSTAR-GARCH, ESTAR-GARCH and ESTAR-LSTGARCH, each of the data sets is adjusted using the following trimming algorithm:

1. Calculate the standard deviation for the sample;

2. If an observation is 4 times larger than the standard deviation, it is reduced to 4 times the standard deviation;

21

3. If an observation is between 3 and 4 times larger than the standard deviation, it is reduced to 3 times the standard deviation;

4. If an observation is between 2.5 and 3 times larger than the standard deviation, it is reduced to 2.5 times the standard deviation;

5. Repeat steps 1 to 4 above for every observation in the sample.

An LSTAR-GARCH model, as defined in equations (2.8) and (2.13), is estimated using both the adjusted and unadjusted S&P, Heng Seng and Nikkei data. The Newton algorithm is used in each case. The estimates can be found in the following Table 3.

Table 3: LSTAR-GARCH Estimates for Adjusted and Unadjusted Data

	S&P		Heng Seng		Nikkei	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
$\hat{\phi_{11}}$	1.7833**	-1.7092**	-0.2238**	-0.5797**	-0.2796**	-0.6837**
$\hat{\phi_{12}}$	1.1204**	2.1950**	2.0523**	-4.4924**	-0.1930**	0.8175**
$\hat{\phi_{21}}$	-0.2862**	1.5478**	0.2260**	0.6832**	0.0013**	0.3890**
$\hat{\phi_{22}}$	-1.3476**	-0.2446**	-1.8170**	-4.9560**	0.0126	0.1554**
$\hat{\gamma}$	2.4515**	2.9175**	4.3080**	15.6068**	1.6278**	1.0431**
$\hat{c}$	0.8634**	0.1849**	-0.0162	-0.1252**	-2.0104**	0.7372**
$\hat{\omega}$	1.17e-6	3.87e-7**	7.08e-6**	5.44e-6**	3.22e-6**	2.05e-6**
$\hat{\alpha}$	0.0753**	0.0378**	0.1488**	0.1022**	0.1475**	0.1017**
$\hat{\beta}$	0.9166**	0.9579**	0.8394**	0.8750**	0.8519**	0.8909**

As shown in Table 3, ϕ_{12} exceeds 1 using both the adjusted and unadjusted S&P data, which suggests that the first regime follows a non-stationary process. The same is also true for Heng Seng. However, as no results on non-stationary STAR-type models are available, it is difficult to interpret these estimates.

In all three cases, the values of $\hat{\alpha}$ decreased and those of $\hat{\beta}$ increased when the data were adjusted, which agrees with other findings in the literature

(see, for example, Verhoeven and McAleer (1999)). However, the effects of extreme observations and outliers on the estimates of STAR are not entirely clear, though it appears that, if such data have a positive (negative) effect on $\hat{c}$, they will have a negative (positive) effect on $\hat{\gamma}$. This is an unusual result as there is no obvious reason why the threshold value should be related to the transition rate. However, for S&P and Heng Seng, the threshold values are closer to 0 with the adjusted data, suggesting that the adjusted data exhibit asymmetric behaviour. Moreover, it appears that if $\hat{\phi_{12}}$ increases (decreases) after the data are adjusted, then $\hat{\phi_{22}}$ will also increase (decrease), which suggests that extreme observations and outliers have the same effects on the coefficients of y_{t-1} in both regimes.

A similar analysis is conducted to examine the effects of extreme observations and outliers for ESTAR-GARCH. The estimates are given in Table 4.

Table 4: ESTAR-GARCH Estimates for Adjusted and Unadjusted Data

	S&P		Heng Seng		Nikkei	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
$\hat{\phi_{11}}$	-0.7497**	-0.5490**	-0.9955**	-0.4233**	-0.4259**	-0.2144**
$\hat{\phi_{12}}$	0.0790	-2.3958**	-0.9307**	0.0004	0.1168**	0.6474**
$\hat{\phi_{21}}$	0.0008**	1.8336**	0.0013**	0.0013**	0.0011**	0.3901**
$\hat{\phi_{22}}$	0.0360	0.9974**	0.1269**	0.1126**	0.0137	0.1911**
$\hat{\gamma}$	0.9474**	0.7736**	0.7997**	1.2412**	1.6688**	0.8317**
$\hat{c}$	-3.0813**	-0.5821**	-4.3532**	-2.5382**	-2.1846**	0.7273**
$\hat{\omega}$	1.22e-6**	3.87e-7**	7.27e-7**	6.07e-6**	3.22e-6**	2.05e-6**
$\hat{\alpha}$	0.0767**	0.0378**	0.1506**	0.1076**	0.1475**	0.1017**
$\hat{\beta}$	0.9149**	0.9579**	0.8370**	0.8669**	0.8519**	0.8909**

The estimates of ESTAR-GARCH using the unadjusted data seem more plausible than those for LSTAR-GARCH in Table 3. All the regimes follow a stationary AR(1) process, but the low threshold values suggest that the second regime would dominate the first for all three indices. Furthermore, the estimates for the GARCH component are very similar to those for

LSTAR-GARCH and, for Nikkei, the GARCH estimates are identical! This suggests that the choice of transition function for the conditional mean does not affect the GARCH estimates, reflecting the block-diagonal nature of the information matrix for STAR-GARCH models.

The effects of extreme observations and outliers on the transition rate, $\hat{\gamma}$, and the threshold value, $\hat{c}$, for ESTAR-GARCH seem to be different from LSTAR-GARCH. In particular the inverse relationship between $\hat{\gamma}$ and $\hat{c}$ is no longer valid. However, the estimated threshold values increased when the adjusted data were used. For all three data sets, the estimated threshold values using the adjusted data were closer to 0 than for the unadjusted data. Furthermore, the sign of the estimated threshold value changed from negative to positive for Nikkei.

However, the effects of extreme observations and outliers on the estimates of the transition rates are unclear. The effects of such data on $\hat{\phi_{12}}$ and $\hat{\phi_{22}}$ for ESTAR-GARCH are different from LSTAR-GARCH. In this case, if extreme observations and outliers have positive (negative) effects on $\hat{\phi_{12}}$, they have negative (positive) effects on $\hat{\phi_{22}}$ for both S&P and Heng Seng. However, extreme observations and outliers have negative effects on both $\hat{\phi_{12}}$ and $\hat{\phi_{22}}$ for Nikkei.

There does not seem to be a clear pattern between the estimates using the adjusted and unadjusted data. However, due to the increase in the threshold value, the second regime no longer dominates the first, so that the adjusted data exhibit regime switching behaviour. Empirical evidence suggests that the estimate of the threshold value is sensitive to the sign and magnitude of outliers. If the magnitude of positive outliers is greater (smaller) than their negative counterparts, the threshold estimate will be greater (smaller) than 0. This result explains the low threshold estimates, because the magnitude of the negative outliers is often larger than the positive outliers for all three indices, and also the increase in the threshold estimates when the magnitude of the outliers is reduced.

These results also show that the effects of outliers and extreme observations on the estimates of STAR-GARCH are sensitive to the choice of transition function. Moreover, the convergence of the estimates also seems

to be sensitive to the choice of transition function for STAR-STGARCH. The choice of transition function, and their convergence of the algorithm, are summarised in Table 5.

Table 5: Convergence with Different Transition Functions

Transition function		Unadjusted data		
mean	variance	S&P	Heng Seng	Nikkei
Logistic	Logistic	Yes	Yes	No
Logistic	Exponential	No	Yes	No
Exponential	Logistic	Yes	Yes	Yes
Exponential	Exponential	No	No	No

As Table 5 shows, the exponential/logistic combination converges for all three indices. The estimates for the three data sets are given in Table 6.

The estimates of the extreme threshold values, which indicate that one regime dominates the other, suggest that two regimes are unnecessary for either the conditional mean or the conditional variance for S&P and Nikkei. For the same reason, two regimes for the conditional variance are also unnecessary for Heng Seng. However, the estimated threshold value of the conditional mean, $\hat{c}$, for Heng Seng is close to 0, which indicates the data exhibit asymmetric behaviour and two regimes are present for the conditional mean.

Although all the estimates are highly significant, this is based on the assumption of asymptotic normality, which cannot be checked as no regularity conditions are available. Moreover, the estimates from a single regime are similar to those using an AR(1)-GARCH(1,1) model for S&P and Nikkei, which reinforces the conclusion that two regimes are not present.

Table 6: ESTAR-LSTARGARCH Estimates for S&P, Heng Seng and Nikkei

	S&P	Heng Seng	Nikkei
$\hat{\phi_{11}}$	-0.4167**	-0.1124**	1.5339**
$\hat{\phi_{12}}$	0.0862**	-0.6694**	0.0991**
$\hat{\phi_{21}}$	0.0007**	0.8880**	-0.7509**
$\hat{\phi_{22}}$	0.0359**	-1.8741**	1.2881**
$\hat{\gamma}$	1.5140**	2.3182**	0.3103**
$\hat{c}$	-2.3757**	-0.2280**	-1.8928**
$\hat{\omega_1}$	1.184e-06**	6.88e-6**	3.15e-06**
$\hat{\alpha_1}$	0.7662**	0.1488**	0.1465**
$\hat{\beta_1}$	0.9149**	0.8394**	0.8529**
$\hat{\omega_2}$	4.362e-03**	1.24e-3**	3.896e-03**
$\hat{\alpha_2}$	0.1186**	0.0170**	0.0331**
$\hat{\beta_2}$	0.00008464	0.0096	0.035**
$\hat{\xi}$	3.1153**	2.7183**	3.1029**
$\hat{d}$	2.9595**	2.5338**	2.9643**

Estimates for ESTAR-LSTGARCH for two sets of adjusted data are given in Table 7, because the Newton method did not converge using the adjusted Heng Seng data. However, the estimates did converge using the BFGS algorithm, which supports the findings in Section 3 that the estimates are sensitive to the choice of algorithm.

It appears that extreme observations and outliers have little impact on $\hat{d}$, which suggests there is no regime switching behaviour in the GARCH component for either adjusted or unadjusted data. For each data set, only the first regime is required. Interestingly, $\hat{\alpha_1}$ decreased and $\hat{\beta_1}$ increased after the data were adjusted. The same outcome holds for $\hat{\alpha_2}$ and $\hat{\beta_2}$ for S&P, but not for Nikkei.

Furthermore, the effects of extreme observations and outliers on $\hat{\gamma}$ and $\hat{c}$ are also unclear. Both $\hat{\phi_{12}}$ and $\hat{\phi_{22}}$ decreased for S&P and Nikkei when the adjusted data were used, which is surprising as the effects of extreme observations and outliers for ESTAR-STGARCH are expected to be similar

to those for ESTAR-GARCH. This result may arise because the information matrix of STAR-STGARCH is no longer block diagonal with regard to the parameters of the conditional mean and the conditional variance.

For S&P, two regimes were required before the data were adjusted, but only the second regime is significant for the adjusted data. The estimated threshold value increased for Nikkei when the adjusted data were used, but the second regime was still insignificant.

Table 7: ESTAR-LSTGARCH Estimates for the Adjusted S&P and Nikkei

	S&P	Nikkei
$\hat{\phi_{11}}$	-0.0590**	-0.7316**
$\hat{\phi_{12}}$	-0.8052**	-0.2209**
$\hat{\phi_{21}}$	1.5970**	0.0013**
$\hat{\phi_{22}}$	0.0017	0.0108
$\hat{\gamma}$	1.7790**	1.5160**
$\hat{c}$	-0.1433**	-2.1709**
$\hat{\omega_1}$	3.75e-7**	2.00e-6**
$\hat{\alpha_1}$	0.0378**	0.1021**
$\hat{\beta_1}$	0.9579**	0.8906**
$\hat{\omega_2}$	1.10e-3**	2.38e-3**
$\hat{\alpha_2}$	0.0363**	0.0687**
$\hat{\beta_2}$	0.0154**	0.0080**
$\hat{\xi}$	2.9537**	2.9638**
$\hat{d}$	2.7692**	2.8433**

6   Conclusion

This paper provided a survey of recent developments for analysing the GARCH, STAR, STAR-GARCH and STAR-STGARCH models. The difficulties in evaluating these models because of the absence of structural and statistical properties, particularly, the regularity conditions for consistency and asymptotic normality, were emphasized.

Empirical evidence using the S&P, Heng Seng and Nikkei indexes showed that the QMLE for STAR-GARCH models are sensitive to the choice of optimisation algorithm. This does not agree with previous results in the literature. It was also shown that the estimates for STAR-GARCH and STAR-STGARCH are also highly sensitive to extreme observations and outliers. Furthermore, the effects of extreme observations and outliers on the estimates of STAR-GARCH depend on the choice of transition function.

The effects of extreme observations and outliers on the estimates for STAR-STGARCH are unclear, but the effects are not the same as for STAR-GARCH, which may arise because the information matrix is no longer block diagonal. Furthermore, the convergence of the estimates is sensitive to the choice of algorithm. This sensitivity could arise through model misspecification, as well as through the properties of the log-likelihood functions.

References

[1] 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.

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

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

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

[5] C.H. Chen, editor. *Pattern Recognition and Signal Processing*. Amsterdam: Sijhoff and Noordoff, 1978.

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

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

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

[9] D. van Dijk, T. Teräsvirta, and P.H. Franses. Time Varying Smooth Transition Autoregressive Models. Submitted. 2000b.

[10] 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.

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

[12] R.F. Engle and D.L. McFadden, editors. *Aspects of Modelling Nonlinear Time Series*. Amsterdam: Elsevier Science, 1994.

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

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

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

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

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

[18] 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.

[19] 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.

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

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

[22] D.G. Luenberger. *Linear and Non-linear Programming*. Addison-Wesley Publishing Company, 1989.

[23] 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.

[24] 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.

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

[26] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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