FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Estimation of STAR-GARCH Models with Iteratively Weighted Least Squares

Murat Midilic

January 2016
2016/918

D/2016/7012/02

Estimation of STAR-GARCH Models with Iteratively Weighted Least Squares

Murat Midilic *

Ghent University

January 15, 2016

Abstract

This study applies the Iteratively Weighted Least Squares (IWLS) algorithm to a Smooth Transition Autoregressive (STAR) model with conditional variance. Monte Carlo simulations are performed to measure the performance of the algorithm, to compare its performance with the performances of established methods in the literature, and to see the effect of initial value selection method. Simulation results show that low bias and mean squared error are received for the slope parameter estimator from the IWLS algorithm when the real value of the slope parameter is low. In an empirical illustration, STAR-GARCH model is used to forecast daily US Dollar/Australian Dollar and FTSE Small Cap index returns. 1-day ahead out-of-sample forecast results show that forecast performance of the STAR-GARCH model improves with the IWLS algorithm and the model performs better that the benchmark model.

Keywords: STAR, GARCH, iteratively weighted least squares, Australian Dollar, FTSE

JEL classification: C15, C51, C53 C58, C87, F31

---

*Department of Financial Economics, Ghent University, Sint-Pietersplein 5, 9000 Ghent, Belgium. Tel.: +32 9 264 89 78; fax: +32 9 264 89 95; e-mail address: murat.midilic@ugent.be. The study was partly written during the author's stay at the Center for Research in Econometric Analysis of Time Series, CREATES, Aarhus University. I would like to thank Timo Teräsvirta, Cristina Amado, Michael Frömmel, Joris Wauters, Xing Han, Huichou Huang, and participants of the 2015 NBER-NSF Time Series Conference, Vienna, and International Conference on Computational and Financial Econometrics 2015, London, for their comments. I would like thank CREATES for their hospitality during my stay. This research was supported by the Postgraduate Scholarship of the Ministry of National Education of the Republic of Turkey, and the research stay was supported by the Faculty of Economics and Business Administration, Ghent University. All the remaining errors are mine.

1. Introduction

Nonlinear models with heteroskedastic variance have been used in analysing financial time series data for a long time. One special type of models under this family is the Smooth Transition Autoregressive (STAR) model with conditional variance. Unlike STAR models, which are estimated by nonlinear least squares (NLS), these models are estimated by maximum likelihood estimation (MLE), or quasi-maximum likelihood estimation (QMLE). However, in real world applications the numerical features of these models and computational aspects of the method used may lead to a complicated estimation procedure. For both NLS and MLE, the experience shows that the final estimates are highly dependent on the starting values (for STAR models, see Teräsvirta (1994), for STAR models with heteroskedastic variance, see Lundbergh et al. (1999) and Dijk et al. (2002)), the bias of the slope parameter estimator is higher than the biases of other parameter estimators, and estimation of the slope parameter gets more difficult as the parameter value gets higher. This study presents a potential solution to these problems by using Iteratively Weighted Least Squares (IWLS) and compares its performance with other established algorithms.

As pointed out by Teräsvirta (1994), Lundbergh et al. (1999), Dijk et al. (2002); one problem with the estimation of STAR type models is to find sensible starting points for the algorithm. Due to the highly nonlinear nature of STAR type models, estimation results are sensitive to starting points (Chan and McAleer, 2003). Moreover, according to Teräsvirta (1994) and Chan and McAleer (2002), the degree of difficulty in estimation differs depending on the type of the transition function used in the model. A comparison of the models with logistic transition function and exponential transition function shows that STAR models with logistic transition function proved to be more problematic than the other.

Simulation studies of Chan and Theoharakis (2011) show that the main reason for these problems seem to be the complex nature of the log-likelihoods of these functions stemmed by slope parameter(s). For STAR-GARCH, Chan and Theoharakis (2011) show how the log-likelihood function behave around optimum values of the parameters. They illustrate graphically that the log-likelihood function might be flat for exponential transition functions or lumpy for logistic transition functions around the optimum value of the slope parameter(s). These results can be regarded as supporting evidence for solution methods that are robust to local optima in estimating STAR-type models.

One common feature of the studies that use STAR models with conditional variance is that the models are estimated by maximum likelihood (ML).[1] Although, in theory ML

[1] Other solution methods are available for STAR-type models with homoskedastic errors. Teräsvirta

estimation is the correct way to handle the estimation problem and gives consistent estimates; from a numerical point of view, the algorithm used in solving the ML can be sensitive to starting values and might have poor performance in dealing with local optima issue; so the algorithm might be the reason for the problems described above. Therefore, an approach targeting robustness of the ML estimation of nonlinear models would be a potential solution to the referred problems. One such an approach is the IWLS estimation of nonlinear models with conditional variance. Mak (1993) and Mak et al. (1997) show that the maximum likelihood problems can be transformed into a problem that can be solved by IWLS, and the performance of the algorithm is compatible with or better than the traditionally used ML algorithms in the literature for the set of models we are interested in.

The purpose of this study is to show the performance of the IWLS algorithm in estimating STAR models with conditional variance in a basic setup. For this purpose, STAR-GARCH model is chosen to be the model used in the study. Monte Carlo (MC) simulations are carried with a STAR-GARCH model to show the performance of the algorithm conditional on different initial values. Robustness of the algorithm to initial value selection is checked by carrying simulations with randomly generated initial values and initial values provided by a heuristic algorithm, which has been used in the STAR model estimation literature. Brooks et al. (2001) show that there might be significant differences in the results from different software for the same problem. Therefore, the performance of IWLS is also compared with other functions that are commonly used for maximum likelihood and quasi-maximum likelihood estimations. These are *fmincon* function of MATLAB, and *maxLik* function of R. Special attention is paid to the slope and location parameters of the transition function since in practice, these parameters are found to be the most difficult to estimate in the literature. In order to account the effect of the dynamics in the variance component of the model, simulations are carried with several GARCH specifications. Practical implications of using the IWLS algorithm are studied with empirical applications.

The contribution of the study to the literature is twofold. First, in a basic setup, it shows that for the STAR-GARCH model, it is possible to have slope parameter estimators with smaller bias and variance with the IWLS algorithm. Second, as an empirical contribution, daily exchange rates and stock indices are forecasted by the STAR-GARCH model; thus, the study contributes to the exchange rate and stock index forecasting dis-

et al. (2010) discuss estimating the models by dividing the parameter vector into two subsets in order to reduce numerical burden. The method is first proposed by Leybourne et al. (1998). Even though it eases the numerical burden of the STAR model estimations, Maugeri (2014) show that there are cases in which the method gives "biased and inconsistent" results and maximum likelihood algorithms should be preferred.

cussions in the literature.

Simulation studies convey three key results. The first result is that when the real value of the slope parameter is low, IWLS performs better than other methods in estimating the parameter while IWLS does not perform worse in estimating the slope parameter when the real value of the parameter is high. The second result states that for the IWLS algorithm, bias of the slope parameter estimator from randomly generated initial values is smaller than the bias of the estimator received from the benchmark initial value selection method. According to the third result, estimation performances of the methods change as the real value of the persistency parameters in the variance equation change. In cases where performances of other methods deteriorate, IWLS performs better in estimating these parameters.

In the empirical part of the study, daily US Dollar (USD)/Australian Dollar (AUD) exchange rate and Financial Times Stock Exchange Small Cap (FTSE SC) returns are forecasted by using the STAR-GARCH model. According to the out-of-sample forecast error performance and prediction accuracy tests, the IWLS algorithm performs better than the benchmark random walk (RW) model as well as the competing algorithms. For the exchange rate forecasts, the statistical significance of the performance of the IWLS algorithm is robust to different predictive accuracy tests while robustness is not observed for the stock index forecasts, though most of the tests give significant results. Empirical exercises suggest that traditional methods used in the literature might give misleading results in the sense that even though the smooth transition model performs better than the benchmark model, the algorithms cannot demonstrate the performance. IWLS is shown to correct this problem in practice.

The study is organised as follows. The model and the IWLS algorithm are described in the next section. Section 3 presents the simulations, summarises results, and discusses their implications. In Section 4, the STAR-GARCH model is used in empirical illustrations to forecast daily exchange rate and stock index returns. The final section summarises and concludes.

2. STAR-GARCH Model and IWLS Estimation

To the best of our knowledge, STAR-type models have never been estimated by the IWLS algorithm in the literature; so the aim of the study is to first show its performance in a basic setup. STAR-GARCH model has been selected because of its simplicity within the available extensions of STAR-type models in the literature and it has a wide range of empirical applications that include estimation of daily and intra-daily S&P 500 index (Chan and McAleer, 2002, 2003), daily major exchange rates (Westerhoff and Reitz, 2003), and monthly commodity prices (Reitz and Westerhoff, 2007). The IWLS algorithm can be extended to other STAR-type models with conditional variance based on the potential success of the algorithm in this basic setup.

This section first describes the STAR-GARCH model and gives the non-trivial stationary conditions both for the mean and variance equations. Then, the IWLS algorithm is given.

2.1. STAR-GARCH Model

Consider a STAR(p) model with a GARCH(q_1,q_2) component and with $t = 1-p, 1-(p-1), ..., T-1, T$ as the following:

$$y_t = x_t\phi^{(1)} + x_t\phi^{(2)}G_t(.) + \varepsilon_t, \tag{1}$$

$$\varepsilon_t = v_t\sqrt{h_t},\ v_t \sim N(0,1), \tag{2}$$

$$h_t = \omega + \sum_{i=1}^{q_1}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q_2}\beta_j h_{t-j}, \tag{3}$$

where $x_t = (1, y_{t-1}, y_{t-2}, ..., y_{t-p})$ is a $1\times(p+1)$ vector; $\phi^{(1)} = (\phi_0^{(1)}, \phi_1^{(1)}, \phi_2^{(1)}, ..., \phi_p^{(1)})'$, and $\phi^{(2)} = (\phi_0^{(2)}, \phi_1^{(2)}, \phi_2^{(2)}, ..., \phi_p^{(2)})'$ are $(p+1)\times 1$ vectors of coefficients; $G_t(.)$ is the transition function; ω is the constant of the variance process; α_i is the ARCH coefficient for $i = 1, .., q1$; β_j is the GARCH coefficient for $j = 1, .., q2$, and h_t is the conditional variance. The errors in the mean equation, ε_t, are assumed to be orthogonal to x_t. The logistic transition function, $G_t(.)$, is given by:

$$G_t(y_{t-d}; e^{\eta}, c) = \frac{1}{1+e^{-e^{\eta}(y_{t-d}-c)}}, \tag{4}$$

where y_{t-d}, $d > 0$, is the transition variable, e^{η} is the slope parameter, and c is the location parameter. The slope parameter is expressed as an exponential function. In contrast to the traditional literature which gives the slope parameter as $\gamma > 0$; following Goodwin et al. (2011) and Hurn et al. (2014), the parameter is written as a monotonic transformation of γ. By doing so, the interval for the parameter of interest, η, can take values in the interval $(-\infty, \infty)$. In this way, one eliminates the non-negativity restriction on the parameter and the search for the slope parameter focuses on a smaller range of values because of the exponential mapping of η to γ (Hurn et al., 2014). The interval for $G_t(.)$ stays the same with the transformation: as $\eta \to -\infty$, $G_t(.) \to 0.5$ and the model approaches to a linear AR(p)-GARCH(q_1,q_2) model. As $\eta \to \infty$, $G_t(.) \to 1$ and the model approaches to a threshold AR (TAR(p))-GARCH(q_1,q_2) model. The corresponding log-likelihood function for observation t of the model is

$$l_t = -\frac{1}{2} lnh_t - \frac{\varepsilon_t^2}{2h_t}. \tag{5}$$

The model is assumed to be weakly stationary. For nonlinear models with heteroskedastic conditional error, Meitz and Saikkonen (2008) give stationary conditions and conclude the stationary conditions for the mean and the variance can be checked separately. According to Meitz and Saikkonen (2008), one of the following two conditions is sufficient for the mean to be stationary:

$$\sum_{i=1}^{p} max\left\{\left|\phi_i^{(1)}\right|, \left|\phi_i^{(1)} + \phi_i^{(2)}\right|\right\} < 1, \tag{6a}$$

$$\rho(\{\mathbf{A}_1, \mathbf{A}_2\}) < 1. \tag{6b}$$

In condition (6b), $\rho(\{\mathbf{A}_1, \mathbf{A}_2\})$ is the joint spectral radius of $\mathbf{A}_1 = \bar{\mathbf{A}}_p([\phi_1^{(1)} ... \phi_p^{(1)}]')$ and $\mathbf{A}_2 = \bar{\mathbf{A}}_p([\phi_1^{(1)} + \phi_1^{(2)} ... \phi_p^{(1)} + \phi_p^{(2)}]')$ where $\bar{\mathbf{A}}_p$ is defined as the following:

$$\bar{\mathbf{A}}_p(a) = \begin{bmatrix} a_1 & a_2 & ... & a_{p-1} & a_p \\ 1 & 0 & ... & 0 & 0 \\ 0 & 1 & ... & 0 & 0 \\ . & . & ... & . & . \\ . & . & ... & . & . \\ . & . & ... & . & . \\ 0 & 0 & ... & 1 & 0 \end{bmatrix}.$$

In the simulations, both stationarity conditions are checked. In practice, it is most convenient to use the condition (6a) since computation of the joint spectral density can be burdensome as p increases.

In the MC simulations, the conditional variance is assumed to follow a GARCH(1,1) model. The weak stationarity condition of the conditional variance is thus $\alpha_1 + \beta_1 < 1$. Non-negativity of the conditional variance requires $\omega > 0$, $\alpha_1 \geq 0$, and $\beta_1 \geq 0$. Finally, the conditional variance is assumed to be fourth moment stationary which, when the errors are normal, implies the condition $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$ (for the fourth moment conditions of a GARCH processes see He and Teräsvirta (1999a) and He and Teräsvirta (1999b)).

2.2. IWLS Estimation of the STAR-GARCH Model

IWLS is a robust regression algorithm that goes back to Beaton and Tukey (1974) and was first used for robust polynomial fitting. In order to illustrate the basic setup for IWLS assume that y is a $T \times 1$ vector of observations, $\mathbf{X}$ is a $T \times n$ matrix of regressors, β is an $n \times 1$ vector of parameters, and $\mathbf{W}$ is an $n \times n$ diagonal weight matrix. Then, parameter estimates at the r^{th} IWLS iteration at can be written as follows:

$$\hat{\beta}_r = \hat{\beta}_{r-1} + (\mathbf{X}'\mathbf{W}_r\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}_r(y - \mathbf{X}\hat{\beta}_{r-1})). \tag{7}$$

where the weight matrix $\mathbf{W}_r$ is a function of the residuals from the previous step. This is why IWLS is sometimes referred to as "iteratively *re*weighted least squares". It is possible to assign a functional structure to the weight matrix by using weight factors (see Beaton and Tukey (1974) for an example with biweight regression). A simple example of the weights is one in which the elements of $(y - \mathbf{X}\hat{\beta}_{r-1})/k$ (k is a scale parameter) form the main diagonal of $\mathbf{W}_r$ which itself is a diagonal matrix. The scale parameter k can be chosen beforehand depending on the data and the purposes of the analysis or can be determined in step $r-1$ to be used in step r. At the initial step, the sum of the weights can be equal to one or a unit weight can be assigned to each element. IWLS can be used with numerical algorithms such as Gauss-Newton and Levenberg-Marquardt, and it has a wide range of application areas besides econometrics such as sparse recovery, face recognition, and magnetic resonance imaging (see Holland and Welsch (1977), Chartrand and Yin (2008), Daubechies et al. (2010)).

Mak et al. (1997) show the superiority of IWLS over the BHHH algorithm in the estimation of an ARCH model. Li and Li (1996) estimate their DTARCH model with IWLS and show that it is faster than the BHHH algorithm. Biases of some parameter estimators are comparable with those of the BHHH estimators, whereas for some other parameters IWLS estimators are less biased of the two. Besides being faster, IWLS is less sensitive to the choice of starting values. Green (1984) notes that "General experience seems to be that choice of starting values for the parameter estimates is not particularly critical". These features of IWLS make it useful in the estimation of STAR models with heteroskedastic variance.

Based on the results of Mak (1993), Mak et al. (1997) derive an IWLS algorithm for a very general class of nonlinear models with heteroskedastic errors. For the clarity of the subsequent application of the algorithm to the STAR-GARCH model, the main results of Mak (1993) and Mak et al. (1997) are restated here.

Let y be a vector of observations with size T, θ be a q-dimensional vector of parameters, and $p(y, \theta)$ be the density function of y under θ. Let θ_0 be the true parameter vector and assume that θ_0 lies in an open parameter space $\Omega = \{\theta\} \subseteq \mathbb{R}^q$. Ω is assumed to contain an estimator of θ_0, $\hat{\theta}$, as a root of

$$f(y,\theta) = 0. \tag{8}$$

The core assumption of the setup is that an estimator of Equation (8) is unbiased, i.e. $E\{f(y,\theta)|\theta\} = 0$ where $E\{.|\theta\}$ denotes the expectation operator. The scale of $f(y,\theta)$ is assumed to be the order of $O_p(T^{-1/2})$. Mak (1993) defines the function $g(.)$ for any θ, $\tilde{\theta}$ as follows:

$$g(\tilde{\theta},\theta) = E\left\{f(y,\theta)|\tilde{\theta}\right\} = 0. \tag{9}$$

An inductive sequence $\left\{\theta_{(r)}\right\}_0^\infty$ is defined as follows. Assume $\theta_{(r)}$ is given. For large T, where $f(y,\theta) \simeq E\left\{f(y,\theta_{(r)})|\theta_0\right\}$, θ_0 can be approximated by a θ which equates the observed value of f with the expected value under θ such that

$$f(y,\theta_{(r)}) = E\left\{f(y,\theta_{(r)})|\theta_{(r+1)}\right\} = g(\theta_{(r+1)},\theta_{(r)}). \tag{10}$$

If we take $r \to \infty$ and use the condition of unbiasedness, we see that if $\left\{\theta_{(r)}\right\}_0^\infty$ converges, it will converge to a root of Equation (8).

Using this result, Mak (1993) reaches the following main results:

$$\left.\frac{\partial g(\tilde{\theta},\theta)}{\partial\tilde{\theta}}\right|_\theta = -E\left\{\left.\frac{\partial f(y,\theta)}{\partial\theta}\right|_\theta\right\}, \tag{11}$$

$$\left.\frac{\partial\psi}{\partial\theta}\right|_{\hat{\theta}} \to 0 \qquad as \qquad T \to \infty, \tag{12}$$

$$Pr(S_n) \to 1 \qquad as \qquad T \to \infty, \tag{13}$$

and

$$\theta_{(2)} - \hat{\theta} = o_p(T^{-1/2}), \tag{14}$$

where S_n is the event that $\left\{\theta_{(r)}\right\}_0^\infty$ converges, and $\psi : \mathbb{R}^q \to \mathbb{R}^q$ is an implicit function so that

$$f(y,\theta) = g\left\{\psi(\theta),\theta\right\}. \tag{15}$$

ψ relates $\theta_{(r)}$ to $\theta_{(r+1)}$ as $\psi(\theta_{(r)}) = \theta_{(r+1)}$. For maximum likelihood estimations in which $f(y,\theta)$ is a vector of partial derivatives and might be complicated for further differentiation, the result given by Equation (11) (*Lemma 1* in Mak (1993)) suggests an alternative method. According to Equation (12) (*Lemma 2* in Mak (1993)), as T gets larger; when evaluated at the real value of θ, change in the implicit function ψ with respect to θ will converge to zero. Result (13) (*Theorem 1* in Mak (1993)) denotes the convergence probability of the sequence $\{\theta_{(r)}\}_0^\infty$ will approach to 1 as T gets larger. Finally, result (14) (*Theorem 2* in Mak (1993)) shows that the proposed algorithm converges very fast just in 2 steps.

If Equation (15) does not have an explicit solution, Mak (1993)) suggests using the following linearisation:

$$g(\tilde{\theta},\theta)+\left[\frac{\partial g(\tilde{\theta},\theta)}{\partial\tilde{\theta}}\bigg|_{\tilde{\theta}=\theta}\right]'(\tilde{\theta}-\theta)=\left[\frac{\partial g(\tilde{\theta},\theta)}{\partial\tilde{\theta}}\bigg|_{\tilde{\theta}=\theta}\right]'(\tilde{\theta}-\theta)=f(y,\theta). \tag{16}$$

Based on these results, Mak et al. (1997) start building an IWLS algorithm for nonlinear models with heteroskedastic errors first by defining $f(y,\theta)$ as follows:

$$f(y,\theta)=\frac{\partial lnp(y,\theta)}{\partial\theta} \tag{17}$$

where $lnp(y,\theta)$ is the density of y. The nonlinear model which is used for deriving the results is given as the following:

$$y_t=\mu(z_t,y_{t-1},y_{t-2},...,y_{t-p},\theta)+\varepsilon_t, \tag{18}$$

where z_t is a vector of regressors and ε_t is conditionally normally distributed with mean 0, $E(\varepsilon_t)=0$, and variance given by

$$h_t=h(z_t,y_{t-1},y_{t-2},...,y_{t-p},\theta). \tag{19}$$

It has to be noted that the model specification is very general and for derivation of the further results the model is assumed to satisfy some regularity conditions. Even though such a general specification is feasible, there might be some practical concerns while implementing the method for estimating a specific nonlinear model.

Set $y=(y_T,y_{T-1},y_{T-2},...)$, then we have $f(y,\theta)$ after differentiating the log-likelihood function as the following:

$$f(y,\theta)=-\frac{1}{2}\sum\frac{\partial h_t}{\partial\theta}\left\{\frac{1}{h_t}-\frac{(y_t-\mu_t)^2}{h_t^2}\right\}+\sum\frac{\partial\mu_t}{\partial\theta}\frac{(y_t-\mu_t)}{h_t}. \tag{20}$$

From Equation (15), we have:

$$g(\tilde{\theta},\theta)=-\frac{1}{2}\sum\frac{\partial h_t}{\partial\theta}\left\{\frac{1}{h_t}-\frac{\tilde{h}_t+(\tilde{\mu}_t-\mu_t)^2}{h_t^2}\right\}+\sum\frac{\partial\mu_t}{\partial\theta}\frac{(\tilde{\mu}_t-\mu_t)}{h_t}, \tag{21}$$

where $\tilde{\mu}_t = \mu(z_t, y_{t-1}, y_{t-2}, ..., y_{t-p}, \tilde{\theta})$ and $\tilde{h}_t = h(z_t, y_{t-1}, y_{t-2}, ..., y_{t-p}, \tilde{\theta})$. According to (11), Fisher's information matrix can be derived first by differentiating $g(\tilde{\theta}, \theta)$ with respect to $\tilde{\theta}$, which gives

$$\frac{\partial g(\tilde{\theta}, \theta)}{\partial \tilde{\theta}} = -\frac{1}{2} \sum \left\{ -\frac{1}{h_t^2} \left(\frac{\partial \tilde{h}_t}{\partial \tilde{\theta}} \right) - \frac{2(\tilde{\mu}_t - \mu_t)}{h_t^2} \left(\frac{\partial \tilde{\mu}_t}{\partial \tilde{\theta}} \right) \right\} \left(\frac{\partial h_t}{\partial \theta} \right)' + \sum \frac{1}{h_t} \left(\frac{\partial \tilde{\mu}_t}{\partial \tilde{\theta}} \right) \left(\frac{\partial \mu_t}{\partial \theta} \right)'. \tag{22}$$

Therefore, at $\tilde{\theta} = \theta$, the Fisher's information matrix can be written as

$$\mathbf{I}(\theta) = \frac{1}{2} \sum \frac{1}{h_t^2} \left(\frac{\partial h_t}{\partial \theta} \right) \left(\frac{\partial h_t}{\partial \theta} \right)' + \sum \frac{1}{h_t} \left(\frac{\partial \mu_t}{\partial \theta} \right) \left(\frac{\partial \mu_t}{\partial \theta} \right)'. \tag{23}$$

The IWLS algorithm is derived by using the Equations (20)-(22) and (16). Inserting Equations (20)-(22) in (16), we have

$$\sum \frac{1}{2} \frac{1}{h_t^2} \frac{\partial h_t}{\partial \theta} \left\{ \left(\frac{\partial h_t}{\partial \theta} \right)' (\tilde{\theta} - \theta) + h_t - (y_t - \mu_t)^2 \right\} + \sum \frac{1}{h_t} \frac{\partial \mu_t}{\partial \theta} \left\{ \left(\frac{\partial \mu_t}{\partial \theta} \right)' (\tilde{\theta} - \theta) - (y_t - \mu_t) \right\} = 0. \tag{24}$$

The terms of (24) can be rearranged in order to have:

$$\sum W_{1t} z_{1t} \{ y_{1t} - z_{1t}' \tilde{\theta} \} + \sum W_{2t} z_{2t} \{ y_{2t} - z_{2t}' \tilde{\theta} \} = 0, \tag{25}$$

where

$$W_{1t} = \frac{1}{h_t}, \quad z_{1t} = \frac{\partial \mu_t}{\partial \theta}, \quad y_{1t} = \left(\frac{\partial \mu_t}{\partial \theta} \right)' \theta + (y_t - \mu_t),$$
$$W_{2t} = \frac{1}{h_t^2}, \quad z_{2t} = \frac{\partial h_t}{\partial \theta}, \quad y_{2t} = \left(\frac{\partial h_t}{\partial \theta} \right)' \theta - h_t + (y_t - \mu_t)^2.$$

The weights W_{1t} and W_{2t} are based on the conditional variance h_t. Intuitively, as the variance at a certain point in time increases, the weight assigned to that certain point will decrease.

At the $(r+1)$th step, $\theta_{(r+1)}$ is computed as the weighted least squares estimate:

$$\theta_{(r+1)} = \left\{ \sum_{m=1}^{2} \sum_{t} W_{mt} z_{mt} z_{mt}' \right\}^{-1} \left\{ \sum_{m=1}^{2} \sum_{t} W_{mt} z_{mt} y_{mt} \right\}, \tag{26}$$

where $m = 1, 2$ represent the mean and variance equations respectively, and θ in the definitions of W_{mt}, z_{mt}, and y_{mt} are replaced with $\theta_{(r)}$.

There are some advantages of the algorithm. It only uses the first derivatives of the log-likelihood function and this gives a computational advantage over the algorithms that also need second order derivatives. Second, see (14), the algorithm promises a fast convergence. Third, it is a robust algorithm. In this context, this feature of the algorithm is reflected on the weights that would be low for possible large shocks (i.e. possible outliers in the sample). Fourth, Green (1984) notes that initial values are not very critical. For the examples given in Mak (1993), initial values are found to be independent of the starting values. However, Green (1984) also notes cases in which the final IWLS estimates may depend on the starting values.

On the other hand, the behaviour of the algorithm is not tractable except for simple cases. According to Green (1984), the algorithm can be seen as a fixed point problem and if the algorithm converges, the results will be a solution to the likelihood equations.

3. Simulation Study

The aim of the simulations is to evaluate the performance of the IWLS algorithm in estimating the STAR-GARCH model, compare the performance of the algorithm with other functions/algorithms used in the literature, and define cases in which it performs better or worse, if any.

3.1. Simulation Design

The IWLS algorithm is compared with the *fmincon* function of MATLAB, and *maxLik* function of R. Estimation methods are compared in terms of mean, standard deviation, median, bias and mean squared error (MSE). Distributional properties of the estimates are also compared with several distribution comparison tests as robustness checks.

Before going further in describing the simulation setup and results, it is worth noting the similarities and differences between these methods to better comment on the results. *fmincon*[2] is a minimisation function of MATLAB that can handle linear and nonlinear constraints, equality and inequality constraints, and parameter boundaries. In ML estimations, one should insert the negative of the log likelihood function as the objective function. Interior-point, trust-region-reflective, sequential quadratic programming, and active-set algorithms are optional in the function. In the simulations, interior-point algorithm has been used with *fmincon* since active-set algorithm requires a user-supplied gradient, and trust-region-reflective and sequential quadratic programming algorithms are

[2] For a detailed description of the function and its features, please see http://www.mathworks.com/help/optim/ug/fmincon.html

not large scale algorithms, which would create problems for long time series. *maxLik*[3] function is the maximisation function of the maxLik package of R that is designed for ML applications. Newton Raphson, BHHH, BHGS, Simulated ANNealing, Conjugate Gradients, and Nelder-Mead algorithms are available as optional within the function. *maxLik* can also handle linear constraints and parameter boundaries. Inequality constraints are only allowed for the Nelder-Mead algorithm; so this algorithm has been used with *maxLik* during the study.

The IWLS algorithm cannot handle constraints or boundaries. Instead, following the recommendation of Mak et al. (1997), if a parameter value at step r happens to be out of bounds, it is replaced with a value that is close to the boundary.[4] This is done by using the smallest distance between two numbers that is defined by the software. For MATLAB, the smallest distance is identified as 2.2204×10^{-16} and when, for instance, a parameter value is above the boundary at a specific step r, then the value if replace by the boundary value minus 2.2204×10^{-16} to be used in the next step.[5]

Parameters of the simulated series are specified as the following:

$$y_t = 0 - 0.35y_{t-1} + 0.55y_{t-2} + (0.02 + 0.20y_{t-1} - 0.25y_{t-2})G(y_{t-1}; e^{\eta}, 0.02),$$
$$h_t = 0.001 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}.$$

This model parameters are used with two η specifications in order to observe the effect of the slope parameter on estimation performances. Selected η values are $ln5$ and $ln100$, which correspond to slope parameter values of 5 and 100 in the traditional notation in the literature and these are commonly used values to show behaviours of the STAR models for different slope parameters.

This parameterisation is used with 9 different GARCH parameters starting with $(\alpha, \beta) = (0.09, 0.90)$ and increase (decrease) β (α) by 0.01 until $(\alpha, \beta) = (0.01, 0.98)$ for each η values used. The purpose of this exercise is to compare estimation performances of the methods for low values of α since as it is pointed out by Zivot (2009) for GARCH estimations, when some parameters are close to be unidentified[6], maximum likelihood estimations may not be reliable.

[3] For a detailed description of the function and its features, please see http://www.http://cran.r-project.org/web/packages/maxLik/maxLik.pdf

[4] For their problem, Mak et al. (1997) replace with 0; but for the problem at hand, replacing with a value that is close to the boundary is seen to work better since value of 0 for some parameters might lead the function to be undefined. Handling the constraints and boundaries can be seen as the major disadvantage of IWLS against other algorithms.

[5] This procedure implies that the boundaries are defined as real boundary value $\pm 2.2204 \times 10^{-16}$ in practice; therefore, if a parameter estimate is found to be exactly at this value, then that estimation is discarded from the simulations.

[6] In the case of GARCH, β is unidentified when $\alpha = 0$.

Figure (1) depicts selected representative series from the simulated data. The figure includes plots for y_t and h_t against time, and G_t against the transition variable y_{t-1}. As it can be seen on the plots of the transition variables, when the value of the slope parameter is low, transition from one state to another state is smoother. In addition to that, these figures show the effect of different GARCH specifications. For higher values of α, the volatility clusters in the data are more apparent.

A crucial part of estimating STAR-GARCH (or STAR type models) is the sensitivity of the final estimates to the starting points. For IWLS algorithm; Green (1984), Mak et al. (1997), and Li and Li (1996) state that initial points are not found to be important. The relevance of this claim for the STAR-GARCH model is tested by comparing estimates from two different sets of initial values: estimates of the Simulated Annealing (SA) algorithm and randomly generated initial values.

The SA algorithm is a heuristic algorithm that can be used for optimisation when traditional algorithms fail to converge or to get values close to the global maximum when there is local maxima problem in the function of interest. Given the initial temperature, maximum temperature, temperature reduction function, acceptance criteria, and stopping criterion[7](Goffe et al., 1994; Brooks and Morgan, 1995); the SA algorithm can be sketched in five steps: (i) generate a random point calculate the loss function for this point (e.g. log-likelihood function), (ii) select another point and decide if this point stays or not based on an acceptance criterion, (iii) decrease the temperature based on the temperature reduction function, (iv) reanneal, and (v) end the algorithm when the stopping criterion is satisfied. The acceptance and temperature reduction functions used in the study can be given as follows:

$$\text{acceptance} = e^{\frac{Loss_{new} - Loss_{old}}{\text{Temp}}}, \tag{27}$$

$$\text{Temp} = \text{Temp}_0 * 0.95^{n_{SA}}, \tag{28}$$

where acceptance is the acceptance probability, $Loss_{new}$ and $Loss_{old}$ denote loss function values for the old and new points to be decided on, $n_{SA} = 1, \ldots, N_{SA}$ is the number of annealing, Temp_0 is the starting temperature, Temp is the new temperature at annealing number n_{SA}.

Goffe et al. (1994) compare the SA algorithm with a simplex algorithm, a conjugate gradient algorithm, and a quasi-Newton algorithm and show that the SA algorithm performs better than the others. The performance of the algorithm is supported by Brooks

[7] The terminology of the SA algorithm comes from the physics and material sciences literature. In this context, we can consider temperature only as a parameter of acceptance. High temperature leads to high acceptance probability. Temperature is systematically decreased with the temperature reduction function; so that the probability of accepting a point as the optimum gets lower.

and Morgan (1995) who measure the performance of the algorithm with a problem in which maximum likelihood estimations fail to converge. It is also noted that the SA algorithm might fail to find the global optimum when the problem at hand is very complex (Goffe et al., 1994).[8,9] In the simulations of this study, the SA algorithm is started by the randomly generated initial values which are then also used to estimate the model.

A second set of initial values randomly generated with respect to the stationary conditions given for the mean and variance of the STAR-GARCH model. For the slope parameter, initial values are chosen randomly in the interval $[1,4]$, and for the location parameter, initial values are chosen to be the mean of the series that are estimated.

The number of replications for each study is 1000. The parameter bias and MSE of the estimations are calculated as follows:

$$\text{bias} = (\sum_{n=1}^{1000} (\hat{\theta}_n - \theta))/1000,$$

$$\text{MSE} = (\sum_{n=1}^{1000} (\hat{\theta}_n - \theta)^2)/1000,$$

where $\hat{\theta}_n$ is the parameter estimates from the n^{th} estimation.

3.2. Simulation Results

Tables 1-4 give mean, median, standard deviation, bias, and MSE for the parameter estimates for simulations with GARCH specifications $(\alpha, \beta) = \{(0.09, 0.90), (0.01, 0.98)\}$ of both low and high values of the slope parameter. Results from simulations with other GARCH specifications are similar to the ones reported here.[10] The tables also include average iteration numbers, which show that on average the IWLS algorithm has lower number of iterations than other methods. Figure 2 shows the fast convergence of the IWLS algorithm at the first two steps which justifies the theory of Mak (1993).

A number of observations on the simulations results can be listed. First, *maxLik* performs poorly compared to other methods. The bias and MSE of mean and variance parameters are larger than others with the *maxLik* function for all GARCH and slope parameter specifications. This function performs well only in the estimation of the slope and location parameters that in one case is the best of all three methods.

Second; in the estimation of the mean parameters except the slope and location pa-

8 For an application of the algorithm and other heuristic algorithms for STAR type models, please see Schleer (2015). For other applications of the SA algorithm, please see Nakatsuma (2000) and Bernard et al. (2012)

9 Despite the desirable features of the SA, it is not chosen to be the estimation algorithm because as the parameter number increases, the algorithm gets slower.

10 These results are available upon request.

rameters, the IWLS algorithm and *fmincon* perform similarly. None of the methods systematically outperforms the other with respect to bias and MSE.

The performance of the methods in estimating the slope parameter gets worse when the value of the parameter is increased. The IWLS algorithm performs best in estimating the slope parameter when the real value of the parameter is low. The performance of the algorithm is better with randomly generated initial values and the bias drops to 0.003 with the $(\alpha, \beta) = (0.09, 0.90)$ specification. This performance of the algorithm is beaten by the *maxLik* function only for two GARCH specifications with SA initial values and sometimes this function performs close to the IWLS; but the performance of the *maxLik* is shadowed by its comparatively lower performance in estimating other mean parameters. On the contrary; when the slope parameter is high, none of the methods shows significantly better performance in estimating the slope parameter. In this case, *fmincon* performs best and the IWLS algorithm is the second best. Figures 3 and 4 plot kernel distribution of the slope parameter estimates. As these plots show, the IWLS estimates are clustered around the real value of the parameter when $\eta = ln5$, while the distribution of *fmincon* estimates has a fatter right tail. When $\eta = ln100$, the differences between the distributions become more pronounced. Estimates from *fmincon* are clustered closer to $ln100$; but still all methods give high estimator bias.

Significance of the differences between slope parameter estimates are further scrutinised by nonparametric distribution comparison tests. Differences are analysed with two-sample Kolmogorov-Smirnov (KS) (Kolmogorov, 1933; Smirnov, 1948), Ansari-Bradley (AB) (Ansari et al., 1960), Wilcoxon rank sum (RS) (Wilcoxon, 1945), and Aslan-Zech (AS) (Aslan and Zech, 2005) tests. The null hypothesis of KS and AZ tests is that the two same samples come from the same continuous distribution and the alternative is that they do not. In the AB test, the null hypothesis is that the samples come from the same distribution against the alternative that they have the same mean but different variances. When medians of two samples considered are not equal, Ansari et al. (1960) suggest subtracting the medians before the test. The null hypothesis of the RS test is that the samples are from distributions with equal medians and the alternative is that they are not. The tests are used to compare IWLS estimates with estimates from other methods both for SA and random initial values, and for comparing estimates from SA initial values and random initial values for each method.

For the cases reported in the study, results for the distribution comparison tests are given in Tables (5)-(8). The tables first list results for comparisons of the IWLS estimates with estimates from other methods and then results for the comparison of estimates with the SA initial values and randomly generated initial values. According to the p-values of the tests, the null hypotheses can be rejected consistently for the comparison of IWLS with *fmincon* when the slope parameter is low. In this case, difference of IWLS estimates from *maxLik* estimates cannot be significantly rejected only with the RS test for the $(\alpha, \beta) = (0.01, 0.98)$ specification and SA initial values. When the slope parameter is high, the null hypotheses of the test are significantly rejected most of the time but the evidence against the difference between the distributional properties of the slope parameter estimators from IWLS and other methods is not robust. Results for the comparison of SA and randomly generated initial values are function specific. For IWLS, the null hypotheses are consistently rejected only for the case $(\alpha, \beta) = (0.09, 0.90)$ and the slope parameter is high. For *fmincon*, they are consistently rejected when the slope parameter is low; and finally for *maxLik*, they are consistently rejected for all cases. Therefore, for the comparison of SA and random initial values, the results suggest that there is not significant evidence supporting the argument that they give significantly different results with IWLS; but with *maxLik* the difference is significant while for *fmincon*, the method for generating initial values matter only when the slope parameter is low.

For both low and high value specifications of the slope parameter, location parameter of the IWLS algorithm with the randomly generated initial values always has the smallest bias and MSE. Overall, the IWLS gives better estimates for this parameter.

Some differences can be observed in the GARCH persistency parameter estimations. As the β parameter increases, the bias of this parameter estimator increases with all methods. For the $(\alpha, \beta) = (0.01, 0.98)$ specification, the bias of β estimator is around -0.25 from *fmincon* for both slope parameter specifications and initial value generation processes. Nevertheless, the bias from the IWLS algorithm gets to 0.055 at most and MSE is the smallest with the IWLS algorithm.

Overall, the better performance of the IWLS algorithm in estimating the slope parameter, location parameter, and the β parameter suggest that this algorithm might provide better estimation results when the real slope parameter is low, and in all cases it should not do worse than other methods. On the method of initial value selection, results of distribution comparison tests suggest that estimates from random initial values and SA initial values are not significantly different from each other for the slope parameter, which is the parameter that is shown to be the most different within estimation methods. However, there are some caveats to be noted on these results. First, results are derived under the assumption of stationarity for both the mean and the variance equations. If there is significant evidence against stationarity, comparison results given here may not be reliable and since the IWLS algorithm is based on these regularity conditions, the algorithm should not be preferred. Second, the STAR component studied here is a basic model with only one slope and location parameter that correspond to a two-state model. The performance of the IWLS algorithm should also be investigated when the number of states increased.

Third, the current model does not assume any nonlinearities in the variance component; but there are models such as STAR-STGARCH of Lundbergh et al. (1999) who assume that there is a smooth transition component in the variance and the transition variable is the errors of the mean equation.

In order to clarify the implications of the simulation results in practice, the next section includes forecasting exercises with daily exchange rate and stock index returns. Since *maxLik* has a poor performance in the simulations, the empirical exercise does not use this function.

4. Empirical Application

In this part of the study, two daily series are used for forecasting. The section first describes the series used, then the model specification is given. Finally, results of the forecasting exercises and predictive accuracy tests are reported.

According to the most recent survey of the Bank for International settlements, USD and AUD is the fourth most traded currency pair globally with a share of 6.8% of the total $5.3 trillion volume per day (BIS, 2013). The top panel of Figure (5) shows the evolution of the exchange rate during the period under consideration, 26/12/2007-26/08/2015. USD appreciates agains AUD during the global financial crisis and returns back to its pre-crisis level after 2010. It drops to a lower level than the pre-crisis USD/AUD exchange rate. This depreciation period is common for many currencies against USD until 2013 due to the Quantitative Easing (QE) policy of the US Federal Reserve (FED). USD starts to appreciate again in 2013 after by the announcement of end of the QE policy of FED. The middle panel of the same figure gives the returns of the exchange rates, e_t, which is calculated as the following:

$$e_t = log(\frac{E_t}{E_{t-1}}), \tag{29}$$

where E_t denotes the level of the exchange rate at time t.

Exchange rate modelling has been one of the most controversial topics in the international economics and finance literature since the seminal work of Meese and Rogoff (1983) empirically showed that exchange rate models are outperformed by the RW model. Both univariate time series models and models with macroeconomic fundamentals have been the subject of the discussion. Most of the studies focus on the prediction of exchange rate levels and at the monthly frequency or lower.[11] Arguing modelling of exchange rates

[11] Recently, in a detailed literature review, Rossi (2013) shows that the performance of in-sample and out-of-sample predictive power of exchange rate models depend on the selection of the benchmark model, exchange rate model considered, selected variables, forecast horizon, and forecast accuracy

in general is beyond the scope of this study. The purpose of the exercise is merely to see the forecasting performance of the IWLS algorithm with respect to other estimation methods and a benchmark model.

The second series to be forecasted by the STAR-GARCH model is FTSE SC daily returns. FTSE SC consists of companies starting from the 351st to the 619th largest listed in the FTSE and these companies make around 2% of the market capitalisation. The motivation in using the index stems from the discussions on the relation between company size and expected returns of investments. Since the work of Banz (1981) who show small companies have higher returns than large ones, role of size have been discussed in the literature and have been relevant for portfolio selection. Recently, Fama and French (2014, 2015) argue the topic in the context of factor models and underline the role of size in asset pricing. An interesting stylised fact about the FTSE SC is volatility of the series that justifies the discussions in the literature. During the first years of the global financial crisis yearly return jumps to 40% in 2009–which was the highest yearly return in the last decade within FTSE 100, 250, and SC indices–from -61% in 2008. Top panel of Figure (6) shows the evolution of the stock index for the period between 26/12/2007 and 26/08/2015. The noted drop in the index in 2008 and rebound in 2009 can be observed in this panel. The daily return is calculated by Equation (29) and is plotted in the middle panel of Figure (6).

Descriptive statistics of both return series are given in Table (9). Normality hypothesis of the series is tested by the Jarque-Bera normality test, and the null of normal distribution is rejected at 1% significance level for both series.

The STAR-GARCH model specification procedure includes several steps and will follow Lundbergh et al. (1999) and Li and Li (1996) who use similar models. The modelling cycle starts with lag selection. In the second step, a linear model with the selected lag is fitted and the errors are tested for remaining autocorrelation with the ARCH-LM test. In the final step, STAR-type nonlinearity is tested and appropriate transition variable is selected, if any.

Typical financial data series include many data points and for these series, traditional lag selection criteria tend to give either high or no lag value at all. Rech et al. (2001) propose a lag selection procedure to solve this problem for series that might have nonlinearities. The method calculates traditional information criteria after running regressions with interactions of the considered variables for a selected polynomial degree. It is shown that as the degree of the polynomial increases, the reliability of the selection procedure increases. Rech et al. (2001) recommend running the regressions for several lag values, calculate SBIC or AIC and then select the appropriate lag. For the lag selection of return series, the maximum lag length and maximum polynomial level are set to 10 and 4

test used.

respectively. SBIC is used for lag selection. For the USD/AUD series, with polynomial levels 1 to 3, the lag is selected to be 3, but at the highest polynomial level, the optimum lag length is selected to be 2. Based on the theoretical results given in Rech et al. (2001), lag length of 2 is used in the study. The same selection procedure gives a lag length of 2 for the FTSE SC return series.

AR(2) model is fitted to the series and error terms from both regressions are tested for autocorrelation with the ARCH-LM test. The maximum lag number for the test is set to 8. According to the results given in Table (10), the null hypothesis of no autocorrelation is rejected at 1% significance level at any given lag for both error terms.

Then, the STAR-type nonlinearity test of Teräsvirta (1994) has been used to decide on the nonlinearity in the data and corresponding delay variable. The results of the tests are given at the bottom of Table (9) for lag values 1 and 2. Similar results are received from tests with both series. The p-values show that linearity is rejected for both lags and it is more significant for lag 2; so the second lag is chosen to be the transition variable for both USD/AUD and FTSE SC returns.

Considering the test results, a STAR-GARCH(2;1,1) model with $d = 2$ is fitted to the series. Estimation results with the full samples are given in Table (11) for USD/AUD and in Table (12) for FTSE SC. According to full sample estimation results, estimate for the slope parameter is the highest with the IWLS algorithm that is started with SA initial values for both series. Another observation on the results is that ARCH effect is higher in FTSE SC series.

For the forecasting exercise, the length of the initial sample is chosen to be 1500 and 500 1-day ahead forecasts are calculated. Mean forecasting error (MFE), mean square forecasting error (MSFE), mean absolute forecasting error (MAFE), *Theil's U* and Pesaran and Timmermann (1992) (PT) test of directional forecasting for all models and methods used in the exercise are given in Tables (13) and (14) for USD/AUD and FTSE SC respectively. RW with a drift is used as the benchmark model for comparisons and results of the models for MFE, MSFE, and MAFE are given with respect to the performance of the RW. *Theil's U* has been calculated as follows:

$$\textit{Theil's } U = \frac{\sqrt{\frac{1}{K}\sum_{k=1}^{K}(e_{L+k} - \hat{e}_{L+k})^2}}{\sqrt{\frac{1}{K}\sum_{k=1}^{K}(e_{L+k})^2} + \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{e}_{L+k})^2}} \tag{30}$$

where L is the size of the initial sample, $k = 1,..,K$ denotes the number of forecasts and K is the total number of forecasts (i.e. $K + L = T$), $\hat{e}_{L+k}$ is the forecast value of e_{L+k} at time $L + k$. The statistic gives a measure of the forecast performance corrected by the real values. The statistic takes values between 0 and 1 and as the performance of a model gets better, the statistic approaches to 0. Finally, PT is a nonparametric test statistic that measures the directional forecasting ability of a model. The null hypothesis of the PT test is that the model at hand is not able to forecast the direction of changes and the test statistic has a standard normal distribution.

Forecasts statistics for the conditional mean of USD/AUD show that the best performing method is the IWLS algorithm with random initial values. *Theil's U* statistics show that nonlinear the STAR-GARCH model performs better than the linear models irrespective of the method and initial value. The best performance is from *fmincon* function with SA initial values. However, IWLS algorithm with random initial values is the only method that beats the RW model for every metric. MSFE and MAFE of the method is slightly below the values from RW while every other method does worse than the benchmark model. For the PT, IWLS with random initial values has the highest value but it is not significantly able to predict the direction of changes in the data.

In case of FTSE SC conditional mean forecasts, the IWLS algorithm with random initial values again gives the best performance based on MSFE; but AR(2)-GARCH model also outperforms the RW model. Based on the *Theil's U*, IWLS with random initial values performs better than the AR(2)-GARCH model. For the directional forecasting, the models cannot give statistically significant results in this case either.

The significance of the difference between the RW and other models are further tested with the predictive ability tests. The first test to be used in the analysis is the Diebold et al. (1995) (DM) test of equal predictive accuracy. The test statistic of the DM test has a standard normal distribution. However, in cases where two nested models are compared with the test, it is shown that asymptotically the test statistic does not have a standard normal distribution and rejects null too often (McCracken, 2007). For comparison of forecasts from nested models, Clark and West (2007) (CW) propose a test that uses and "adjusted" MSFE term which is the sample average of the squares of the differences between two model forecasts. CW is the second test to compare predictive ability of forecasts.

As shown by Rossi (2013) for the case of exchange rates, results on the forecast performance of models might change depending the predictive ability tests used. In order to see the robustness of the DM and CW test results, predictive ability of the models are tested by the encompassing and mean squared error based test statistics.[12] These tests are ENC_t (Harvey et al., 1998), ENC_F (Clark and McCracken, 2001), MSE_t, and MSE_F (McCracken, 2007). Let $\hat{\varepsilon}_{0,t+1}$ denote the 1-day ahead forecast error from the benchmark model (i.e. RW) and $\hat{\varepsilon}_{1,t+1}$ denote the forecast error from the model to be compared. De-

[12] Corradi and Swanson (2002) propose a nonparametric test for comparison of from nonlinear models and linear models. The test is also recommended by Teräsvirta (2006) for STAR-type models. This test could also be used in checking the predictive accuracy of the models

fine $\hat{d}_{t+1} = \hat{\varepsilon}^2_{0,t+1} - \hat{\varepsilon}^2_{1,t+1}$, $\hat{c}_{t+h} = \hat{\varepsilon}_{0,t+1}(\hat{\varepsilon}_{0,t+1} - \hat{\varepsilon}_{1,t+h})$, and $\hat{\sigma}^2_1 = (K-1+1)\sum_{t=L}^{T-1} \hat{\varepsilon}^2_{1,t+1}$. Then the test statistics can be written as the following:

$$\text{ENC\_t} = \frac{K^{-1/2}\sum_{t=L}^{T-1} \hat{c}_{t+1}}{\hat{S}_{cc}^{1/2}}, \tag{31}$$

$$\text{ENC\_F} = \frac{\sum_{t=L}^{T-1} \hat{c}_{t+1}}{\hat{\sigma}^2_1}, \tag{32}$$

$$\text{MSE\_t} = \frac{K^{-1/2}\sum_{t=L}^{T-1} \hat{d}_{t+1}}{\hat{S}_{dd}^{1/2}}, \tag{33}$$

$$\text{MSE\_F} = \frac{\sum_{t=L}^{T-1} \hat{d}_{t+1}}{\hat{\sigma}^2_1}, \tag{34}$$

where $\hat{S}_{cc}$ and $\hat{S}_{dd}$ denote Newey and West (1987) heteroskedasticity and autocorrelation consistent (HAC) variance for $\hat{c}_{t+h}$ and $\hat{d}_{t+h}$ respectively. In cases where the assumptions of the test statistics are violated, Clark and McCracken (2010) propose to use a bootstrapping method in which "artificial" samples are used to calculate the test statistics. In order to create the artificial samples, first the benchmark model is estimated and the estimates are stored, then the forecast errors of the bigger model is fitted with an moving average model and the estimates of the moving average model is used to create artificial forecast errors by changing the moving average model with draws form standard normal distribution. Finally, the artificial forecast errors are added to the fitted values of the benchmark model to create the artificial samples. In this study, this procedure is repeated for 10,000 times to calculate the test statistics and the p-values for each is calculated.

Test statistics for the predictive ability test are given in Tables (15) and (16); and accompanying bootstrapped p-values for ENC_t, ENC_F, MSE_t, and MSE_F test statistics are given in Tables (20) and (21). For the case of the exchange rate series, DM test statistics suggest that *fmincon* forecasts with SA initial values are significantly different from the RW forecasts. The negative values of these test statistics imply that the performance of the RW model is better than the performance of these models. On the other hand, the DM test statistic is also negative for IWLS with SA initial values but not significant; and even though the DM test is not significant for the IWLS with random initial values, other test statistics suggest that IWLS forecasts with random initial values are significantly better than the RW forecasts. The significance level is lower with the ENC_F test which is, according to Busetti and Marcucci (2013), the most powerful of the given tests used here.

For the case of stock index series; according to the CW test, AR(2)-GARCH and IWLS with random initial values perform significantly better than the RW. However, in this case neither of the performances is robust to different tests. Performance of IWLS with random initial values is significantly better according to the MSE_t and MSE_F tests while the performance of the AR(2)-GARCH is significantly better according to ENC_t and MSE_t test statistics. ENC_F test does not report any significant result in this case.

Forecasts tests statistics for the variance component of the models are given in Tables (17) and (18) and predictive accuracy test statistics are given in Table (19). We do not have a nested model issue for the volatility forecasts; so only DM test is used for comparison of accuracy. In both cases, STAR-GARCH models cannot outperform the benchmark model based on the MSFE statistic and on the contrary to the conditional mean forecasts, results for the *Theil's U* statistics are mixed. DM test statistics suggest that in the case of USD/AUD, IWLS with SA initial values is the only method that does not significantly perform worse than the benchmark model; and in the case of FTSE SC, the only method that does not significantly perform worse than the benchmark is IWLS with random initial values. *fmincon* with both SA and random initial values perform significantly worse than the benchmark in both cases. Differences in the IWLS and *fmincon* suggest that, as shown in the simulation studies, estimation method has practical implications also for the conditional variance component.

Empirical studies show that the IWLS algorithm with random initial values give better forecasts than the benchmark model and more relevant to the argumentation of this study, it gives better forecast results than the *fmincon* function.

5. Conclusions

This study considers application of the IWLS algorithm to STAR-GARCH models in order to find a solution to initial value selection and slope parameter estimation in STAR-type models. Performance of the algorithm is compared with other methods in a simulation study for different values of slope parameter, GARCH persistency parameters, and initial value selection procedure. Real data application of the algorithm is shown in an empirical exercise with USD/AUD daily exchange rate and FTSE SC stock index returns.

Simulation studies show the cases when IWLS algorithm performs better than considered maximum likelihood estimation functions based on bias and MSE of the parameter estimators. MSE and bias of the slope parameter estimator are shown to be the lowest with the IWLS algorithm with randomly generated initial values when the real value of the slope parameter is low. When the real value of the slope parameter is high, MSE and bias of the slope parameter turns out to be high with all methods; thus, the IWLS algorithm does not deliver desired results in this case. In addition to that, for the GARCH parameter estimators, the IWLS algorithm give better results compared to other methods as the real value of the ARCH decreases. Practical implications of these results are shown with two empirical cases in which the IWLS algorithm significantly performs better than the benchmark model and *fmincon* function.

Another result received from the simulation study is that for the slope parameter estimator, initial value selection method delivers significantly different results for *fmincon* and *maxLik* but not for the IWLS algorithm. However, in the forecasting exercises, it is observed that the IWLS algorithm with randomly generated initial values outperforms IWLS with SA initial values, which might be taken as evidence for the use of random initial values with IWLS as it is argued in the literature.

The model used in this study is a basic one and there are some regularity conditions. Further research is needed to see the performance of the algorithm with more complex models such as STAR model with more than one location parameters or models that also include nonlinearities in the conditional variance. The algorithm can easily be extended to estimate such kind of models as long as their first order derivatives are provided.

In conclusion, the simulation study and empirical applications show that the IWLS algorithm can be chosen as an estimation method for STAR-type models with conditional variance by concerning less about initial value selection.

References

Ansari, A. R., Bradley, R. A., et al. (1960). Rank-sum tests for dispersions. *The Annals of Mathematical Statistics*, 31(4):1174–1189.

Aslan, B. and Zech, G. (2005). Statistical energy as a tool for binning-free, multivariate goodness-of-fit tests, two-sample comparison and unfolding. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 537(3):626–636.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3–18.

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16(2):147–185.

Bernard, J.-T., Dufour, J.-M., Khalaf, L., and Kichian, M. (2012). An identification-robust test for time-varying parameters in the dynamics of energy prices. *Journal of Applied Econometrics*, 27(4):603–624.

BIS (2013). Triennial central bank survey foreign exchange turnover in april 2013: preliminary global results. Technical report, Bank for International Settlements.

Brooks, C., Burke, S. P., and Persand, G. (2001). Benchmarks and the accuracy of garch model estimation. *International Journal of Forecasting*, 17(1):45–56.

Brooks, S. P. and Morgan, B. J. (1995). Optimization using simulated annealing. *The Statistician*, pages 241–257.

Busetti, F. and Marcucci, J. (2013). Comparing forecast accuracy: a monte carlo investigation. *International Journal of Forecasting*, 29(1):13–27.

Chan, F. and McAleer, M. (2002). Maximum likelihood estimation of STAR and STAR-GARCH models: Theory and Monte Carlo evidence. *Journal of Applied Econometrics*, 17(5):509–534.

Chan, F. and McAleer, M. (2003). Estimating smooth transition autoregressive models with GARCH errors in the presence of extreme observations and outliers. *Applied Financial Economics*, 13(8):581–592.

Chan, F. and Theoharakis, B. (2011). Estimating m-regimes STAR-GARCH model using QMLE with parameter transformation. *Mathematics and Computers in Simulation*, 81(7):1385–1396.

Chartrand, R. and Yin, W. (2008). Iteratively reweighted algorithms for compressive sensing. In *Acoustics, speech and signal processing, 2008. ICASSP 2008. IEEE international conference on*, pages 3869–3872. IEEE.

Clark, T. E. and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105(1):85–110.

Clark, T. E. and McCracken, M. W. (2010). Testing for unconditional predictive ability. *Federal Reserve Bank of St. Louis Working Paper No*.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Corradi, V. and Swanson, N. R. (2002). A consistent test for nonlinear out of sample predictive accuracy. *Journal of Econometrics*, 110(2):353–381.

Daubechies, I., DeVore, R., Fornasier, M., and Güntürk, C. S. (2010). Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38.

Diebold, F. X., Mariano, R. S., et al. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

Dijk, D. v., Teräsvirta, T., and Franses, P. H. (2002). Smooth transition autoregressive models—a survey of recent developments. *Econometric Reviews*, 21(1):1–47.

Fama, E. F. and French, K. R. (2014). Dissecting anomalies with a five-factor model. *Fama-Miller Working Paper*.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Goffe, W. L., Ferrier, G. D., and Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, 60(1):65–99.

Goodwin, B. K., Holt, M. T., and Prestemon, J. P. (2011). North American oriented strand board markets, arbitrage activity, and market price dynamics: a smooth transition approach. *American Journal of Agricultural Economics*, 93(4):993–1014.

Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society B*, 46(2):149–192.

Harvey, D. S., Leybourne, S. J., and Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business & Economic Statistics*, 16(2):254–259.

He, C. and Teräsvirta, T. (1999a). Fourth moment structure of the GARCH (p, q) process. *Econometric Theory*, 15(06):824–846.

He, C. and Teräsvirta, T. (1999b). Properties of moments of a family of GARCH processes. *Journal of Econometrics*, 92(1):173–192.

Holland, P. W. and Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics-Theory and Methods*, 6(9):813–827.

Hurn, A., Silvennoinen, A., and Teräsvirta, T. (2014). A smooth transition logit model of the effects of deregulation in the electricity market. Technical report, National Centre for Econometric Research.

Kolmogorov, A. (1933). Sulla determinazione empirica delle leggi di probabilita. *Giorn. Ist. Ital. Attuari*, 4:1–11.

Leybourne, S., Newbold, P., and Vougas, D. (1998). Unit roots and smooth transitions. *Journal of Time Series Analysis*, 19(1):83–97.

Li, C. and Li, W. (1996). On a double-threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics*, 11(3):253–274.

Lundbergh, S., Teräsvirta, T., et al. (1999). Modelling economic high-frequency time series. Technical report, Tinbergen Institute.

Mak, T., Wong, H., and Li, W. (1997). Estimation of nonlinear time series with conditional heteroscedastic variances by iteratively weighted least squares. *Computational Statistics & Data Analysis*, 24(2):169–178.

Mak, T. K. (1993). Solving non-linear estimation equations. *Journal of the Royal Statistical Society. Series B. Methodological*, 55(4):945–955.

Maugeri, N. (2014). Some pitfalls in smooth transition models estimation: A monte carlo study. *Computational Economics*, 44(3):339–378.

McCracken, M. W. (2007). Asymptotics for out of sample tests of granger causality. *Journal of Econometrics*, 140(2):719–752.

Meese, R. A. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14(1):3–24.

Meitz, M. and Saikkonen, P. (2008). Stability of nonlinear AR-GARCH models. *Journal of Time Series Analysis*, 29(3):453–475.

Nakatsuma, T. (2000). Bayesian analysis of ARMA–GARCH models: A markov chain sampling approach. *Journal of Econometrics*, 95(1):57–69.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica (1986-1998)*, 55(3):703.

Pesaran, M. H. and Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4):461–465.

Rech, G., Teräsvirta, T., and Tschernig, R. (2001). A simple variable selection technique for nonlinear models. *Communications in Statistics-Theory and Methods*, 30(6):1227–1241.

Reitz, S. and Westerhoff, F. (2007). Commodity price cycles and heterogeneous speculators: a star–garch model. *Empirical Economics*, 33(2):231–244.

Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, 51(4):1063–1119.

Schleer, F. (2015). Finding starting-values for the estimation of vector STAR models. *Econometrics*, 3(1):65–90.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, pages 279–281.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425):208–218.

Teräsvirta, T. (2006). Forecasting economic variables with nonlinear models. *Handbook of Economic Forecasting*, 1:413–457.

Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010). *Modelling nonlinear economic time series*. Oxford University Press Oxford.

Westerhoff, F. H. and Reitz, S. (2003). Nonlinearities and cyclical behavior: the role of chartists and fundamentalists. *Studies in Nonlinear Dynamics & Econometrics*, 7(4).

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, pages 80–83.

Zivot, E. (2009). Practical issues in the analysis of univariate GARCH models. In *Handbook of Financial Time Series*, pages 113–155. Springer.

Table 1: Monte Carlo simulation results ($(\alpha,\beta) = (0.09, 0.90)$ and $\eta = ln5$)

		SA			random		
		IWLS	fmincon	maxLik	IWLS	fmincon	maxLik
$\phi_0^{(1)}$	mean	0.030	0.025	0.150	0.022	0.025	0.130
	std.	0.085	0.064	0.192	0.074	0.066	0.271
	median	0.007	0.009	0.081	0.005	0.010	0.034
	bias	0.030	0.025	0.150	0.022	0.025	0.130
	MSE	0.008	0.005	0.059	0.006	0.005	0.090
$\phi_1^{(1)}$	mean	-0.305	-0.311	-0.159	-0.318	-0.311	-0.153
	std.	0.121	0.098	0.233	0.104	0.095	0.241
	median	-0.331	-0.327	-0.208	-0.337	-0.324	-0.221
	bias	0.045	0.039	0.191	0.032	0.039	0.197
	MSE	0.017	0.011	0.091	0.012	0.011	0.097
$\phi_2^{(1)}$	mean	0.533	0.525	0.503	0.530	0.518	0.391
	std.	0.071	0.061	0.183	0.066	0.064	0.266
	median	0.522	0.519	0.489	0.519	0.514	0.429
	bias	-0.017	-0.025	-0.047	-0.020	-0.032	-0.159
	MSE	0.005	0.004	0.036	0.005	0.005	0.096
$\phi_0^{(2)}$	mean	-0.015	-0.003	-0.199	-0.008	-0.001	-0.135
	std.	0.139	0.117	0.290	0.120	0.122	0.401
	median	0.002	0.002	-0.143	0.002	0.004	-0.105
	bias	-0.035	-0.023	-0.219	-0.028	-0.021	-0.155
	MSE	0.020	0.014	0.132	0.015	0.015	0.185
$\phi_1^{(2)}$	mean	0.162	0.159	0.093	0.172	0.156	0.032
	std.	0.113	0.101	0.277	0.100	0.101	0.351
	median	0.176	0.171	0.101	0.185	0.170	0.032
	bias	-0.038	-0.041	-0.107	-0.028	-0.044	-0.168
	MSE	0.014	0.012	0.088	0.011	0.012	0.152
$\phi_2^{(2)}$	mean	-0.228	-0.219	-0.181	-0.222	-0.214	-0.116
	std.	0.110	0.094	0.275	0.104	0.101	0.395
	median	-0.213	-0.207	-0.174	-0.209	-0.202	-0.140
	bias	0.022	0.031	0.069	0.028	0.036	0.134
	MSE	0.013	0.010	0.081	0.012	0.012	0.174
η	mean	1.527	1.863	1.397	1.613	2.038	2.546
	std.	1.415	1.377	1.405	1.355	1.435	1.198
	median	1.163	1.401	1.049	1.268	1.617	2.532
	bias	-0.082	0.254	-0.213	0.003	0.428	0.937
	MSE	2.008	1.958	2.016	1.835	2.242	2.310
c	mean	0.030	0.039	-0.019	0.029	0.067	0.118
	std.	0.197	0.216	0.439	0.177	0.270	0.717
	median	0.015	0.033	-0.071	0.027	0.052	0.125
	bias	0.010	0.019	-0.039	0.009	0.047	0.098
	MSE	0.039	0.047	0.194	0.031	0.075	0.523
ω	mean	0.002	0.001	0.005	0.002	0.001	0.015
	std.	0.008	0.001	0.011	0.005	0.001	0.039
	median	0.001	0.001	0.002	0.001	0.001	0.003
	bias	0.001	0.000	0.004	0.001	0.000	0.014
	MSE	0.000	0.000	0.000	0.000	0.000	0.002
α	mean	0.096	0.095	0.133	0.096	0.095	0.174
	std.	0.033	0.012	0.071	0.029	0.012	0.121
	median	0.092	0.095	0.111	0.092	0.095	0.132
	bias	0.006	0.005	0.043	0.006	0.005	0.084
	MSE	0.001	0.000	0.007	0.001	0.000	0.022
β	mean	0.892	0.888	0.806	0.892	0.888	0.678
	std.	0.036	0.015	0.179	0.029	0.015	0.288
	median	0.896	0.888	0.874	0.896	0.889	0.825
	bias	-0.008	-0.012	-0.094	-0.008	-0.012	-0.222
	MSE	0.001	0.000	0.041	0.001	0.000	0.132
	Aver. iterations	15.763	57.056	325.347	16.887	67.076	513.803

Monte Carlo simulation results for the STAR-GARCH model with $(\alpha,\beta) = (0.09, 0.90)$ and $\eta = ln5$. For each parameter; the mean, standard deviation, median, bias and MSE of estimates are given. Sample size is 2000 and the number of replications is 1000.

Table 2: Monte Carlo simulation results ($(\alpha,\beta) = (0.01, 0.98)$ and $\eta = ln5$)

		SA			random		
		IWLS	fmincon	maxLik	IWLS	fmincon	maxLik
$\phi_0^{(1)}$	mean	0.032	0.024	0.133	0.025	0.022	0.143
	std.	0.097	0.064	0.177	0.083	0.061	0.269
	median	0.007	0.009	0.060	0.003	0.009	0.045
	bias	0.032	0.024	0.133	0.025	0.022	0.143
	MSE	0.010	0.005	0.049	0.007	0.004	0.093
$\phi_1^{(1)}$	mean	-0.311	-0.319	-0.191	-0.320	-0.322	-0.162
	std.	0.115	0.094	0.211	0.111	0.092	0.241
	median	-0.337	-0.336	-0.244	-0.341	-0.336	-0.239
	bias	0.039	0.031	0.159	0.030	0.028	0.188
	MSE	0.015	0.010	0.070	0.013	0.009	0.093
$\phi_2^{(1)}$	mean	0.536	0.524	0.503	0.533	0.520	0.402
	std.	0.068	0.057	0.148	0.066	0.058	0.240
	median	0.525	0.517	0.488	0.522	0.513	0.434
	bias	-0.014	-0.026	-0.047	-0.017	-0.030	-0.148
	MSE	0.005	0.004	0.024	0.005	0.004	0.079
$\phi_0^{(2)}$	mean	-0.018	-0.003	-0.176	-0.011	-0.001	-0.133
	std.	0.148	0.116	0.267	0.132	0.110	0.378
	median	0.001	0.002	-0.138	0.004	0.001	-0.114
	bias	-0.038	-0.023	-0.196	-0.031	-0.021	-0.153
	MSE	0.023	0.014	0.109	0.018	0.013	0.166
$\phi_1^{(2)}$	mean	0.169	0.169	0.122	0.178	0.173	0.050
	std.	0.109	0.098	0.261	0.104	0.098	0.342
	median	0.183	0.180	0.139	0.184	0.185	0.068
	bias	-0.031	-0.031	-0.078	-0.022	-0.027	-0.150
	MSE	0.013	0.010	0.074	0.011	0.010	0.139
$\phi_2^{(2)}$	mean	-0.233	-0.220	-0.184	-0.228	-0.213	-0.078
	std.	0.099	0.086	0.218	0.095	0.088	0.341
	median	-0.219	-0.207	-0.173	-0.215	-0.198	-0.100
	bias	0.017	0.030	0.066	0.022	0.037	0.172
	MSE	0.010	0.008	0.052	0.010	0.009	0.146
η	mean	1.478	1.951	1.574	1.591	2.191	2.624
	std.	1.332	1.371	1.390	1.367	1.415	1.155
	median	1.117	1.516	1.324	1.213	1.947	2.586
	bias	-0.131	0.342	-0.036	-0.018	0.581	1.015
	MSE	1.790	1.995	1.931	1.866	2.338	2.363
c	mean	0.030	0.059	0.043	0.031	0.065	0.097
	std.	0.204	0.207	0.428	0.185	0.221	0.662
	median	0.025	0.048	0.046	0.029	0.056	0.106
	bias	0.010	0.039	0.023	0.011	0.045	0.077
	MSE	0.042	0.044	0.184	0.034	0.051	0.443
ω	mean	0.006	0.027	0.053	0.005	0.027	0.050
	std.	0.016	0.032	0.034	0.015	0.032	0.034
	median	0.001	0.004	0.058	0.001	0.003	0.052
	bias	0.005	0.026	0.052	0.004	0.026	0.049
	MSE	0.000	0.002	0.004	0.000	0.002	0.004
α	mean	0.023	0.034	0.061	0.024	0.033	0.140
	std.	0.021	0.013	0.061	0.033	0.013	0.145
	median	0.021	0.032	0.046	0.021	0.032	0.091
	bias	0.013	0.024	0.051	0.014	0.023	0.130
	MSE	0.001	0.001	0.006	0.001	0.001	0.038
β	mean	0.925	0.702	0.457	0.936	0.704	0.451
	std.	0.143	0.313	0.314	0.117	0.315	0.318
	median	0.965	0.931	0.407	0.965	0.932	0.416
	bias	-0.055	-0.278	-0.523	-0.044	-0.276	-0.529
	MSE	0.024	0.175	0.372	0.016	0.175	0.380
	Aver. iterations	24.642	58.851	250.351	27.381	71.213	418.890

Monte Carlo simulation results for the STAR-GARCH model with $(\alpha,\beta) = (0.01, 0.98)$ and $\eta = ln5$. For each parameter; the mean, standard deviation, median, bias and MSE of estimates are given. Sample size is 2000 and the number of replications is 1000.

Table 3: Monte Carlo simulation results ($(\alpha, \beta) = (0.09, 0.90)$ and $\eta = ln100$)

		SA			random		
		IWLS	fmincon	maxLik	IWLS	fmincon	maxLik
$\phi_0^{(1)}$	mean	0.007	0.005	0.145	0.006	0.008	0.119
	std.	0.037	0.022	0.173	0.030	0.034	0.278
	median	0.003	0.002	0.079	0.003	0.003	0.044
	bias	0.007	0.005	0.145	0.006	0.008	0.119
	MSE	0.001	0.001	0.051	0.001	0.001	0.091
$\phi_1^{(1)}$	mean	-0.333	-0.337	-0.161	-0.335	-0.334	-0.163
	std.	0.074	0.061	0.233	0.070	0.063	0.239
	median	-0.343	-0.342	-0.212	-0.343	-0.339	-0.223
	bias	0.017	0.013	0.189	0.015	0.016	0.187
	MSE	0.006	0.004	0.090	0.005	0.004	0.092
$\phi_2^{(1)}$	mean	0.552	0.550	0.538	0.553	0.544	0.409
	std.	0.048	0.038	0.176	0.043	0.050	0.274
	median	0.553	0.551	0.527	0.553	0.549	0.436
	bias	0.002	0.000	-0.012	0.003	-0.006	-0.141
	MSE	0.002	0.001	0.031	0.002	0.003	0.095
$\phi_0^{(2)}$	mean	0.016	0.020	-0.183	0.017	0.019	-0.099
	std.	0.060	0.039	0.278	0.047	0.055	0.412
	median	0.019	0.020	-0.131	0.019	0.020	-0.072
	bias	-0.004	0.000	-0.203	-0.003	-0.001	-0.119
	MSE	0.004	0.001	0.118	0.002	0.003	0.184
$\phi_1^{(2)}$	mean	0.174	0.177	0.085	0.178	0.172	0.023
	std.	0.082	0.070	0.274	0.073	0.078	0.339
	median	0.177	0.180	0.092	0.178	0.177	0.017
	bias	-0.026	-0.023	-0.115	-0.022	-0.028	-0.177
	MSE	0.007	0.005	0.088	0.006	0.007	0.146
$\phi_2^{(2)}$	mean	-0.265	-0.260	-0.253	-0.265	-0.254	-0.141
	std.	0.073	0.054	0.258	0.061	0.067	0.395
	median	-0.261	-0.259	-0.248	-0.261	-0.256	-0.167
	bias	-0.015	-0.010	-0.003	-0.015	-0.004	0.109
	MSE	0.006	0.003	0.067	0.004	0.005	0.168
η	mean	2.581	3.431	1.360	2.785	3.430	2.550
	std.	1.363	1.075	1.425	1.198	1.053	1.133
	median	2.526	3.572	0.975	2.720	3.553	2.561
	bias	-2.024	-1.174	-3.245	-1.820	-1.175	-2.055
	MSE	5.954	2.534	12.559	4.747	2.488	5.505
c	mean	0.026	0.028	0.012	0.023	0.055	0.094
	std.	0.098	0.085	0.422	0.061	0.209	0.711
	median	0.021	0.022	-0.019	0.024	0.023	0.084
	bias	0.006	0.008	-0.008	0.003	0.035	0.074
	MSE	0.010	0.007	0.178	0.004	0.045	0.511
ω	mean	0.001	0.001	0.005	0.001	0.001	0.014
	std.	0.005	0.001	0.011	0.004	0.001	0.037
	median	0.001	0.001	0.002	0.001	0.001	0.003
	bias	0.000	0.000	0.004	0.000	0.000	0.013
	MSE	0.000	0.000	0.000	0.000	0.000	0.002
α	mean	0.092	0.096	0.132	0.092	0.096	0.174
	std.	0.022	0.012	0.063	0.015	0.012	0.120
	median	0.091	0.095	0.110	0.091	0.095	0.134
	bias	0.002	0.006	0.042	0.002	0.006	0.084
	MSE	0.001	0.000	0.006	0.000	0.000	0.021
β	mean	0.895	0.887	0.806	0.895	0.888	0.684
	std.	0.024	0.015	0.176	0.016	0.015	0.281
	median	0.897	0.888	0.873	0.896	0.888	0.828
	bias	-0.005	-0.013	-0.094	-0.005	-0.012	-0.216
	MSE	0.001	0.000	0.040	0.000	0.000	0.125
	Aver. iterations	13.077	52.991	331.108	10.040	62.154	517.008

Monte Carlo simulation results for the STAR-GARCH model with $(\alpha, \beta) = (0.09, 0.90)$ and $\eta = ln100$. For each parameter; the mean, standard deviation, median, bias and MSE of estimates are given. Sample size is 2000 and the number of replications is 1000.

Table 4: Monte Carlo simulation results ($(\alpha, \beta) = (0.01, 0.98)$ and $\eta = ln100$)

		SA			random		
		IWLS	fmincon	maxLik	IWLS	fmincon	maxLik
$\phi_0^{(1)}$	mean	0.010	0.009	0.148	0.008	0.009	0.138
	std.	0.036	0.027	0.172	0.033	0.029	0.259
	median	0.006	0.006	0.076	0.004	0.005	0.055
	bias	0.010	0.009	0.148	0.008	0.009	0.138
	MSE	0.001	0.001	0.051	0.001	0.001	0.086
$\phi_1^{(1)}$	mean	-0.330	-0.331	-0.167	-0.333	-0.332	-0.167
	std.	0.066	0.058	0.217	0.064	0.057	0.252
	median	-0.337	-0.336	-0.229	-0.339	-0.336	-0.240
	bias	0.020	0.019	0.183	0.017	0.018	0.183
	MSE	0.005	0.004	0.080	0.004	0.004	0.097
$\phi_2^{(1)}$	mean	0.553	0.551	0.540	0.554	0.550	0.417
	std.	0.043	0.038	0.143	0.041	0.039	0.242
	median	0.552	0.552	0.540	0.554	0.552	0.446
	bias	0.003	0.001	-0.010	0.004	-0.000	-0.133
	MSE	0.002	0.001	0.021	0.002	0.002	0.076
$\phi_0^{(2)}$	mean	0.012	0.013	-0.180	0.013	0.014	-0.119
	std.	0.058	0.041	0.271	0.055	0.044	0.366
	median	0.015	0.015	-0.134	0.016	0.015	-0.084
	bias	-0.008	-0.007	-0.200	-0.007	-0.006	-0.139
	MSE	0.003	0.002	0.114	0.003	0.002	0.153
$\phi_1^{(2)}$	mean	0.176	0.176	0.087	0.179	0.176	0.033
	std.	0.075	0.070	0.238	0.074	0.071	0.332
	median	0.175	0.176	0.104	0.180	0.177	0.050
	bias	-0.024	-0.024	-0.113	-0.021	-0.024	-0.167
	MSE	0.006	0.005	0.069	0.006	0.006	0.138
$\phi_2^{(2)}$	mean	-0.265	-0.262	-0.247	-0.264	-0.259	-0.118
	std.	0.066	0.055	0.217	0.062	0.055	0.345
	median	-0.261	-0.260	-0.251	-0.263	-0.259	-0.160
	bias	-0.015	-0.012	0.003	-0.014	-0.009	0.132
	MSE	0.005	0.003	0.047	0.004	0.003	0.137
η	mean	2.623	3.305	1.472	2.674	3.353	2.659
	std.	1.383	1.144	1.330	1.312	1.124	1.158
	median	2.568	3.457	1.165	2.593	3.495	2.649
	bias	-1.983	-1.300	-3.134	-1.931	-1.253	-1.946
	MSE	5.841	2.998	11.587	5.451	2.831	5.126
c	mean	0.024	0.029	0.032	0.021	0.032	0.109
	std.	0.094	0.094	0.419	0.092	0.126	0.646
	median	0.021	0.022	0.003	0.022	0.022	0.066
	bias	0.004	0.009	0.012	0.001	0.012	0.089
	MSE	0.009	0.009	0.176	0.008	0.016	0.425
ω	mean	0.005	0.027	0.053	0.004	0.026	0.053
	std.	0.014	0.032	0.034	0.014	0.032	0.039
	median	0.001	0.004	0.057	0.001	0.004	0.053
	bias	0.004	0.026	0.052	0.003	0.025	0.052
	MSE	0.000	0.002	0.004	0.000	0.002	0.004
α	mean	0.022	0.034	0.058	0.022	0.033	0.141
	std.	0.013	0.013	0.051	0.014	0.013	0.148
	median	0.021	0.032	0.045	0.020	0.032	0.089
	bias	0.012	0.024	0.048	0.012	0.023	0.131
	MSE	0.000	0.001	0.005	0.000	0.001	0.039
β	mean	0.928	0.705	0.459	0.941	0.710	0.440
	std.	0.140	0.313	0.307	0.108	0.312	0.324
	median	0.965	0.931	0.416	0.965	0.932	0.385
	bias	-0.052	-0.275	-0.521	-0.039	-0.270	-0.540
	MSE	0.022	0.173	0.366	0.013	0.170	0.397
	Aver. iterations	23.220	55.510	265.642	21.220	67.647	416.994

Monte Carlo simulation results for the STAR-GARCH model with $(\alpha, \beta) = (0.01, 0.98)$ and $\eta = ln100$. For each parameter; the mean, standard deviation, median, bias and MSE of estimates are given. Sample size is 2000 and the number of replications is 1000.

Table 5: Distribution comparisons ($(\alpha, \beta) = (0.09, 0.90)$ and $\eta = ln5$)

	KS	AB	RS	AZ
IWLS random vs.				
fmincon	0.000	0.000	0.000	0.000
maxLik	0.000	0.000	0.000	0.000
IWLS SA vs.				
fmincon	0.000	0.000	0.000	0.000
maxLik	0.000	0.000	0.002	0.000
random vs. SA				
IWLS	0.075	0.360	0.023	0.077
fmincon	0.012	0.000	0.010	0.107
maxLik	0.000	0.002	0.000	0.000

P-values of the distribution comparison tests for the STAR-GARCH model with $(\alpha, \beta) = (0.09, 0.90)$ and $\eta = ln5$. The columns give results for two-sample Kolmogorov-Smirnov test (KS), Ansari-Bradley test (AB), Wilcoxon rank sum test (RS), and Aslan-Zech test (AZ). Null hypotheses of the tests are given in the text.

Table 6: Distribution comparisons ($(\alpha, \beta) = (0.01, 0.98)$ and $\eta = ln5$)

	KS	AB	RS	AZ
IWLS random vs.				
fmincon	0.000	0.000	0.000	0.000
maxLik	0.000	0.000	0.000	0.000
IWLS SA vs.				
fmincon	0.000	0.000	0.000	0.000
maxLik	0.000	0.000	0.491	0.000
random vs. SA				
IWLS	0.075	0.455	0.026	0.512
fmincon	0.001	0.000	0.000	0.001
maxLik	0.000	0.000	0.000	0.000

P-values of the distribution comparison tests for the STAR-GARCH model with $(\alpha, \beta) = (0.01, 0.98)$ and $\eta = ln5$. The columns give results for two-sample Kolmogorov-Smirnov test (KS), Ansari-Bradley test (AB), Wilcoxon rank sum test (RS), and Aslan-Zech test (AZ). Null hypotheses of the tests are given in the text.

Table 7: Distribution comparisons ($(\alpha, \beta) = (0.09, 0.90)$ and $\eta = ln100$)

	KS	AB	RS	AZ
IWLS random vs.				
fmincon	0.000	0.247	0.000	0.000
maxLik	0.014	0.686	0.000	0.004
IWLS SA vs.				
fmincon	0.000	0.018	0.000	0.000
maxLik	0.000	0.116	0.000	0.000
random vs. SA				
IWLS	0.001	0.090	0.000	0.001
fmincon	0.999	0.771	0.850	1.000
maxLik	0.000	0.000	0.000	0.000

P-values of the distribution comparison tests for the STAR-GARCH model with $(\alpha, \beta) = (0.09, 0.90)$ and $\eta = ln100$. The columns give results for two-sample Kolmogorov-Smirnov test (KS), Ansari-Bradley test (AB), Wilcoxon rank sum test (RS), and Aslan-Zech test (AZ). Null hypotheses of the tests are given in the text.

Table 8: Distribution comparisons ($(\alpha, \beta) = (0.01, 0.98)$ and $\eta = ln100$)

	KS	AB	RS	AZ
IWLS random vs.				
fmincon	0.000	0.311	0.000	0.000
maxLik	0.105	0.284	0.897	0.015
IWLS SA vs.				
fmincon	0.000	0.017	0.000	0.000
maxLik	0.000	0.760	0.000	0.000
random vs. SA				
IWLS	0.459	0.148	0.405	0.478
fmincon	0.855	0.954	0.403	1.000
maxLik	0.000	0.010	0.000	0.000

P-values of the distribution comparison tests for the STAR-GARCH model with $(\alpha, \beta) = (0.01, 0.98)$ and $\eta = ln100$. The columns give results for two-sample Kolmogorov-Smirnov test (KS), Ansari-Bradley test (AB), Wilcoxon rank sum test (RS), and Aslan-Zech test (AZ) respectively. Null hypotheses of the tests are given in the text.

Table 9: Descriptive statistics

	Min.	Max.	Mean	Var.	Skewness	Kurtosis	Normality
USD/AUD	-0.0670	0.0883	0.0001	0.0001	0.8623	15.5294	0.0001
FTSE SC	-0.0615	0.0377	0.0001	0.0001	-0.8860	9.1194	0.0001
	LM-test for nonlinearity of STAR type						
	d=1	d=2					
USD/AUD	1.62×10^{-7}	$9.78\times10^{-13*}$					
FTSE SC	0.097	$8.23\times10^{-6*}$					

Descriptive statistics for the daily USD/AUD returns. The column "Normality" gives the p-value for the Jarque-Bera normality test.

Table 10: ARCH-LM test

USD/AUD		FTSE SP	
statistic	p-value	statistic	p-value
53.05	3.24×10^{-13}	114.58	0
358.01	0	151.59	0
443.75	0	203.43	0
473.27	0	285.53	0
478.87	0	305.25	0
479.16	0	331.36	0
480.12	0	345.79	0
484.00	0	356.31	0

ARCH-LM test results for the AR(2) model errors

Table 11: Estimation results-USD/AUD

	RW	AR(2)	IWLS (SA)	fmincon (SA)	IWLS (R)	fmincon (R)
$\phi_0^{(1)}$	4.83×10^{-5}	4.99×10^{-5}	0.0001	0.0001	0.0005	4.79×10^{-5}
	(0.0001)	(0.0012)	(7.82×10^{-6})	(4.51×10^{-5})	(1.07×10^{-5})	(4.93×10^{-6})
$\phi_1^{(1)}$		-0.0224	-0.0103	0.0708	-0.0122	-0.0196
		(0.6679)	(0.0007)	(0.0015)	(0.0006)	(0.0005)
$\phi_2^{(1)}$		0.0001	-0.0281	0.0091	-0.0022	0.0089
		(0.7879	(0.0009)	(0.0021)	(0.0011)	(0.0006)
$\phi_0^{(2)}$			-6.88×10^{-5}	-6.83×10^{-5}	-0.0014	-0.0315
			(1.33×10^{-5})	(4.78×10^{-5})	(2.58×10^{-5})	(0.0002)
$\phi_1^{(2)}$			-0.0329	-0.1027	-0.0117	-0.1576
			(0.001)	(0.0018)	(0.0011)	(0.003)
$\phi_2^{(2)}$			0.1017	-0.0062	0.0947	0.5421
			(0.0013)	(0.0021)	(0.0016)	(0.004)
η			5.936	1.3532	2.042	1.7535
			(1.5383)	(0.067)	(0.057)	(0.0275)
c			0.0016	-0.0107	0.0041	0.0304
			(4.64×10^{-5})	(0.0002)	(9.24×10^{-5})	(6.50×10^{-5})
ω	3.91×10^{-7}	3.60×10^{-7}	3.81×10^{-7}	1.10×10^{-6}	3.78×10^{-7}	1.02×10^{-6}
	(1.24×10^{-5})	(1.91×10^{-5})	(1.05×10^{-7})	(1.57×10^{-7})	(1.04×10^{-7})	(1.53×10^{-7})
α	0.0768	0.0766	0.0756	0.0877	0.0741	0.0854
	(0.3889)	(0.5447)	(0.0012)	(0.0016)	(0.0012)	(0.0018)
β	0.9214	0.9221	0.9224	0.8982	0.9234	0.9018
	(0.0272)	(0.7891)	(0.0013)	(0.0021)	(0.0013)	(0.0022)
N. Ite.	17	24	7	28	10	57

Estimation results for the full sample of USD/AUD returns. Standard deviations are in the parentheses.

Table 12: Estimation results-FTSE SC

	RW	AR(2)	IWLS (SA)	fmincon (SA)	IWLS (R)	fmincon (R)
$\phi_0^{(1)}$	0.0005	0.0004	1.39×10^{-4}	-7.69×10^{-4}	-1.67×10^{-3}	-0.009
	(0.0001)	(0.0001)	(9.95×10^{-6})	(3.07×10^{-5})	(4.44×10^{-5})	(0.0001)
$\phi_1^{(1)}$		0.097	0.1469	0.1995	0.1987	0.3807
		(0.0253)	(0.0007)	(0.0011)	(1.13×10^{-3})	(0.0018)
$\phi_2^{(1)}$		0.0642	0.0945	-4.59×10^{-5}	-2.85×10^{-2}	-0.2884
		(0.0267)	(0.001)	(0.0017)	(0.0021)	(0.0034)
$\phi_0^{(2)}$			5.05×10^{-4}	1.67×10^{-3}	2.50×10^{-3}	0.0098
			(1.44×10^{-5})	(3.89×10^{-5})	(5.74×10^{-5})	(0.0001)
$\phi_1^{(2)}$			-0.0341	-0.1496	-0.102	-0.2895
			(0.0011)	(0.0016)	(0.0016)	(0.0019)
$\phi_2^{(2)}$			-0.1137	-0.0121	-0.0008	0.3485
			(0.0016)	(0.0019)	(0.0023)	(0.0034)
η			3.2368	1.2127	0.9956	2.2101
			(0.1262)	(0.025)	(0.025)	(0.0292)
c			1.09×10^{-4}	-3.89×10^{-3}	-4.39×10^{-3}	-1.81×10^{-2}
			(3.97×10^{-5})	(7.53×10^{-5})	(9.67×10^{-5})	(3.56×10^{-5})
ω	1.22×10^{-6}	1.24×10^{-6}	1.26×10^{-6}	1.54×10^{-6}	1.25×10^{-6}	1.53×10^{-6}
	(3.97×10^{-7})	(3.54×10^{-7})	(1.88×10^{-7})	(2.07×10^{-7})	(1.88×10^{-7})	(2.07×10^{-7})
α	0.1302	0.1309	0.1327	0.1388	0.1318	0.1369
	(0.0198)	(0.0199)	(0.0035)	(0.0038)	(0.0036)	(0.0038)
β	0.8556	0.8541	0.8518	0.8395	0.8529	0.8414
	(0.0224)	(0.0217)	(0.0043)	(0.0048)	(0.0043)	(0.0048)
N. Ite.	14	18	23	41	87	72

Estimation results for the full sample of FTSE Small Cap index returns. Standard deviations are in the parentheses.

Table 13: Forecasts statistics-USD/AUD (conditional mean)

	MFE	MSFE	MAFE	*Theil's U*	PT
RW	0.0006	3.72×10^{-5}	0.0045	0.9839	-0.4866
AR(2)	1.0002	1.0000	1.0002	0.9844	-0.2145
IWLS (SA)	1.1344	1.0259	1.0302	0.8618	-0.6627
fmincon (SA)	1.6050	1.6258	1.1239	0.7289	0.0819
IWLS (random)	0.8637	0.9911	0.9964	0.9025	1.2945
fmincon (random)	1.0419	1.0206	1.0076	0.9171	-1.2478

Forecasts statistics for the 500 1-day ahead forecasts with the daily USD/AUD returns. The columns respectively give mean forecasting error (MFE), mean square forecasting error (MSFE), mean absolute forecasting error (MAFE), *Theil's U* and Pesaran and Timmermann (1992) (PT) test statistic of directional forecasting.

Table 14: Forecasts statistics-FTSE SC (conditional mean)

	MFE	MSFE	MAFE	*Theil's U*	PT
RW	-0.0005	2.58×10^{-5}	0.0035	0.9027	0.0000
AR(2)	0.9100	0.9965	1.0001	0.9066	1.1168
IWLS (SA)	0.9358	1.0431	1.0357	0.8403	-0.6893
fmincon (SA)	1.0111	1.8748	1.1609	0.7151	-1.1157
IWLS (random)	0.8113	0.9902	1.0127	0.8177	0.8206
fmincon (random)	0.9500	1.0415	1.0180	0.8021	0.3126

Forecasts statistics for the 500 1-day ahead forecasts with the daily FTSE Small Cap returns. The columns respectively give mean forecasting error (MFE), mean square forecasting error (MSFE), mean absolute forecasting error (MAFE), *Theil's U* and Pesaran and Timmermann (1992) (PT) test statistic of directional forecasting.

Table 15: Predictive accuracy tests-USD/AUD (conditional mean)

	DM	CW	ENC_t	MSE_t	ENC_F	MSE_F
AR(2)	-0.04	0.06	0.06	-0.04	0.00	-0.01
IWLS (SA)	-1.43	0.33	0.26	-1.44	1.16	-12.64
fmincon (SA)	-1.84*	-0.99	-0.99	-1.84	-9.94	-192.45
IWLS (R)	0.84	1.97**	1.95***	0.84***	4.95**	4.50***
fmincon (R)	-1.74*	-1.00	-1.00	-1.75	-2.59	-10.07

Predictive accuracy test statistics for the daily USD/AUD return forecasts. The columns respectively give statistics for Diebold-Mariano (DM), Clark-West (CW), ENC_t, MSE_t, ENC_F, and MSE_F tests. P-values for the ENC_t, MSE_t, ENC_F, and MSE_F tests are calculated by bootstrapping. * p<0.10, ** p<0.05, *** p<0.01

Table 16: Predictive accuracy tests-FTSE SC (conditional mean)

	DM	CW	ENC_t	MSE_t	ENC_F	MSE_F
AR(2)	1.37	1.81*	1.50***	1.37***	0.96	1.75
IWLS (SA)	-1.73	-0.37	-0.26	-1.73	-1.50	-20.66
fmincon (SA)	-1.66	-0.42	-0.42	-1.66	-6.09	-233.30
IWLS (R)	0.43	1.63*	2.06	0.43**	12.67	4.96**
fmincon (R)	-0.73	0.60	0.60	-0.73	5.54	-19.90

Predictive accuracy test statistics for the daily FTSE Small Cap return forecasts. The columns respectively give statistics for Diebold-Mariano (DM), Clark-West (CW), ENC_t, MSE_t, ENC_F, and MSE_F tests. P-values for the ENC_t, MSE_t, ENC_F, and MSE_F tests are calculated by bootstrapping. * p<0.10, ** p<0.05, *** p<0.01

Table 17: Forecasts statistics (conditional variance), USD/AUD

	MFE	MSFE	MAFE	*Theil's U*
RW	-3.79$\times 10^{-6}$	4.52$\times 10^{-9}$	4.19$\times 10^{-5}$	0.5402
AR(2)	0.969	0.9987	0.9979	0.5407
IWLS (SA)	1.3332	1.0281	1.0236	0.5380
fmincon (SA)	4.4476	2.2516	1.2720	0.5959
IWLS (R)	1.2064	1.0298	1.0193	0.5422
fmincon (R)	3.9215	2.2363	1.1998	0.5714

Forecasts statistics for the 500 1-day ahead forecasts with the daily USD/AUD return realised variance. The columns respectively give mean forecasting error (MFE), mean square forecasting error (MSFE), mean absolute forecasting error (MAFE), and *Theil's U*.

Table 18: Forecasts statistics (conditional variance), FTSE SC

	MFE	MSFE	MAFE	*Theil's U*
RW	-2.24$\times 10^{-6}$	9.11$\times 10^{-9}$	2.96$\times 10^{-5}$	0.6957
AR(2)	0.8736	1.0125	0.9993	0.702
IWLS (SA)	1.0795	1.0243	1.0148	0.7017
fmincon (SA)	4.3977	1.2421	1.2139	0.6551
IWLS (R)	1.6562	1.0971	1.0554	0.6796
fmincon (R)	5.8406	1.6196	1.3314	0.6471

Forecasts statistics for the 500 1-day ahead forecasts with the daily FTSE SC return realised variance. The columns respectively give mean forecasting error (MFE), mean square forecasting error (MSFE), mean absolute forecasting error (MAFE), and *Theil's U*.

Table 19: Predictive accuracy test, DM (conditional variance)

	USD/AUD	FTSE SC
AR(2)	1.33	-1.62
IWLS (SA)	-1.03	-1.96**
fmincon (SA)	-2.35**	-1.84*
IWLS (R)	-1.74*	-1.38
fmincon (R)	-1.73*	-1.84*

DM test statistics for the daily USD/AUD and FTSE SC returns conditional variance forecasts. * p<0.10, ** p<0.05, *** p<0.01

Table 20: Predictive accuracy tests p-values-USD/AUD

		ENC_t	MSE_t	ENC_F	MSE_F
AR(2)	90^{th}	0.15	0.05	0.03	0.01
	95^{th}	0.18	0.07	0.08	0.01
	99^{th}	0.47	0.08	0.60	0.01
IWLS (SA)	90^{th}	2.99	1.92	33.81	12.58
	95^{th}	3.08	2.16	41.90	13.37
	99^{th}	3.35	2.39	53.43	14.47
fmincon (SA)	90^{th}	1.88	1.86	369.10	287.38
	95^{th}	1.89	1.87	391.28	303.53
	99^{th}	1.94	1.87	404.23	310.51
IWLS (R)	90^{th}	0.23	-0.55	1.54	-1.87
	95^{th}	0.33	-0.43	3.34	-1.26
	99^{th}	0.82	-0.26	9.82	-0.61
fmincon (R)	90^{th}	2.27	1.89	23.17	11.10
	95^{th}	2.32	1.98	30.85	11.64
	99^{th}	2.52	2.07	40.22	12.18

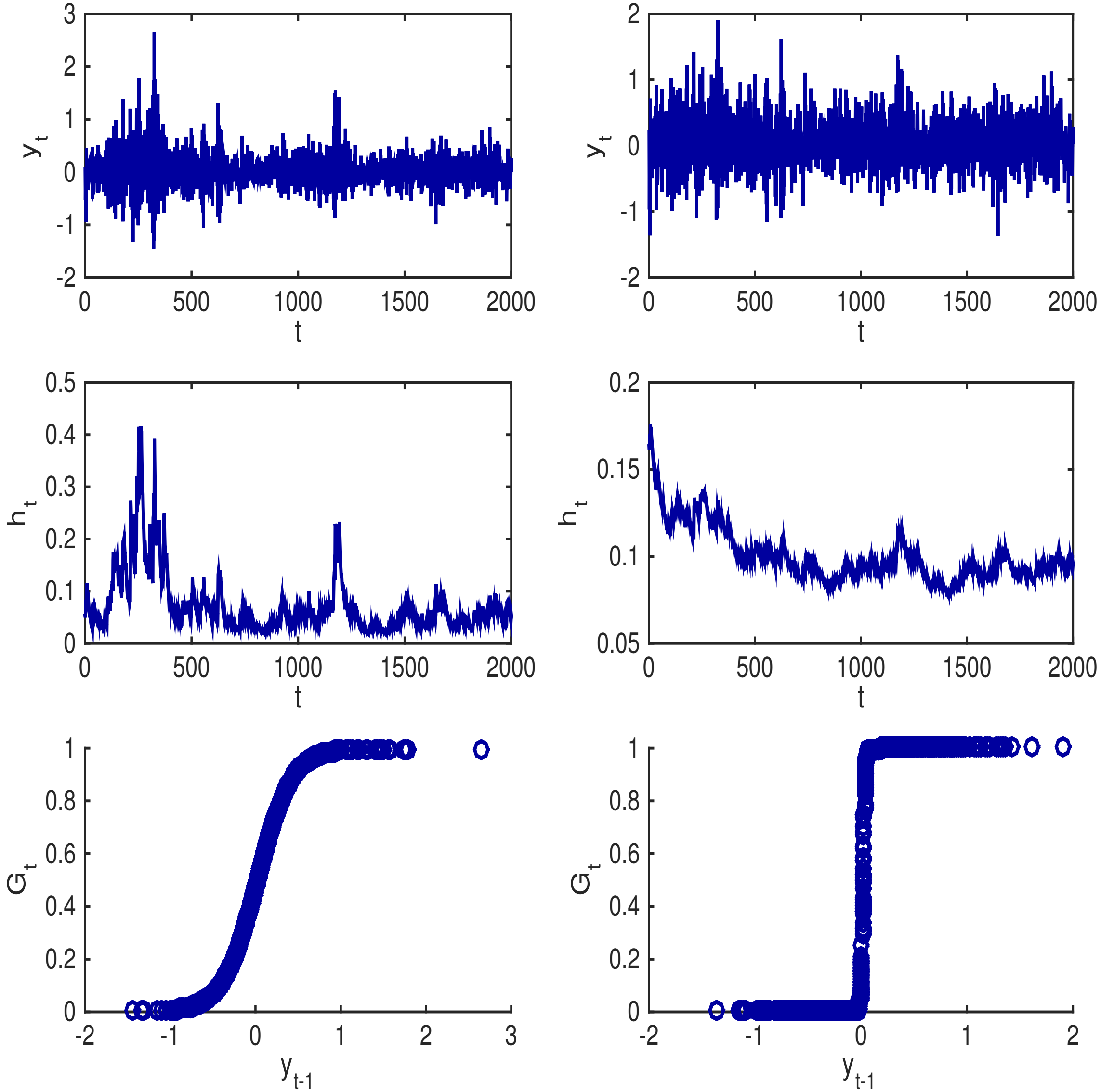
Bootstrapped percentile values for ENC_t, MSE_t, ENC_F, and MSE_F test statistics. Number of bootstrap is 10,000.

Table 21: Predictive accuracy tests p-values-FTSE SC

		ENC_t	MSE_t	ENC_F	MSE_F
AR(2)	90^{th}	1.23	1.04	2.57	2.33
	95^{th}	1.32	1.07	4.59	2.95
	99^{th}	1.48	1.12	13.38	4.43
IWLS (SA)	90^{th}	3.16	1.56	48.62	16.95
	95^{th}	3.77	1.97	100.48	30.44
	99^{th}	4.78	2.77	281.12	119.21
fmincon (SA)	90^{th}	1.68	1.65	472.91	389.80
	95^{th}	1.69	1.66	502.61	410.99
	99^{th}	1.80	1.67	527.83	428.04
IWLS (R)	90^{th}	2.09	-0.49	34.16	-3.79
	95^{th}	2.78	-0.09	72.96	-0.72
	99^{th}	4.01	0.81	196.32	21.28
fmincon (R)	90^{th}	1.35	0.79	47.07	17.42
	95^{th}	1.56	0.91	61.87	19.62
	99^{th}	2.63	1.03	95.78	23.81

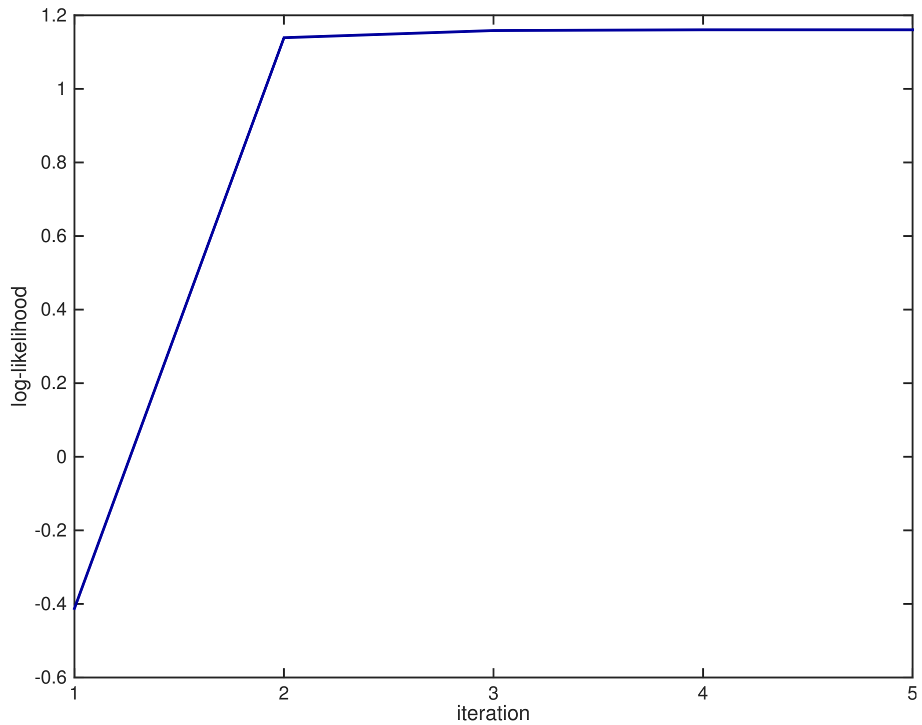
Bootstrapped percentile values for ENC_t, MSE_t, ENC_F, and MSE_F test statistics. Number of bootstrap is 10,000.

Figure 1: Simulated data



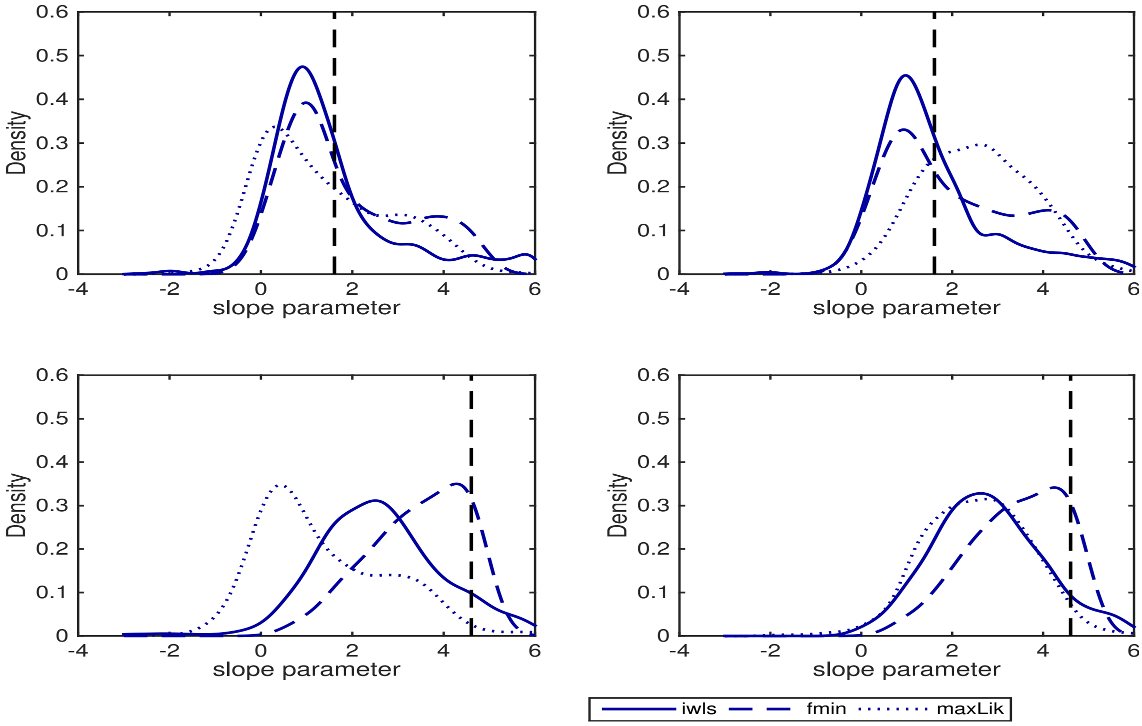
Representative plots of the simulated data with $(\alpha, \beta) = (0.09, 0.90)$ with $\eta_{re} = ln5$ (on the left) and $(\alpha, \beta) = (0.01, 0.98)$ specifications with $\eta_{re} = ln100$ (on the right) . Top panels: y_t against t, middle panels: h_t against t, bottom panels: G_t against y_{t-1}. Sample size is 2000.

Figure 2: IWLS log-likelihood function



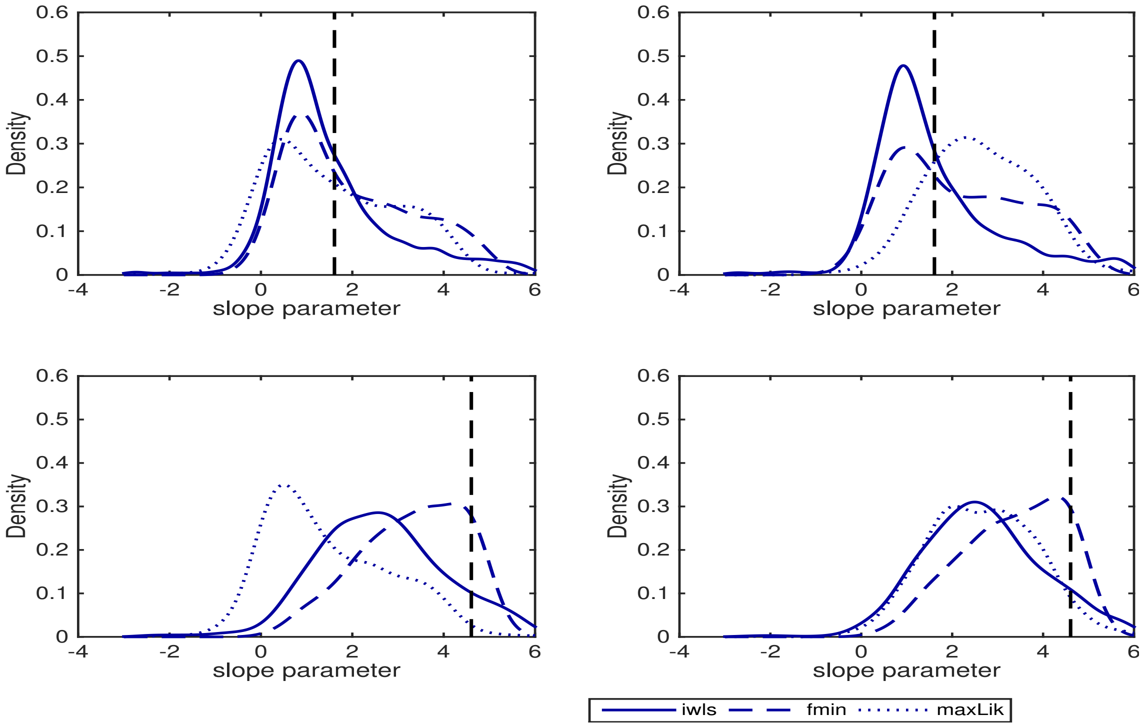
Evaluation of the log-likelihood function values during the IWLS estimation. This is a representative plot from the simulations with $(\alpha, \beta) = (0.09, 0.90)$ specification and $\eta_{re} = ln5$.

Figure 3: Estimated η $((\alpha,\beta) = (0.09, 0.90))$



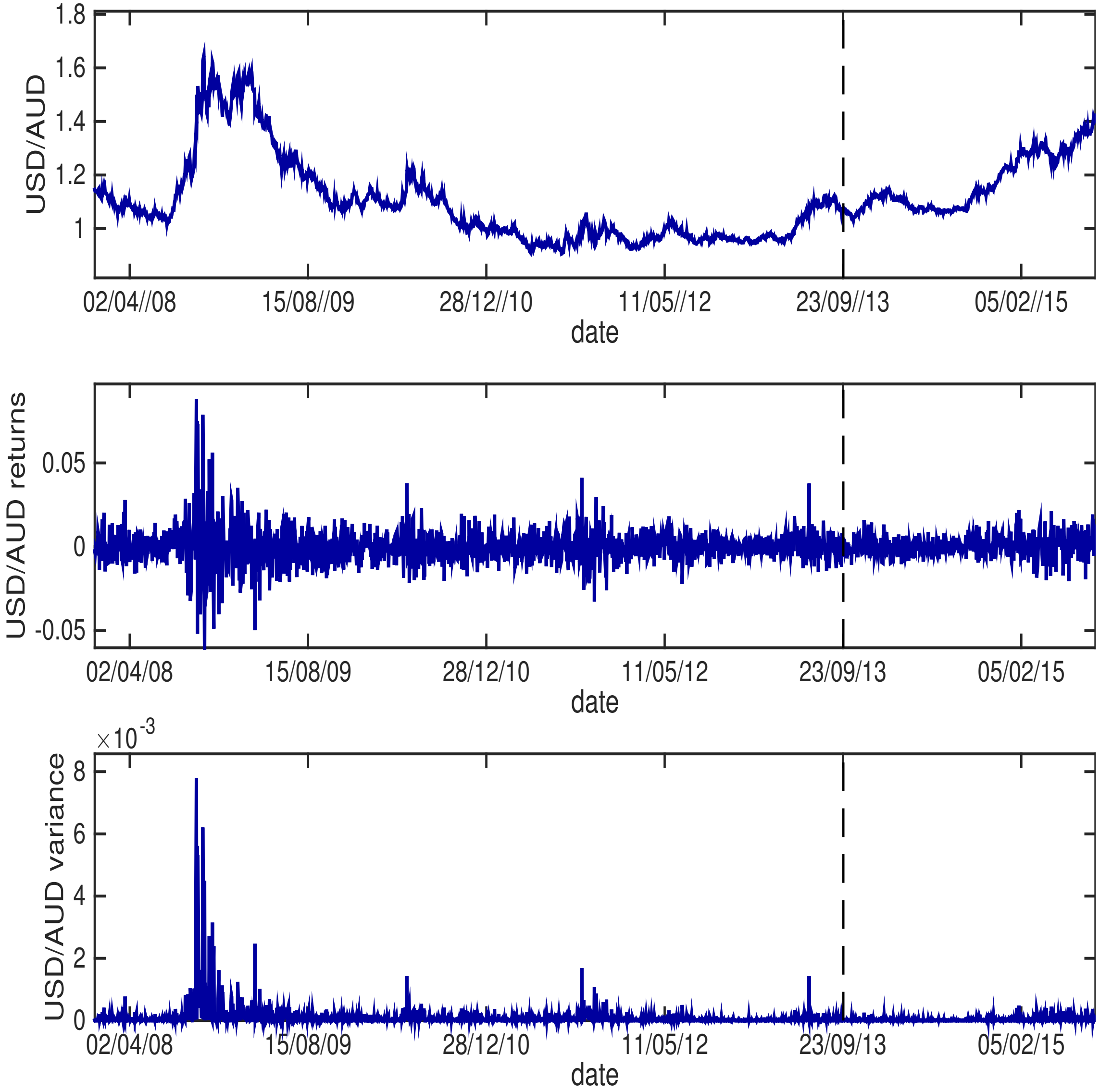
Kernel density plots of the slope parameter estimations for the $(\alpha,\beta) = (0.09, 0.90)$ specification. Top left panel: real slope parameter value is $ln5$, with SA initial values, top right panel: real slope parameter values is $ln5$, random initial values; bottom left panel: real slope parameter value is $ln100$, with SA initial values, bottom right panel: real slope parameter value is $ln100$, with random initial values. Vertical dashed line represents the real value of the slope parameter.

Figure 4: Estimated η $((\alpha,\beta) = (0.01, 0.98))$



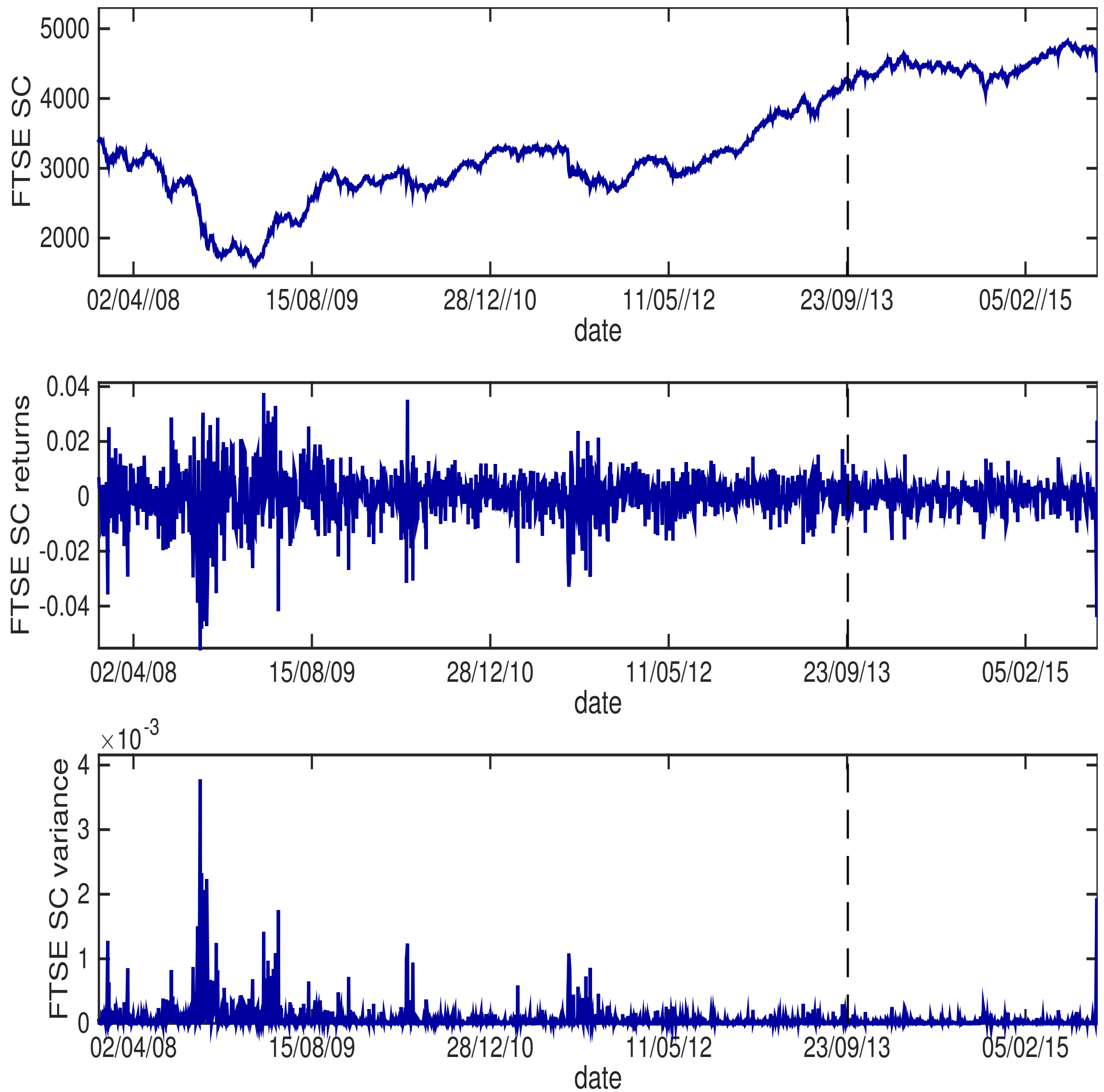
Kernel density plots of the slope parameter estimations for the $(\alpha,\beta) = (0.01, 0.98)$ specification. Top left panel: real slope parameter value is $ln5$, with SA initial values, top right panel: real slope parameter values is $ln5$, random initial values; bottom left panel: real slope parameter value is $ln100$, with SA initial values, bottom right panel: real slope parameter value is $ln100$, with random initial values. Vertical dashed line represents the real value of the slope parameter.

Figure 5: USD/AUD data



Plots for the daily USD/AUD exchange rates. Top panel: daily exchange rate levels, middle panel: daily exchange rate returns, bottom panel: realised variance of the daily exchange rate returns. The dashed vertical line marks the start of the forecasting period.

Figure 6: FTSE SC data



Plots for the daily FTSE Small Cap stock index. Top panel: daily index levels, middle panel: daily index returns, bottom panel: realised variance of the daily index returns. The dashed vertical line marks the start of the forecasting period.