SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2011-1

Modelling Volatility by Variance Decomposition

Cristina Amado and Timo Teräsvirta

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Modelling Volatility by Variance Decomposition

Cristina Amado*

University of Minho and NIPE

Campus de Gualtar, 4710-057 Braga, Portugal

Timo Teräsvirta†

CREATES, School of Economics and Management, Aarhus University

Building 1322, DK-8000 Aarhus, Denmark

January 2011

*E-mail: camado@eeg.uminho.pt

†E-mail: tterasvirta@econ.au.dk

Acknowledgements: An earlier version of this paper was circulated under the title "Modelling Conditional and Unconditional Heteroskedasticity with Smoothly Time-Varying Structure". This research has been supported by the Danish National Research Foundation. Material from this paper has been presented at the International Symposium on Econometric Theory and Applications, Xiamen, April 2006; 5th Annual International Conference 'Forecasting Financial Markets and Economic Decision-making', Lodz, May 2006; 13th International Conference on 'Forecasting Financial Markets', Marseille, May-June 2006; 26th International Symposium on Forecasting, Santander, June 2006; Workshop 'Volatility day', Stockholm, November 2006; Nordic Econometric Meeting, Tartu, May 2007; Symposium on 'Long Memory', Aarhus, June-July 2007; LACEA-LAMES, Bogotá, October 2007; 63rd European Meeting of the Econometric Society, Milan, August 2008; Symposium for Dag Tjøstheim, Bergen, October 2010, and in the seminars at Banca d'Italia, Rome, CEMFI, Madrid, Chinese University of Hong Kong, CREST (Malinvaud Seminar), Paris, ECARES, Brussels, European University Institute, Florence, Humboldt University, Berlin, Leibniz University, Hannover, Leonard N. Stern School of Business at New York University, Maastricht University, Monash University, Clayton, Stockholm School of Economics, Universidad Carlos III de Madrid, University of Minho, Braga, University of Vilnius and Uppsala University. We would like to thank the participants at these occasions for their comments, Stefan Lundbergh, Mika Meitz, Anders Rahbek and Esther Ruiz for useful discussions and suggestions, and Miguel Portela for allowing us to use his software for carrying out the specification strategy simulations. The responsibility for any errors and shortcomings in this article remains ours.

Abstract

In this paper, we propose two parametric alternatives to the standard GARCH model. They allow the variance of the model to have a smooth time-varying structure of either additive or multiplicative type. The suggested parameterisations describe both nonlinearity and structural change in the conditional and unconditional variances where the transition between regimes over time is smooth. The main focus is on the multiplicative decomposition that decomposes the variance into an unconditional and conditional component. A modelling strategy for the time-varying GARCH model based on the multiplicative decomposition of the variance is developed. It is heavily dependent on Lagrange multiplier type misspecification tests. Finite-sample properties of the strategy and tests are examined by simulation. An empirical application to daily stock returns and another one to daily exchange rate returns illustrate the functioning and properties of our modelling strategy in practice. The results show that the long memory type behaviour of the sample autocorrelation functions of the absolute returns can also be explained by deterministic changes in the unconditional variance.

JEL classification: C12; C22; C51; C52

Key words: Conditional heteroskedasticity; Structural change; Lagrange multiplier test; Misspecification test; Nonlinear time series; Time-varying parameter model.

1 Introduction

Modelling time-varying volatility of financial returns has been a flourishing field of research for a quarter of a century following the introduction of the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle (1982) and the Generalized ARCH (GARCH) model developed by Bollerslev (1986). These basic models have since been generalised in many ways, see Teräsvirta (2009) for a recent survey. The increasing popularity of GARCH models has been mainly due to their ability to describe the dynamic structure of volatility clustering of stock return series, specifically over short periods of time. However, one may expect that economic or political events or changes in institutions cause the structure of volatility to change over time. This means that the assumption of stationarity may be inappropriate under the evidence of structural changes in financial return series. Recently, Mikosch and Stărică (2004) argued that stylized facts in financial return series such as the long-range dependence and the 'integrated GARCH effect' can be well explained by unaccounted structural breaks in the unconditional variance; see also Lamoureux and Lastrapes (1990). Diebold (1986) was the first to suggest that occasional level shifts in the intercept of the GARCH model can bias the estimates towards the parameters of an integrated GARCH model.

Another line of research has focussed on explaining nonstationary behaviour of volatility by long-memory models, such as the Fractionally Integrated GARCH (FIGARCH) model by Baillie, Bollerslev, and Mikkelsen (1996). The FIGARCH model is not the only way of handling the 'integrated GARCH effect' in return series. Baillie and Morana (2009) generalized the FIGARCH model by allowing a deterministically changing intercept. Hamilton and Susmel (1994) and Cai (1994) suggested a Markov-switching ARCH model for the purpose, and their model has later been generalized by others. One may also assume that the GARCH process contains sudden deterministic switches and try and detect them; see Berkes, Gombay, Horváth, and Kokoszka (2004) who proposed a method of sequential switch or change-point detection.

Yet another way of dealing with high persistence would be to explicitly assume that the volatility process is 'smoothly' nonstationary and model it accordingly. Dahlhaus and Subba Rao (2006) introduced a time-varying ARCH process for modelling nonstationary volatility. Their tvARCH model is asymptotically locally stationary at every point of observation but it is globally nonstationary because of time-varying parameters. Van Bellegem and von Sachs (2004) and, later, Engle and Gonzalo Rangel (2008) assumed that the variance of the process of interest can be decomposed into two components, a stationary and a nonstationary one. The former authors fitted the deterministic component nonparametrically to the squared observations, whereas the latter described the nonstationary component by using splines. Both assumed that the stationary component follows a GARCH process. For a similar approach using a different version of splines, see Brownlees and Gallo (2010). Mishra, Su, and Ullah (2010) did not explicitly mention nonstationarity but used the multiplicative decomposition to correct the potential misspecification due to a 'rough' parametric GARCH specification by a smooth nonparametric component. Yet another multiplicative decomposition was introduced in Osiewalski (2009) and Osiewalski and Pajor (2009). Some of these developments are described in detail in van Bellegem (2011) and Teräsvirta (2011).

In this paper, we introduce two nonstationary GARCH models for situations in which volatility appears to be nonstationary. First, we propose an additive time-varying parameter model, in which a directly time-dependent component is added to the GARCH specification. In the second alternative, the variance is multiplicatively decomposed into the stationary and nonstationary component as in van Bellegem and von Sachs (2004) or Engle and Gonzalo Rangel (2008). We show that the multiplicative decomposition is a special case of the general additive decomposition. These two alternatives are quite flexible representations of volatility and can describe many types of nonstationary behaviour. We emphasize the role of model building in this approach. The model is first specified, which includes determining the structure of the deterministic component and then estimated. After parameter estimation, the model is evaluated by misspecification tests, following the ideas in Eitrheim and Teräsvirta (1996) and Lundbergh and Teräsvirta (2002).

The outline of this paper is as follows. In Section 2 we present the new Time-Varying (TV-) GARCH or GJR-GARCH model and highlight some of its properties. Maximum likelihood estimation of the model is discussed in Section 3. The modelling strategy is presented in Section 4. Specification and misspecification tests for the TV-GARCH model are considered in Section 5. Section 6 contains simulation results on the empirical performance of the tests and the specification strategy. In Section 7 we apply our modelling cycle to both stock and exchange rate returns. Finally, Section 8 contains concluding remarks.

2 The model

Let the model for an asset or index return y_t be

$$y_t = \mu_t + \varepsilon_t$$

where $\{\varepsilon_t\}$ is an innovation sequence with the conditional mean $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and a potentially time-varying conditional variance $\mathsf{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2$, and $\mathcal{F}_{t-1}$ is the σ-field generated by the available information until $t-1$. We assume that $\mathsf{E}(y_t|\mathcal{F}_{t-1}) = 0$ and focus solely on σ_t^2. Let

$$\varepsilon_t = \zeta_t \sigma_t \tag{1}$$

where $\{\zeta_t\}$ is a sequence of independent random variables with mean zero and variance one. Furthermore, assume that σ_t^2 is a time-varying representation measurable with respect to $\mathcal{F}_{t-1}$ with either an additive structure

$$\sigma_t^2 = h_t + g_t \tag{2}$$

or a multiplicative one

$$\sigma_t^2 = h_t g_t. \tag{3}$$

The function h_t is a component describing conditional heteroskedasticity in the observed process y_t, whereas g_t introduces nonstationarity. Since we are going apply our model to stock return series, where asymmetry of the response to shocks becomes an issue, we assume that h_t follows the stationary GJR-GARCH(p,q) model

$$h_t = \alpha_0 + \sum_{i=1}^{q} (\alpha_i + \lambda_i I(\varepsilon_{t-i} < 0))\varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{4}$$

where $I(A)$ is the indicator variable: $I(A) = 1$ when A is true, and zero otherwise. Then the GJR-GARCH(p,q) model is nested in (2) when $g_t \equiv 0$ and in (3) when $g_t \equiv 1$. Note that when (3) holds, ε_{t-i}^2 is replaced by $\phi_{t-i}^2 = \varepsilon_{t-i}^2 / g_{t-i}$, $i = 1, ..., q$, in (4). Both (2) and (3) combined with (1) define a time-varying parameter GARCH model.

In order to characterize smooth changes in the conditional variance we assume that the parameters in (4) vary smoothly over time. This is done for example by defining the function g_t in (2) as follows:

$$g_t = \{\alpha_0^* + \sum_{i=1}^{q} (\alpha_i^* + \lambda_i^* I(\varepsilon_{t-i} < 0))\varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j^* h_{t-j}\} G(t^*; \gamma, \mathbf{c}), \tag{5}$$

where $G(t^*; \gamma, \mathbf{c})$ is the so-called transition function which is a continuous and non-negative function bounded between zero and one. Furthermore, the rescaled time $t^* = t/T$, where T is the number of observations. A suitable choice for $G(t^*; \gamma, \mathbf{c})$ is the general logistic transition function

$$G(t^*; \gamma, \mathbf{c}) = \left(1 + \exp\left\{-\gamma \prod_{k=1}^{K} (t^* - c_k)\right\}\right)^{-1}, \ \gamma > 0, \ c_1 \leq ... \leq c_K. \tag{6}$$

This function is such that the parameters of the GJR-GARCH model (1)-(2) fluctuate smoothly over time between $(\alpha_i, \lambda_i, \beta_j)$ and $(\alpha_i + \alpha_i^*, \lambda_i + \lambda_i^*, \beta_j + \beta_j^*)$, $i = 0, 1, ..., q$, $j = 1, ..., p$. The slope parameter γ controls the degree of smoothness of the transition function. When $\gamma \longrightarrow \infty$, the switch from one set of parameters to another in (2) is abrupt, that is, the process contains structural breaks at $c_1, c_2, ..., c_K$. The order $K \in \mathbb{Z}_+$ determines the shape of the transition function. Typical choices for the transition function in practice are $K = 1$ and $K = 2$. These are illustrated in Figure 1 for a set of values for γ, c_1, and c_2. Large values of γ increase the velocity of transition from 0 to 1 as a function of t^*. The TV-GARCH model with $K = 1$ is suitable for describing return processes whose volatility dynamics are different before and after the smooth structural change. When $K = 2$, the parameters change but eventually return towards their original values as a function of time.

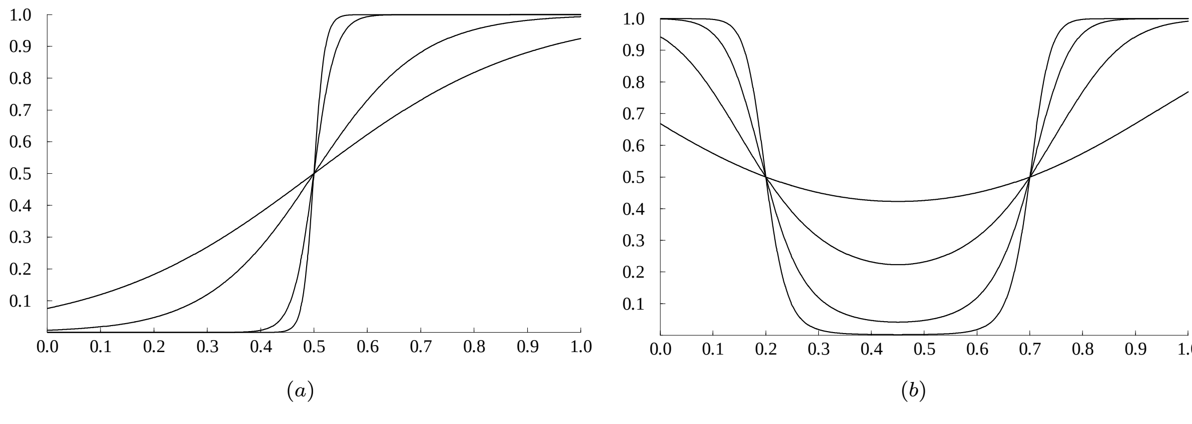


Figure 1. Plots of the logistic transition function (6) for: (a) $K = 1$ with location parameter $c_1 = 0.5$; and (b) $K = 2$ with location parameters $c_1 = 0.2$ and $c_2 = 0.7$ for $\gamma = 5, 10, 50$, and 100, where the lowest value of γ corresponds to the smoothest function.

More generally, one can define an extended version of the additive TV-GJR-GARCH model by allowing more than one transition function in g_t. The result becomes

$$g_t = \sum_{l=1}^{r} \{\alpha_{0l} + \sum_{i=1}^{q} (\alpha_{il} + \lambda_{il} I(\varepsilon_{t-i} < 0))\varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{jl} h_{t-j}\} G_l(t^*; \gamma_l, \mathbf{c}_l) \tag{7}$$

where $G_l(t^*; \gamma_l, \mathbf{c}_l)$, $l = 1, ..., r$, are logistic functions as in (6) with smoothness parameter γ_l and a threshold parameter vector $\mathbf{c}_l$. The parameters in (4) and (7) satisfy the restrictions $\alpha_0 + \sum_{l=1}^{j} \alpha_{0l} > 0$, $\alpha_i + \lambda_i/2 + \sum_{l=1}^{j}(\alpha_{il} + \lambda_i/2) > 0$, $i = 1, ..., q$, and $\beta_i + \sum_{l=1}^{j} \beta_{il} \geq 0$, $i = 1, ..., p$, all for any $j \in \{1, ..., r\}$. These conditions are sufficient for $g_t > 0$ for all t.

The model (1), (2), (4) and (5) or, more generally (7), is an additive TV-GJR-GARCH models whose intercept, ARCH and GARCH parameters are all time-varying. This implies that the model is capable of accommodating systematic changes both in the 'baseline volatility' (or unconditional variance) and in the amplitude of volatility clusters. Such changes cannot be explained by a constant parameter GARCH model.

Function (7) with $r > 1$ is extremely flexible, which is likely to make the model difficult to estimate in practice. A more applicable but still flexible model is obtained by only letting the intercept change smoothly over time. This leads to the following definition for g_t:

$$g_t = \sum_{l=1}^{r} \alpha_{0l} G_l(t^*; \gamma_l, \mathbf{c}_l). \tag{8}$$

It may be mentioned that Baillie and Morana (2009) recently proposed a GARCH model which also has a deterministically time-varying intercept. It is modelled using the flexible functional form of Gallant (1984) based on the Fourier decomposition. Their model differs from our time-varying-intercept GARCH model with (8) in the sense that it is in other respects a FIGARCH model, and the authors called it the Adaptive FIGARCH model.

In the stationary GJR-GARCH(p, q) model, the unconditional variance of the returns is constant over time, that is, $\mathsf{E}\varepsilon_t^2 = \alpha_0/(1 - \sum_{i=1}^{q}(\alpha_i + \lambda_i/2) - \sum_{j=1}^{p} \beta_j) < \infty$ if and only if $\sum_{i=1}^{q}(\alpha_i + \lambda_i/2) + \sum_{j=1}^{p} \beta_j < 1$. This assumption is not consistent with the behaviour of the volatility of the stock market returns, however, if the dynamic behaviour of volatility changes over time. The additive TV-GARCH model with a time-varying intercept is capable of generating changes in the dynamic behaviour of the unconditional variance. The model (2), (4) and (8) can be seen as a GARCH(p, q) model with a stochastic time-varying intercept fluctuating smoothly over time. Such a model can thus explain both the clustering of volatility and smooth changes in the conditional variance.

Consider again the model (1), (2), (4) and (7) and assume that $\alpha_{0l} = \alpha_0 \delta_l$, $\alpha_{il} + \lambda_{il} = (\alpha_i + \lambda_i)\delta_l$, $i = 1, ..., q$; $\beta_{jl} = \beta_j \delta_l$, $j = 1, ..., p$, for $l = 1, ..., r$. Imposing these restrictions on (7) and rewriting (2) yields

$$\sigma_t^2 = h_t \{1 + \sum_{l=1}^{r} \delta_l G_l(t^*; \gamma_l, \mathbf{c}_l)\} \tag{9}$$

which is the multiplicative representation (3) as $g_t = 1 + \sum_{l=1}^{r} \delta_l G_l(t^*; \gamma_l, \mathbf{c}_l)$. Equation (9) is thus a special case of the additive TV-GJR-GARCH model (2), (4) and (7). The proportionality factors δ_l, $l = 1, ..., r$, have to satisfy certain restrictions for $g_t > 0$ for all t. The argument for favouring this representation of g_t is based on the fact that g_t as a linear combination of logistic functions is a universal approximator. It is possible to approximate any function $H(t^*)$ satisfying mild regularity conditions arbitrarily accurately with g_t in the sense that there exists an $r \leq r_0 < \infty$ such that $|H(t^*) - g_t| < \varepsilon/T$ for any $\varepsilon > 0$ and for all t. This indicates that g_t is a very flexible function capable of describing many types of change in the unconditional variance.

The multiplicative model has a straightforward interpretation. Writing it in terms of (1) and (3) as follows:

$$\phi_t = \varepsilon_t / g_t^{1/2} = \zeta_t h_t^{1/2}, \qquad t = 1, ..., T \tag{10}$$

it is seen that ϕ_t has a constant unconditional variance $\mathsf{E}h_t$ and, moreover, that ϕ_t has a standard stationary GJR-GARCH(p,q) representation. Turning (10) around, one finds that

$$\psi_t = \varepsilon_t / h_t^{1/2} = \zeta_t g_t^{1/2}, \qquad t = 1, ..., T$$

form a sequence of independent but not identically distributed observations, as the unconditional variance of ψ_t changes smoothly as a function of time. In the following we shall focus on the multiplicative TV-GJR-GARCH model and leave the additive version of the model for further research.

3 Estimation of multiplicative TV-GJR-GARCH models

3.1 Estimation applying maximisation by parts

The conditional (quasi) log-likelihood function of the multiplicative TV-GJR-GARCH model has the following form:

$$L_T(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\varepsilon}) = k - (1/2)\sum_{t=1}^{T}(\ln h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) + \ln g_t(\boldsymbol{\theta}_1)) - (1/2)\sum_{t=1}^{T}\frac{\varepsilon_t^2}{h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) g_t(\boldsymbol{\theta}_1)} \qquad (11)$$

where $\boldsymbol{\theta}_1 = (\boldsymbol{\delta}', \boldsymbol{\gamma}', \mathbf{c}')'$ with $\boldsymbol{\delta} = (\delta_1, ..., \delta_r)'$, $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_r)'$, $\gamma_l > 0$, $l = 1, ..., r$, and $\mathbf{c} = (c_1, ..., c_r)'$, $c_1 < ... < c_r$. (For notational simplicity, we assume that each transition function in $g_t(\boldsymbol{\theta}_1)$ is a simple logistic function: $K = 1$ in (6).) Furthermore, $\boldsymbol{\theta}_2 = (\alpha_0, \boldsymbol{\alpha}', \boldsymbol{\lambda}', \boldsymbol{\beta}')'$ where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_q)'$, $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_q)'$, and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$. Estimation of the parameters by straightforward maximization of (11) tends to be numerically very difficult. A solution to this problem is to apply maximization by parts; see Song, Fan, and Kalbfleisch (2005) and Fan, Pastorello, and Renault (2007) for discussion. The estimation proceeds as follows:

1. Estimate $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ by assuming $\boldsymbol{\theta}_2 = \alpha_0$, obtain $\boldsymbol{\theta}_1^{(1)}$ (and $\boldsymbol{\theta}_2^{(0)} = \widehat{\alpha}_0^{(0)}$). At this stage, it is useful to set $h_t \equiv 1$, define

$$g_t(\alpha_0, \boldsymbol{\theta}_1) = \alpha_0 + \sum_{l=1}^{r}\delta_l^* G_l(t^*; \gamma_l, c_l) \qquad (12)$$

   where $\alpha_0 > 0$, in (11), and then estimate α_0, $\delta_1^*, ..., \delta_r^*$, $\boldsymbol{\gamma}$ and $\mathbf{c}$. Finally, the estimate of $\boldsymbol{\delta}^{(1)}$ is obtained as follows: $\delta_l^{(1)} = \delta_l^{*(1)} / \alpha_0^{(0)}$, $l = 1, ..., r$.

2. Estimate $\boldsymbol{\theta}_2$ including α_0, letting $g_t(\boldsymbol{\theta}_1) = g_t(\boldsymbol{\theta}_1^{(1)})$.

3. Estimate $\boldsymbol{\theta}_1$ assuming

$$g_t(\boldsymbol{\theta}_1) = 1 + \sum_{l=1}^{r}\delta_l G_l(t^*; \gamma_l, c_l)$$

   while setting $h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = h_t(\boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_2^{(1)})$. This yields $\boldsymbol{\theta}_1^{(2)}$. The important detail here is that $\boldsymbol{\theta}_1$ is re-estimated holding the parameter estimates in $h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ unchanged.

Repeat Steps 2 and 3 until convergence.

Song, Fan, and Kalbfleisch (2005) showed that under regularity conditions, the resulting estimators are consistent and asymptotically normal. In order to prove that this is the case even for the TV-GJR-GARCH model, we make the following assumptions.

AG1. The parameter space $\Theta_1 = \{\alpha_0 \times \Delta \times \Gamma \times C\}$ is compact, where $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_r)' \in \Gamma$, $\gamma_l > 0$, $l = 1, ..., r$, and $\mathbf{c} = (c_1, ..., c_r)' \in C$. The true parameter $\boldsymbol{\theta}_1^0$ is an interior point of Θ_1.

AG2. The elements of $\boldsymbol{\delta} \in \Delta$ are restricted such that $\max_{j=1,\ldots,q} \delta_j \leq M_\delta < \infty$, and

$$\inf_{\boldsymbol{\theta}_1 \in \Theta_1} g_t(\boldsymbol{\theta}_1) \geq g_{\min} > 0 \tag{13}$$

for all t. Furthermore, $\alpha_0 > 0$.

AG3. In (1), $\mathsf{E}|\zeta_t|^{2(2+\phi)} < \infty$ for some $\phi > 0$.

The restrictions $\gamma_l > 0$, $l = 1, ..., r$, in AG1 are identification restrictions required for obtaining a unique maximum value for (11). AG2 stipulates that $g_t(\boldsymbol{\theta}_1)$ is a positive and finite-valued function of t.

Theorem 1 *Let $\widehat{\boldsymbol{\theta}}_{1T}$ be the maximum likelihood estimator of $\boldsymbol{\theta}_1^0$,*

$$\widehat{\boldsymbol{\theta}}_{1T} = \arg\max L_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon})$$

where the quasi log-likelihood function for the model is

$$L_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon}) = \sum_{t=1}^{T} \ell(\boldsymbol{\theta}_1, \varepsilon_t) \tag{14}$$

with

$$\ell(\boldsymbol{\theta}_1, \varepsilon_t) = k - (1/2)\{\ln g_t(\boldsymbol{\theta}_1) + \frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)}\}. \tag{15}$$

Then

$$T^{1/2}(\widehat{\boldsymbol{\theta}}_{1T} - \boldsymbol{\theta}_1^0) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1}(\boldsymbol{\theta}_1^0)\mathbf{B}(\boldsymbol{\theta}_1^0)\mathbf{A}^{-1}(\boldsymbol{\theta}_1^0))$$

where

$$\mathbf{A}(\boldsymbol{\theta}_1^0) = -(1/2) lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \frac{1}{g_t^2(\boldsymbol{\theta}_1^0)} \{\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} \frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1'}\}|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0}$$

and

$$\mathbf{B}(\boldsymbol{\theta}_1^0) = \frac{\mu_4 - 1}{4} lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \frac{1}{g_t^2(\boldsymbol{\theta}_1^0)} \{\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} \frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1'}\}|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0}.$$

when $T \to \infty$.

Proof. See the Appendix. ■

The estimator $\widehat{\boldsymbol{\theta}}_1$ is consistent for $\boldsymbol{\theta}_1^0$ and asymptotically normal but not efficient, because it has been assumed that $h_t = \alpha_0$. It offers, however, an adequate starting-point for further iteration, see Song, Fan, and Kalbfleisch (2005).

In order to show that the maximum likelihood estimators of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are both consistent and asymptotically normal, we have to prove that the parameter vector $\boldsymbol{\theta}_2$ of the model

$$\phi_t = \zeta_t h_t^{1/2} \tag{16}$$

can be consistently estimated at iteration k, and that the estimators of $\boldsymbol{\theta}_2$ are asymptotically normal when $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(k)}$, the consistent estimate of $\boldsymbol{\theta}_1^0$ from the kth iteration. Straumann and Mikosch (2006) recently proved the consistency and asymptotic normality of the maximum likelihood estimators of the parameters of the Asymmetric GARCH(p, q) model

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i^*(|\varepsilon_{t-i}| - \gamma^* \varepsilon_{t-i})^2 + \sum_{i=1}^{p} \beta_i^* h_{t-1} \tag{17}$$

under general conditions. This model can be written in the GJR-GARCH form as pointed out by Meitz and Saikkonen (in press). Note, however, that for $q > 1$, the analogy between (17) and the GJR-GARCH formulation (3) implies the restrictions $\lambda_i = \kappa\alpha_i$, $i = 1, ..., q$, in the latter. We can then use the results of Straumann and Mikosch (2006).

We make the following assumptions:

AH1. In (4), $\alpha_0 > 0$, and $\sum_{i=1}^{q}(\alpha_i + \lambda_i/2) + \sum_{j=1}^{p} \beta_j < 1$, where $\lambda_i = \kappa\alpha_i$, $i = 1, ..., q$. This, together with Assumption AG3 made above, is a sufficient condition for weak stationarity of the GJR-GARCH model.

Remark. This condition is much stronger than what is actually needed. Nevertheless, since idea of the decomposition in Equation (10) is that $\{\phi_t\}$ is a weakly stationary sequence (nonstationarity is modelled by g_t), we use this assumption. See Straumann and Mikosch (2006) for considerably weaker assumptions.

AH2. The polynomials $\sum_{i=1}^{q}(1 + \kappa/2)\alpha_i z^i$ and $1 - \sum_{j=1}^{p} \beta_j z^j$ do not have common roots.

AH3. The parameter space $\Theta_2 = \{\alpha_0, \boldsymbol{\alpha}, \kappa, \boldsymbol{\beta}\}$ is compact and the true parameter value $\boldsymbol{\theta}_2^0$ belongs to the interior of Θ_2.

Theorem 2 *Consider the GJR-GARCH model (16) with (17) and assume that the assumptions AG3 and AH1-3 hold. Then the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{2T}$ of $\boldsymbol{\theta}_2^0$ given $\boldsymbol{\theta}_1^{(k)}$, where $\boldsymbol{\theta}_1^{(k)}$ is fixed, is consistent and*

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_{2T} - \boldsymbol{\theta}_2^0) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{V}_2^{(k)})$$

as $T \to \infty$, where $\mathbf{V}_2^{(k)}$ is the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}_{2T}$ given $\boldsymbol{\theta}_1^{(k)}$.

Proof. Due to the analogy between the GJR-GARCH model and the Asymmetric GARCH model, the desired result follows from Straumann and Mikosch (2006), Theorems 5.5 and 8.1. Assumption AG3 is needed because the proof in that article requires $\mathsf{E}\zeta_t^4 < \infty$. ■

Combining Theorems 1 and 2 and the results in Song, Fan, and Kalbfleisch (2005), Theorem 3, one can conclude that after the kth iteration, the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_T^{(k)} = (\widehat{\boldsymbol{\theta}}_{1T}^{(k)\prime}, \widehat{\boldsymbol{\theta}}_{2T}^{(k)\prime})'$ is consistent and asymptotically normal, that is,

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_T^{(k)} - \boldsymbol{\theta}^0) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{V}^{(k)})$$

as $T \to \infty$, with the asymptotic covariance matrix $\mathbf{V}^{(k)}$ given in Song, Fan, and Kalbfleisch (2005). When $k \to \infty$, $\mathbf{V}^{(k)}$ converges to the asymptotic covariance matrix of $\sqrt{T}\widehat{\boldsymbol{\theta}}_T$, where $\widehat{\boldsymbol{\theta}}_T$ is the final maximum likelihood estimator.

3.2 Numerical aspects

Two remarks are in order regarding numerical aspects of the estimation of TV-GARCH models. The first one concerns the accuracy of the slope estimates when the true parameters γ_l are very large. In order to achieve an accurate estimate for a large γ_l, the number of observations of the transition variable in the neighbourhood of c_l must be very large. This is due to the fact that even large changes in γ_l only have an effect on the transition function in a small neighbourhood of c_l. But then, for the same reason for large γ_l it is sufficient to obtain an estimate that is large; whether or not it is very accurate is not of utmost importance. Note that if $\hat{\gamma}_l$ is large, an 'insignificant' $\hat{\gamma}_l$ is an indication of a large γ_l, not of $\gamma_l \equiv 0$. Besides, because of the identification problem the t-ratio does not have its standard asymptotic distribution when $\gamma_l \equiv 0$. A more serious problem is that large estimates for the slope parameter γ_l may lead to numerical problems when carrying out parameter constancy tests. A simple solution, suggested in Eitrheim and Teräsvirta (1996), is to omit those elements of the score that are partial derivatives with respect to the parameters in the transition function. When γ_l is sufficiently large, this can be done with only a negligible effect on the value of the test statistic.

The second remark has to do with the computation of the derivatives of the log-likelihood function. Many of the existing optimization algorithms require the computation of at least the first and, in some cases, also the second derivatives of the log-likelihood function. It has been a common practice to use numerical derivatives that are relatively fast to compute and reliable and thus avoid the derivation of exact analytic derivatives. Fiorentini, Calzolari, and Panattoni (1996), however, encourage the employment of analytic derivatives, because that leads to fewer iterations than optimization with numerical derivatives. Furthermore, the use of analytic derivatives also improves the accuracy of the estimates of the standard errors of the parameter estimates. Consequently, we use analytic first derivatives in all the computations, both in calculating values of the test statistics and in estimating TV-GARCH models.

4 A three-stage modelling strategy for building TV-GJR-GARCH models

We propose a model-building cycle for TV-GJR-GARCH or TV-GARCH models similar to the specific-to-general strategy for nonlinear models of the conditional mean considered in Teräsvirta (1998) and Teräsvirta, Tjøstheim and Granger (2010, Chapter 16), among others. The cycle or strategy consists of the following three stages:

1. Test the null hypothesis of no conditional heteroskedasticity in $\{\zeta_t\}$. If it is rejected, model the conditional variance h_t as defined in (4) with $p = q = 1$. In some applications, it may be assumed $\lambda_1 = 0$. In financial applications the test of no conditional heteroskedasticity can most often be omitted, however, because return series of sufficiently high frequency are likely to contain at least some conditional heteroskedasticity. The assumption $p = q = 1$ reflects the fact that a first-order GARCH model is very often found a sufficient description of the data.

2. After estimating the GJR-GARCH- or GARCH model, test the null hypothesis of constant variance against a time-varying unconditional variance with a single transition function. This implies testing $\mathrm{H}_{10} : g_t \equiv 1$ in (3) against $\mathrm{H}_{11} : g_t = 1 + \delta_1 G_1(t^*; \gamma_1, \mathbf{c}_1)$. Let α denote the significance level of the test. In case of a rejection, estimate a model with one transition and test $\mathrm{H}_{20} : g_t = 1 + \delta_1 G_1(t^*; \gamma_1, \mathbf{c}_1)$ against $\mathrm{H}_{21} : g_t = 1 + \sum_{l=1}^{2} \delta_l G_l(t^*; \gamma_l, \mathbf{c}_l)$ at the significance level $\alpha\tau$, $0 < \tau < 1$. Continue until the first non-rejection of the null hypothesis. For reasons of parsimony, the significance level is reduced at each step of the testing procedure and converges to zero. After rejecting a null hypothesis, choose the type of transition, $K = 1$ or $K = 2$ in (6). The appropriate test will be described in Section 5.

3. Evaluate the estimated model by means of LM or LM-type diagnostic tests to be discussed in Section 5. They are generalisations of tests proposed by Lundbergh and Teräsvirta (2002). If the model passes all of them, tentatively accept it. Otherwise, respecify the model or consider another family of volatility models.

Modelling could in theory also be initiated by first specifying and estimating the model (1) and (3) with $h_t = \alpha_0$. The ensuing specification tests would then suffer from size distortion due to ignored conditional heteroskedasticity. This drawback could in principle be remedied by a wild bootstrap, but our experiments with two-point distributions along the lines in Davidson, Monticini, and Peel (2007) did not bring satisfactory results. This led us to prefer the order of specification described above.

5 Misspecification testing of multiplicative TV-GJR-GARCH models

In this section we shall consider misspecification tests relevant for Stage 3 of our modelling strategy. One of the tests is of particular interest for Stage 2 and can be viewed as a specification test. The general idea is to construct an augmented version of the TV-GJR-GARCH model by introducing a new component into the original model. This component is a function that is at least twice continuously differentiable with respect to the elements of $\boldsymbol{\theta}_3$, the vector of additional parameters. Misspecification tests considered here may be divided into three categories. The first two correspond to additive and the third one to multiplicative misspecification.

In order to construct the alternative hypotheses, define a new function $f_t(\boldsymbol{\theta}_3)$ which is at least twice continuously differentiable in an open neighbourhood of $\boldsymbol{\theta}_3 = \mathbf{0}$. The three alternative hypotheses are as follows:

1. The deterministic component is additively misspecified:

$$\sigma_t^2 = h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)\{g_t(\boldsymbol{\theta}_1) + f_t(\boldsymbol{\theta}_3)\}.$$

2. The conditional variance component is additively misspecified:

$$\sigma_t^2 = \{h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) + f_t(\boldsymbol{\theta}_3)\}g_t(\boldsymbol{\theta}_1).$$

3. The variance is multiplicatively misspecified:

$$\sigma_t^2 = h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)g_t(\boldsymbol{\theta}_1)f_t(\boldsymbol{\theta}_3).$$

In the first two cases the null hypothesis is $f_t(\boldsymbol{\theta}_3) \equiv 0$, and it is assumed that this identity holds if and only if $\boldsymbol{\theta}_3 = \mathbf{0}$. In the first case one can thus test the hypothesis that the number of transitions in g_t equals r, whereas it is greater than r in the alternative. In the second case one can test a GARCH(p, q) model against either a GARCH$(p + s, q)$ or GARCH$(p, q + s)$, $s > 0$, model. Likewise, testing a linear GARCH model such as a standard GARCH but also a GJR-GARCH model against a smooth transition GARCH model fits into this framework. In the third case $f_t(\boldsymbol{\theta}_3)$ is a positive-valued function such that under the null hypothesis $f_t(\boldsymbol{\theta}_3) \equiv 1$, and this occurs if and only if $\boldsymbol{\theta}_3 = \mathbf{0}$. This design covers the case in which under the null hypothesis $\{\zeta_t\}$ is a sequence of independent standard normal variables as in (1). Under the alternative, ζ_t has an ARCH structure. This implies $\zeta_t = z_t f_t^{1/2}$, where $\{z_t\}$ is a sequence of independent standard normal variables and

$$f_t = 1 + \sum_{j=0}^{k} \nu_j \zeta_{t-j}^2 \tag{18}$$

so $\boldsymbol{\theta}_3 = (\nu_1, ..., \nu_k)'$ with $\nu_j \geq 0$, $j = 1, ..., k$, and at least one inequality is strict. Another possibility is to assume $\varepsilon_t = z_t(h_t g_t f_t)^{1/2}$, where $f_t = 1 + \sum_{j=0}^{k} \nu_j y_{t-j}^2$ such that y_t is a positive-valued stationary stochastic variable, weakly exogenous to the parameters of interest. An interesting special case may be the one in which $g_t \equiv 1$, and the multiplicative structure of the model opens up another way of introducing exogenous information to the model.

In order to consider the first testing problem, assume that the null model is a stationary TV-GJR-GARCH model in which the GARCH component is of order one ($p = q = 1$) in (4). The latter assumption is made for notational convenience, and a higher-order GARCH component is not excluded from consideration.

In Case 1, the log-likelihood for observation t has the form

$$\begin{aligned} \ell(\boldsymbol{\theta},\varepsilon_t) &= k - (1/2)\{\ln h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) + \ln\{g_t(\boldsymbol{\theta}_1) + f_t(\boldsymbol{\theta}_3)\} \\ &\quad + \frac{\varepsilon_t^2}{h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)\{g_t(\boldsymbol{\theta}_1) + f_t(\boldsymbol{\theta}_3)\}}\}. \end{aligned}$$

In order to shorten the notation, set $h_t = h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$, $g_t = g_t(\boldsymbol{\theta}_1)$ and $f_t = f_t(\boldsymbol{\theta}_3)$. The average score vector has the form

$$\mathbf{s}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = (\mathbf{s}_1(\boldsymbol{\theta}, \boldsymbol{\varepsilon})', \mathbf{s}_2(\boldsymbol{\theta}, \boldsymbol{\varepsilon})')' \tag{19}$$

where $\mathbf{s}_1(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = (\mathbf{s}_{11}(\boldsymbol{\theta}, \boldsymbol{\varepsilon})', \mathbf{s}_{12}(\boldsymbol{\theta}, \boldsymbol{\varepsilon})')'$. The components of (19) are

$$\begin{aligned} \mathbf{s}_{11}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) &= (2T)^{-1} \sum_{t=1}^{T} \frac{\partial \ell(\boldsymbol{\theta},\varepsilon_t)}{\partial \boldsymbol{\theta}_1} = (2T)^{-1} \sum_{t=1}^{T} (\zeta_t^2 - 1)(\frac{1}{(g_t + f_t)} \frac{\partial g_t}{\partial \boldsymbol{\theta}_1} + \frac{1}{h_t} \frac{\partial h_t}{\partial \boldsymbol{\theta}_1}) \\ \mathbf{s}_{12}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) &= (2T)^{-1} \sum_{t=1}^{T} \frac{\partial \ell(\boldsymbol{\theta},\varepsilon_t)}{\partial \boldsymbol{\theta}_2} = (2T)^{-1} \sum_{t=1}^{T} (\zeta_t^2 - 1) \frac{1}{h_t} \frac{\partial h_t}{\partial \boldsymbol{\theta}_2} \end{aligned}$$

and

$$\mathbf{s}_2(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = (2T)^{-1} \sum_{t=1}^{T} \frac{\partial \ell(\boldsymbol{\theta},\varepsilon_t)}{\partial \boldsymbol{\theta}_3} = (2T)^{-1} \sum_{t=1}^{T} (\zeta_t^2 - 1)(\frac{1}{(g_t + f_t)} \frac{\partial f_t}{\partial \boldsymbol{\theta}_3} + \frac{1}{h_t} \frac{\partial h_t}{\partial \boldsymbol{\theta}_3})$$

where $\zeta_t^2 = \varepsilon_t^2/\{h_t(g_t + f_t)\}$ and $f_t = 0$ if and only if $\boldsymbol{\theta}_3 = \mathbf{0}$. The partial derivatives of h_t are as follows:

$$
\begin{aligned}
\frac{\partial h_t}{\partial \boldsymbol{\theta}_1} &= -\{\alpha_1 \frac{\varepsilon_{t-1}^2}{(g_{t-1}+f_{t-1})^2} + \gamma_1 \frac{\varepsilon_{t-1}^2 I(\varepsilon_{t-1}<0)}{(g_{t-1}+f_{t-1})^2}\} \frac{\partial g_{t-1}}{\partial \boldsymbol{\theta}_1} + \beta_1 \frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_1} \\
\frac{\partial h_t}{\partial \boldsymbol{\theta}_2} &= \frac{\partial}{\partial \boldsymbol{\theta}_2}\{\alpha_0 + \alpha_1 \frac{\varepsilon_{t-1}^2}{g_{t-1}+f_{t-1}} + \gamma_1 \frac{\varepsilon_{t-1}^2 I(\varepsilon_{t-1}<0)}{g_{t-1}+f_{t-1}} + \beta_1 h_{t-1}\} \\
&= (1, \frac{\varepsilon_{t-1}^2}{g_{t-1}+f_{t-1}}, \frac{\varepsilon_{t-1}^2 I(\varepsilon_{t-1}<0)}{g_{t-1}+f_{t-1}}, h_{t-1})' + \beta_1 \frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_2}
\end{aligned}
$$

and

$$
\frac{\partial h_t}{\partial \boldsymbol{\theta}_3} = -(\alpha_1 \frac{\varepsilon_{t-1}^2}{(g_{t-1}+f_{t-1})^2} + \gamma_1 \frac{\varepsilon_{t-1}^2 I(\varepsilon_{t-1}<0)}{(g_{t-1}+f_{t-1})^2}) \frac{\partial f_{t-1}}{\partial \boldsymbol{\theta}_3} + \beta_1 \frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_3}. \qquad (20)
$$

Furthermore, $\partial g_t/\partial \boldsymbol{\theta}_1$ is defined in Lemma A.1. When g_t with r transitions is tested against $r+1$, $\partial f_t/\partial \boldsymbol{\theta}_3 = \mathbf{t}^* = (t^*, t^{*2}, t^{*3})'$. This follows from the fact that the $(r+1)$st transition function is approximated by its third-order Taylor expansion around $\gamma_{r+1} = 0$, see Eitrheim and Teräsvirta (1996) and Lundbergh and Teräsvirta (2002) for discussion.

Setting $\widehat{h}_t = h_t(\widehat{\boldsymbol{\theta}}_{1T}, \widehat{\boldsymbol{\theta}}_{2T}, \mathbf{0})$ and $\widehat{g}_t = g_t(\widehat{\boldsymbol{\theta}}_{1T})$, and evaluating the average score under H_0: $\boldsymbol{\theta}_3 = \mathbf{0}$ yields

$$
\begin{aligned}
\mathbf{s}_{11}(\widehat{\boldsymbol{\theta}}_{1T}, \widehat{\boldsymbol{\theta}}_{2T}, \mathbf{0}) &= (2T)^{-1} \sum_{t=1}^{T} (\widehat{\zeta}_t^2 - 1)(\frac{1}{\widehat{g}_t} \frac{\partial g_t}{\partial \boldsymbol{\theta}_1}|_{\mathrm{H}_0} + \frac{1}{\widehat{h}_t} \frac{\partial h_t}{\partial \boldsymbol{\theta}_1}|_{\mathrm{H}_0}) \qquad (21) \\
\mathbf{s}_{12}(\widehat{\boldsymbol{\theta}}_{1T}, \widehat{\boldsymbol{\theta}}_{2T}, \mathbf{0}) &= (2T)^{-1} \sum_{t=1}^{T} (\widehat{\zeta}_t^2 - 1) \frac{1}{\widehat{h}_t} \frac{\partial h_t}{\partial \boldsymbol{\theta}_2}|_{\mathrm{H}_0} \qquad (22)
\end{aligned}
$$

and

$$
\mathbf{s}_2(\widehat{\boldsymbol{\theta}}_{1T}, \widehat{\boldsymbol{\theta}}_{2T}, \mathbf{0}) = (2T)^{-1} \sum_{t=1}^{T} (\widehat{\zeta}_t^2 - 1)(\frac{1}{\widehat{g}_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}_3}|_{\mathrm{H}_0} + \frac{1}{\widehat{h}_t} \frac{\partial h_t}{\partial \boldsymbol{\theta}_3}|_{\mathrm{H}_0})
$$

where $\widehat{\zeta}_t^2 = \varepsilon_t^2/(\widehat{h}_t \widehat{g}_t)$, and $\widehat{\boldsymbol{\theta}}_{1T} \to \boldsymbol{\theta}_1^0$ and $\widehat{\boldsymbol{\theta}}_{2T} \to \boldsymbol{\theta}_2^0$ in probability as $T \to \infty$. The partial derivatives, evaluated at H_0, are as follows:

$$
\begin{aligned}
\frac{\partial h_t}{\partial \boldsymbol{\theta}_1}|_{\mathrm{H}_0} &= -\{\widehat{\alpha}_1 + \widehat{\gamma}_1 I(\varepsilon_{t-1}<0)\} \frac{\widehat{\phi}_{t-1}^2}{\widehat{g}_{t-1}} \frac{\partial g_{t-1}}{\partial \boldsymbol{\theta}_1}|_{\mathrm{H}_0} + \widehat{\beta}_1 \frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_1}|_{\mathrm{H}_0} \\
\frac{\partial h_t}{\partial \boldsymbol{\theta}_2}|_{\mathrm{H}_0} &= (1, \widehat{\phi}_{t-1}^2, \widehat{\phi}_{t-1}^2 I(\varepsilon_{t-1}<0), \widehat{h}_{t-1})' + \widehat{\beta}_1 \frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_2}|_{\mathrm{H}_0}
\end{aligned}
$$

where $\widehat{\phi}_t^2 = \varepsilon_t^2/\widehat{g}_t$, and

$$
\frac{\partial h_t}{\partial \boldsymbol{\theta}_3}|_{\mathrm{H}_0} = -\{\widehat{\alpha}_1 + \widehat{\gamma}_1 I(\varepsilon_{t-1}<0)\} \frac{\widehat{\phi}_{t-1}^2}{\widehat{g}_{t-1}} \frac{\partial f_{t-1}}{\partial \boldsymbol{\theta}_3}|_{\mathrm{H}_0} + \widehat{\beta}_1 \frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_3}|_{\mathrm{H}_0}.
$$

where $(\partial f_t/\partial \boldsymbol{\theta}_3)|_{\mathrm{H}_0} = \partial f_t/\partial \boldsymbol{\theta}_3 = \mathbf{t}^*$. Denoting

$$
\mathbf{B}(\boldsymbol{\theta}^0) = \mathsf{E}\mathbf{s}(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0, \mathbf{0})\mathbf{s}(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0, \mathbf{0})' = \begin{bmatrix} \mathbf{B}_{11}(\boldsymbol{\theta}^0) & \mathbf{B}_{12}(\boldsymbol{\theta}^0) \\ \mathbf{B}_{21}(\boldsymbol{\theta}^0) & \mathbf{B}_{22}(\boldsymbol{\theta}^0) \end{bmatrix}
$$

we have that

$$
T\mathbf{s}_2(\widehat{\boldsymbol{\theta}}_{1T}, \widehat{\boldsymbol{\theta}}_{2T}, \mathbf{0})'(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\mathbf{s}_2(\widehat{\boldsymbol{\theta}}_{1T}, \widehat{\boldsymbol{\theta}}_{2T}, \mathbf{0}) \xrightarrow{D} \chi^2(m)
$$

when $\boldsymbol{\theta}_3 = \mathbf{0}$, where m is the dimension of $\boldsymbol{\theta}_3$. Set $\widehat{\mathbf{x}}_{jt} = \widehat{h}_t^{-1}(\partial h_t/\partial \boldsymbol{\theta}_j)|_{\mathrm{H}_0}$, $j = 1,2,3$, $\widehat{\mathbf{u}}_{1t} = \widehat{g_t}^{-1}(\partial g_t/\partial \boldsymbol{\theta}_1)|_{\mathrm{H}_0}$ and $\widehat{\mathbf{v}}_{3t} = \widehat{g_t}^{-1}(\partial f_t/\partial \boldsymbol{\theta}_3)|_{\mathrm{H}_0}$. Furthermore,

$$\boldsymbol{\Sigma}_{ij} = \frac{\mu_4 - 1}{4T}\sum_{t=1}^{T}\widehat{\mathbf{r}}_{it}\widehat{\mathbf{r}}_{jt}', \ i,j = 1,2$$

is a consistent estimator of $\mathbf{B}_{ij}(\boldsymbol{\theta}^0)$ under H_0, where $\widehat{\mathbf{r}}_{1t} = (\widehat{\mathbf{u}}_{1t}' + \widehat{\mathbf{x}}_{1t}', \widehat{\mathbf{x}}_{2t}')'$ and $\widehat{\mathbf{r}}_{2t} = \widehat{\mathbf{v}}_{3t} + \widehat{\mathbf{x}}_{3t}$.

The test can be carried out in stages:

1. Estimate the TV-GJR-GARCH model, save the standardised residuals $\widehat{\zeta}_t$ and construct the 'residual sum of squares' $SSR_0 = \sum_{t=1}^{T}(\widehat{\zeta}_t^2 - 1)^2$.
2. Regress $\widehat{\zeta}_t^2 - 1$ on $\widehat{\mathbf{r}}_{1t}$ and $\widehat{\mathbf{r}}_{2t}$ and form the residual sum of squares SSR_1.
3. Compute the test statistic
$$LM_{\text{add-g}} = T\frac{SSR_0 - SSR_1}{SSR_0}. \tag{23}$$
The statistic (23) has an asymptotic χ^2-distribution with $\dim(\boldsymbol{\theta}_3)$ degrees of freedom when the null hypothesis is valid.

An interesting special case of this test is obtained by assuming $h_t = \alpha_0 > 0$. It may be relevant in situations, in which the error variance of a macroeconometric model can be time-varying but without volatility clustering. The great moderation serves as a recent example. Note, however, that in this case the conditional mean is being modelled as well, whereas it was assumed away in Section 2.

In Case 2, the log-likelihood for observation t has the form

$$\ell(\boldsymbol{\theta},\varepsilon_t) = k - (1/2)[\ln\{h_t(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) + f_t(\boldsymbol{\theta}_3)\} + \ln g_t(\boldsymbol{\theta}_1) + \frac{\varepsilon_t^2}{\{h_t(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) + f(\boldsymbol{\theta}_3)\}g_t(\boldsymbol{\theta}_1)}]$$

and the components of the score are

$$\begin{aligned}
\mathbf{s}_{11}(\boldsymbol{\theta},\boldsymbol{\varepsilon}) &= (2T)^{-1}\sum_{t=1}^{T}\frac{\partial \ell(\boldsymbol{\theta},\varepsilon_t)}{\partial \boldsymbol{\theta}_1} = (2T)^{-1}\sum_{t=1}^{T}(\zeta_t^2 - 1)(\frac{1}{g_t}\frac{\partial g_t}{\partial \boldsymbol{\theta}_1} + \frac{1}{(h_t + f_t)}\frac{\partial h_t}{\partial \boldsymbol{\theta}_1}) \\
\mathbf{s}_{12}(\boldsymbol{\theta},\boldsymbol{\varepsilon}) &= (2T)^{-1}\sum_{t=1}^{T}\frac{\partial \ell(\boldsymbol{\theta},\varepsilon_t)}{\partial \boldsymbol{\theta}_2} = (2T)^{-1}\sum_{t=1}^{T}(\zeta_t^2 - 1)\frac{1}{h_t + f_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}_2}
\end{aligned}$$

and

$$\mathbf{s}_2(\boldsymbol{\theta},\boldsymbol{\varepsilon}) = (2T)^{-1}\sum_{t=1}^{T}\frac{\partial \ell(\boldsymbol{\theta},\varepsilon_t)}{\partial \boldsymbol{\theta}_3} = (2T)^{-1}\sum_{t=1}^{T}(\zeta_t^2 - 1)\frac{1}{(h_t + f_t)}\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}$$

where

$$\begin{aligned}
\frac{\partial h_t}{\partial \boldsymbol{\theta}_1} &= \frac{\partial}{\partial \boldsymbol{\theta}_1}\{\alpha_0 + \alpha_1\frac{\varepsilon_{t-1}^2}{g_{t-1}} + \gamma_1\frac{\varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0)}{g_{t-1}} + \beta_1 h_{t-1}\} \\
&= -\{\alpha_1\frac{\varepsilon_{t-1}^2}{g_{t-1}^2} + \gamma_1\frac{\varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0)}{g_{t-1}^2}\}\frac{\partial g_{t-1}}{\partial \boldsymbol{\theta}_1} + \beta_1\frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_1} \\
\frac{\partial h_t}{\partial \boldsymbol{\theta}_2} &= (1, \frac{\varepsilon_{t-1}^2}{g_{t-1}}, \frac{\varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0)}{g_{t-1}}, h_{t-1})' + \beta_1\frac{\partial h_t}{\partial \boldsymbol{\theta}_2}.
\end{aligned}$$

Analogously, $\zeta_t^2 = \varepsilon_t^2/\{(h_t + f_t)g_t\}$. The partial derivatives, evaluated at H_0, are

$$\begin{aligned}\frac{\partial h_t}{\partial \boldsymbol{\theta}_1}|_{\mathrm{H}_0} &= -(\widehat{\alpha}_1\widehat{\phi}_1^2 + \widehat{\gamma}_1\widehat{\phi}_1^2 I(\varepsilon_{t-1} < 0))\frac{\partial g_{t-1}}{\partial \boldsymbol{\theta}_1}|_{\mathrm{H}_0} + \widehat{\beta}_1\frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_1}|_{\mathrm{H}_0} \\ \frac{\partial h_t}{\partial \boldsymbol{\theta}_2}|_{\mathrm{H}_0} &= (1, \widehat{\phi}_{t-1}^2, \widehat{\phi}_{t-1}^2 I(\varepsilon_{t-1} < 0), \widehat{h}_{t-1})' + \widehat{\beta}_1\frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_2}|_{\mathrm{H}_0}\end{aligned}$$

whereas $\partial f_t/\partial \boldsymbol{\theta}_3$ depends on the alternative hypothesis. For example, if the alternative hypothesis is that $p = 2$ in the GJR-GARCH model, then $f_t = \beta_2 h_{t-2}$, so $\boldsymbol{\theta}_3 = \beta_2$. The blocks $\mathbf{s}_{11}(\widehat{\boldsymbol{\theta}}_{1T}, \widehat{\boldsymbol{\theta}}_{2T}, \mathbf{0})$ and $\mathbf{s}_{12}(\widehat{\boldsymbol{\theta}}_{1T}, \widehat{\boldsymbol{\theta}}_{2T}, \mathbf{0})$ of the average score are the same as (21) and (22), whereas

$$\mathbf{s}_2(\widehat{\boldsymbol{\theta}}_{1T}, \widehat{\boldsymbol{\theta}}_{2T}, \mathbf{0}) = (2T)^{-1}\sum_{t=1}^{T}(\widehat{\zeta}_t^2 - 1)\frac{1}{\widehat{h}_t}\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}|_{\mathrm{H}_0}.$$

The test is carried out in stages as in the previous case, the only difference being that $\widehat{\mathbf{r}}_{2t} = \widehat{h}_t^{-1}(\partial f_t/\partial \boldsymbol{\theta}_3)|_{\mathrm{H}_0}$. To continue the example, if $f_t = \beta_2 h_{t-2}$, then $(\partial f_t/\partial \boldsymbol{\theta}_3)|_{\mathrm{H}_0} = \widehat{h}_{t-2}$. When $g_t \equiv 1$, $\mathbf{s}_{11}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \mathbf{0}$, and the test collapses into the corresponding misspecification test of the (GJR-)GARCH model in Lundbergh and Teräsvirta (2002).

As for Case 3, the components of the average score are

$$\begin{aligned}\mathbf{s}_{11}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) &= (2T)^{-1}\sum_{t=1}^{T}\frac{\partial \ell(\boldsymbol{\theta},\varepsilon_t)}{\partial \boldsymbol{\theta}_1} = (2T)^{-1}\sum_{t=1}^{T}(z_t^2 - 1)(\frac{1}{g_t}\frac{\partial g_t}{\partial \boldsymbol{\theta}_1} + \frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}_1}) \\ \mathbf{s}_{12}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) &= (2T)^{-1}\sum_{t=1}^{T}\frac{\partial \ell(\boldsymbol{\theta},\varepsilon_t)}{\partial \boldsymbol{\theta}_2} = (2T)^{-1}\sum_{t=1}^{T}(z_t^2 - 1)\frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}_2}\end{aligned}$$

and

$$\mathbf{s}_2(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = (2T)^{-1}\sum_{t=1}^{T}\frac{\partial \ell(\boldsymbol{\theta},\varepsilon_t)}{\partial \boldsymbol{\theta}_3} = (2T)^{-1}\sum_{t=1}^{T}(z_t^2 - 1)\frac{1}{f_t}\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}$$

where $z_t^2 = \varepsilon_t^2/(h_t f_t g_t)$. Evaluated under H_0, $\mathbf{s}_{11}(\boldsymbol{\theta}, \boldsymbol{\varepsilon})$ and $\mathbf{s}_{12}(\boldsymbol{\theta}, \boldsymbol{\varepsilon})$ again equal (21) and (22), whereas

$$\mathbf{s}_2(\widehat{\boldsymbol{\theta}}_{1T}, \widehat{\boldsymbol{\theta}}_{2T}, \mathbf{0}) = (2T)^{-1}\sum_{t=1}^{T}(\widehat{\zeta}_t^2 - 1)\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}|_{\mathrm{H}_0}$$

since $z_t \equiv \zeta_t$ and $f_t \equiv 1$ under H_0. The test can be carried out in stages as before, and now $\widehat{\mathbf{r}}_{2t} = (\partial f_t/\partial \boldsymbol{\theta}_3)|_{\mathrm{H}_0}$. When the alternative is 'ARCH in GARCH', $\widehat{\mathbf{r}}_{2t} = (\widehat{\zeta}_{t-1}^2, ..., \widehat{\zeta}_{t-k}^2)'$, and the asymptotic χ^2-statistic has k degrees of freedom. Even here, $g_t \equiv 1$ implies $\mathbf{s}_{11}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \mathbf{0}$, in which case the test collapses into the 'ARCH in GARCH' test in Lundbergh and Teräsvirta (2002), which is asymptotically equivalent to the test of Li and Mak (1994). If, in addition, $h_t \equiv \alpha_0$, we have the test of no ARCH of Engle (1982) which in turn is asymptotically equivalent to the portmanteau test of of McLeod and Li (1983).

6 Simulation study

6.1 Design of the study

In this section, we conduct a small simulation experiment to evaluate the finite-sample properties of the test against a multiplicative TV-GARCH specification. Specifically, we shall investigate the size and power properties of the LM-type test involved in the modelling strategy as well as the success rate of the specification procedures. Sample lengths of 1000, 2500 and 5000 observations are used in all simulations. To avoid the initialization effects, the first 1000 observations have been discarded before generating the actual series. All the computations have been carried out using Ox, version 5.00 (see Doornik (2007)). The behaviour of the test statistics is examined for several data generating processes (DGPs) that can be nested in the following TV-GARCH specification:

$$y_t = \zeta_t (h_t g_t)^{1/2}, \qquad \{\zeta_t\} \sim \text{iid}\mathcal{N}(0,1) \tag{24}$$

The data generating processes are as follows:

DGP (i) $\quad h_t = 0.10 + 0.10\varepsilon_{t-1}^2 + 0.80h_{t-1};\ g_t \equiv 1$

DGP (ii) $\quad h_t = 0.10 + 0.10\varepsilon_{t-1}^2 + 0.85h_{t-1};\ g_t \equiv 1$

DGP (iii) $\quad h_t = 0.05 + 0.05\varepsilon_{t-1}^2 + 0.90h_{t-1};\ g_t \equiv 1$

DGP (iv) $\quad h_t = 0.005 + \{0.05 + 0.10I(\varepsilon_{t-1} < 0)\}\varepsilon_{t-1}^2 + 0.80h_{t-1};\ g_t \equiv 1$

DGP (v) $\quad h_t = 0.10 + 0.10\varepsilon_{t-1}^2 + 0.80h_{t-1}$

$$g_t = 1 + \delta_1 G_1(t^*; \gamma_1, c_1),\ \delta_1 = \{-0.05, 0.05\}$$

$$G_1(t^*; \gamma_1, c_1) = (1 + \exp(-\gamma_1(t^* - c_1)))^{-1}$$

The last design concerns the multiplicative TV-GARCH model where the midpoint of the change in volatility is at $c_1 = 0.5$, whereas the slope parameter γ_1 varies in the interval $\gamma_1 = \{5, 10\}$. Following the suggestion in Bollerslev (1986), recursive computation of h_t is initialized by using the estimated unconditional variance for the pre-sample values $t \leq 0$.

6.2 Results of the study

In this section we shall first report results on the size and power properties of our parameter constancy tests. Then we turn to the results on the performance of our specification strategy.

6.2.1 Size and power simulations

Results of the size simulations are presented graphically in Figure 2. The graphs show the discrepancies in size (the actual minus the nominal size) plotted against the nominal sizes from $0.1\%, 0.3\%, 0.5\%, ..., 10\%$. In each subgraph we present the size discrepancies for DGPs (i)-(iv) and the three sample sizes. In most panels, one can observe that the parameter constancy test is positively size-distorted at the sample size $T = 1000$, but the empirical size becomes more accurate as the sample size increases. For sample sizes typically used for modelling volatility clustering, such as $T = 2500$ and $T = 5000$, the tests are reasonably well-sized. Furthermore, we also investigated the size results of the test statistics when the errors are not identically distributed. The size distortions in the robust version of the test were similar to those reported in Figure 2 and are not presented here. Our main conclusion is that both the non-robust and robust versions of the test statistics are rather good approximations to the finite-sample distributions for $T \geq 2500$.

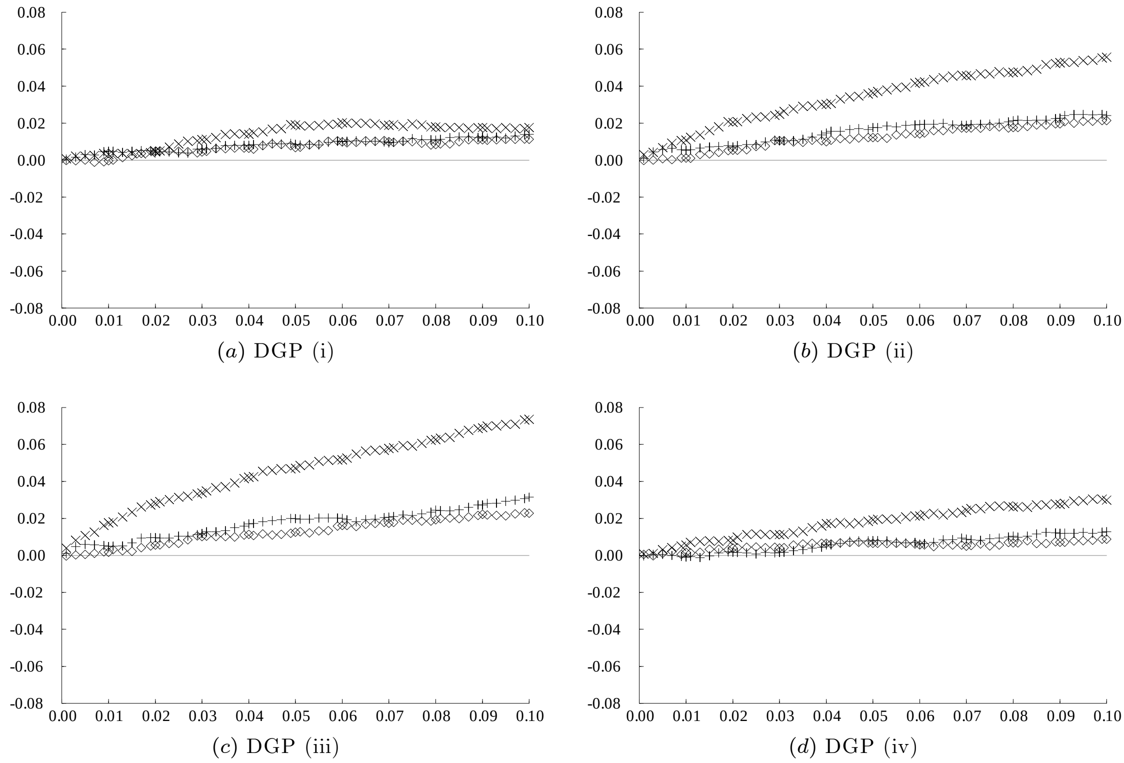


Figure 2. Size discrepancy plots of the parameter constancy test. The size discrepancy is plotted against the nominal size. The lines in each plot correspond to sample sizes 1000 ($\times$), 2500 ($+$), and 5000 ($\Diamond$). The total number of replications equals 5000.

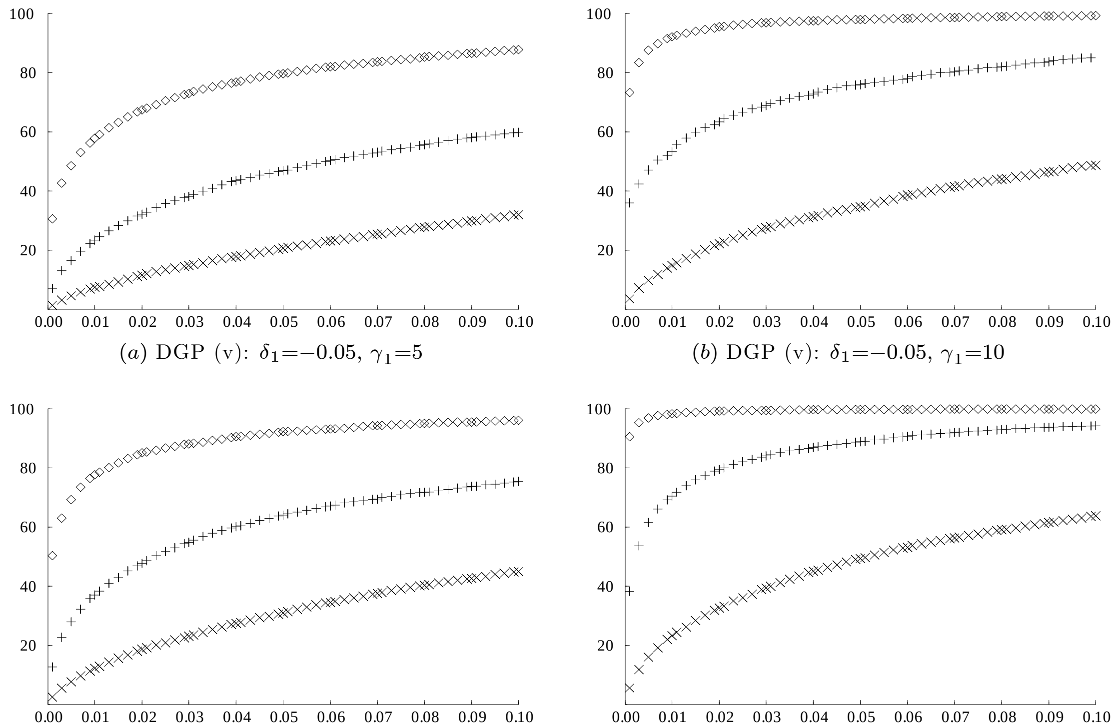
Although there exist several parameter constancy tests in the GARCH literature, none of them can be considered a direct benchmark for our parameter constancy tests. In what follows, we only show results of the power simulations for our test. On each graph of Figure 3, the actual rejection frequencies are plotted against the nominal significance levels $0.1\%, 0.3\%, 0.5\%, ..., 10\%$. Instead of the size-adjusted power curves suggested by Davidson and MacKinnon (1998), we simply report power curves as the tests have good size properties.

Rejection frequencies for the non-robust LM-type test against a multiplicative alternative are shown in Figure 3. The power results have been obtained by generating artificial data from DGP (v). As before, we consider sample sizes of 1000, 2500 and 5000 observations. The rejection frequencies of the LM test statistics shown in the left panel are moderate when $T = 1000$ and increase with the sample size. As expected, the rejection frequencies are an increasing function of the sample size and of the slope parameter γ_1. It is interesting to note that, the LM-type test statistic turns out to be quite powerful even for small values of the parameter δ_1 and short time series.

Again, we compute both the ordinary and robustified versions of the LM test. The behaviour of the robustified version of the test in the power simulations is similar to that of the non-robust version and results are not shown here.

6.2.2 Simulating the model selection strategy

In this section we consider the performance of the specific-to-general specification strategy for the multiplicative TV-GARCH models. This is done by studying the selection frequencies of various models. The specification procedure has been discussed in Section 4. For each DGP, we consider all three sample sizes. As before, the first 1000 observations of each generated series are discarded to minimise the initialization effects. Throughout, we set $\alpha = 0.05$ for both the LM_1 ($\partial f_t/\partial \boldsymbol{\theta}_3 = t^*$ in (20)) and LM_3 ($\partial f_t/\partial \boldsymbol{\theta}_3 = (t^*, t^{*2}, t^{*3})'$ in (20)) versions of the test. The maximum number of transitions considered equals two. Furthermore, $\tau = 1/2$, which means that we halve the significance level of the test at each stage of the sequence.



(a) DGP (v): δ_1=−0.05, γ_1=5

(b) DGP (v): δ_1=−0.05, γ_1=10

(c) DGP (v): δ_1=0.05, γ_1=5

(d) DGP (v): δ_1=0.05, γ_1=10

Figure 3. Power curves of the parameter constancy test. The actual rejection frequency is plotted against the nominal size. The lines in each plot correspond to sample sizes 1000 ($\times$), 2500 ($+$), and 5000 ($\Diamond$). The total number of replications equals 5000.

Results for DGPs (i)-(iv) are reported in Table 1. For each DGP, the total of number of replications equals 5000. The frequencies (in percentages) of the correct number of transitions are shown in boldface. The second column refers to the number of transition functions selected. In general, the statistic LM_1 (has better size properties than LM_3. However, in most cases, the test based on the third-order Taylor expansion also has an empirical size very close to the nominal size except when the sum $\alpha_1 + \beta_1$ is close to one and the sample size is less than 2500 observations.

Results for series generated from a model with a single transition function (DGP (v)) can be found in Table 2. The total number of replications equals 2000. The results concern the case where the centre of the change is located halfway through the sample. Clearly, the constant-parameter GARCH model is chosen too often for parameterizations with smoothest changes and shortest series. As T becomes large, the correct model is chosen more frequently, as may be expected. For large sample sizes, the selection frequencies of the true model become quite high even for very smooth changes. It also becomes easier to identify a single transition when the slope parameter γ_1 increases. Again, the LM_1-test has higher power than LM_3. As expected, the correct model is selected slightly more often for high than for low values of δ_1. Overall, the specification strategy seems to work relatively well for all combinations of parameters considered and for sample sizes $T \geq 2500$.

7 Applications

In this section we shall consider two empirical examples, each of them involving a financial time series. The series in the first example is the Standard and Poor 500 composite index (S&P 500), while the second application concerns the spot exchange rate of the Singapore dollar versus the U.S. dollar (SPD/USD). Both series are observed at a daily frequency and transformed into continuously compounded rates of return.

7.1 Stock index returns

The daily S&P 500 return series originates from the Yahoo-Quotes database. The sample extends from January 2, 1990, to December 31, 1999, which amounts to 2531 observations. The series is plotted in Figure 4. It contains a period of large volatility in the beginning and another at the end of the sample period, whereas the average volatility in the middle of the sample is somewhat lower than in both ends. Volatility clustering can be observed throughout the period.

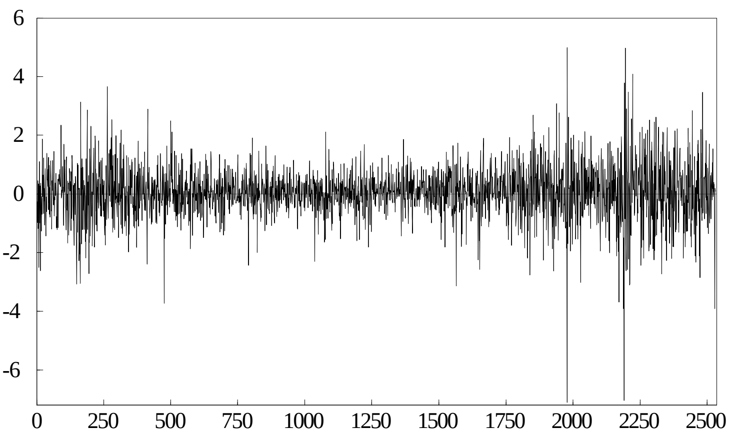


Figure 4. Daily returns of the S&P 500 composite index from January 2, 1990 until December 31, 1999 (2531 observations).

Summary statistics for the series can be found in the second column of Table 3. It is seen that there is both negative skewness and excess kurtosis in the series. Normality of the marginal distribution of the S&P 500 returns is strongly rejected. Robust skewness and kurtosis estimates (see Kim and White (2004) and Teräsvirta and Zhao (in press)) are also provided. The robust skewness measure is positive but very close to zero, which suggests that the asymmetry of the empirical distribution of the returns is due to a small number of outliers. The robust centred kurtosis that has value zero for the normal distribution indicates some excess kurtosis but much less than the conventional measure. This is in line with the robust skewness estimate. As expected, the null hypothesis of no ARCH is strongly rejected.

Since our returns are stock index returns, we begin by fitting a GJR-GARCH(1,1) model to this series. The results can be found in the leftmost columns of Table 4. It is seen that the heteroskedasticity is extremely persistent: the persistence measure $\widehat{\alpha}_1+\widehat{\beta}_1+\widehat{\gamma}_1/2=0.992$: note that the unconditional variance ceases to exist when $\alpha_1+\beta_1+\gamma_1/2=1$. Misspecification tests of this model (other than tests of $g_t\equiv 1$) can be found in Table 6. Both the test against a higher-order GARCH and the one against smooth transition GARCH model strongly reject the model. The latter rejection seems to suggest that the GJR-GARCH model fails to adequately describe the asymmetry of the response of the conditional variance to shocks.

Instead of respecifying this model following the outcomes of these tests we turn our attention to the tests of g_t (type 2 additive misspecification). This is Step 1 in the specification of multiplicative TV-GJR-GARCH models outlined in Section 4. The results can be found in Table 5. The null hypothesis $g_t\equiv 1$ is rejected very strongly as the p-value of the test equals 7×10^{-4}. The test sequence for specifying the structure of the deterministic function g_t points towards $K=2$, as H_{02} is rejected more strongly than either H_{01} or H_{03}. Fitting the TV-GJR-GARCH model with a single transition function and $K=2$ to the series and testing for another transition leads to rejecting the hypothesis of a single transition. The p-value, however, is now considerably larger, equalling 0.023, and the specification test sequence clearly suggests $K=1$ for the second transition. Accepting this outcome, fitting the corresponding TV-GJR-GARCH model to the series and testing for yet another transition yields the p-value 0.497 for the test. It can be concluded that no more than two transitions are needed to adequately characterise the deterministic component of the TV-GJR-GARCH model.

The estimated g_t has the following form:

$$\widehat{g}_t=\{1+\underset{(0.1216)}{1.6034}G_1(t^*;\widehat{\gamma}_1,\widehat{\mathbf{c}}_1)+\underset{(0.3429)}{1.7378}G_2(t^*;\widehat{\gamma}_2,\widehat{c}_2)\} \tag{25}$$

with

$$G_1(t^*;\widehat{\gamma}_1,\widehat{\mathbf{c}}_1)=(1+\exp\{-\underset{(-)}{250}(t^*-\underset{(0.0009)}{0.2055})(t^*-\underset{(0.0031)}{0.6918})\})^{-1} \tag{26}$$

and

$$G_2(t^*;\widehat{\gamma}_2,\widehat{c}_2)=(1+\exp\{-\underset{(-)}{250}(t^*-\underset{(0.0026)}{0.8540})\})^{-1} \tag{27}$$

The graph of $\widehat{g}_t$ is depicted in Figure 6. The two transitions are clearly visible and illustrate how volatility first decreases and then increases over time. A double increase at the end suggests that volatility there is even higher than in the beginning of the series.

The estimated model is subjected to misspecification tests described in Section 5. The results can be found in Table 6. The p-values of the two tests, the one against higher-order GARCH and the one against STGARCH are now considerably higher than previously, which suggests that the main source of misspecification in the GJR-GARCH model was the lack of the proper deterministic component. If we, somewhat arbitrarily, apply the 1% significance level for our tests, the estimated TV-GJR-GARCH model may be deemed adequate. If we, however, wanted to improve the model further, we could add another lag of h_t to the model.

It is of interest to compare the TV-GJR-GARCH model with the GJR-GARCH specification. The log-likelihood of the former model is considerably higher than that of the latter. The most dramatic change in the dynamic properties is that the persistence has decreased remarkably: in the TV-GJR-GARCH model, $\widehat{\alpha}_1+\widehat{\beta}_1+\widehat{\gamma}_1/2=0.917$, which is a strikingly low figure. It is seen that this decrease is mainly due to the decrease in the estimate of β_1, which is the coefficient of h_{t-1}. There is an increase in the value of

γ_1, which is partly attributed to the fact that the estimate of α_1 became statistically insignificant, it was in fact slightly negative, and the term $\varepsilon_{t-1}/g_{t-1}^{1/2}$ was therefore omitted from the model. These changes have the following explanation. When it is assumed that the process is stationary there is only one level (the unconditional variance) to which the conditional variance converges when it is assumed that $\zeta_t = 0$ for $t > t_0$. The convergence then takes a very long time ($\widehat{\alpha}_1 + \widehat{\beta}_1 + \widehat{\gamma}_1/2 = 0.992$ is very close to unity). In the TV-GJR-GARCH model the corresponding level is time-varying, and the rate of convergence to this flexible level can thus be much more rapid than it is in the standard GJR-GARCH model to a fixed level.

The fourth column in Table 3 contains the skewness and kurtosis estimates for $\varepsilon_t/\widehat{g}_t^{1/2}$. The negative skewness remains when measured using the standard nonrobust estimate but, as can be expected from the other results, the excess kurtosis of the final $\varepsilon_t/\widehat{g}_t^{1/2}$ series is considerably less (equal to 2.8) than the original number that equalled 5.3. This is another illustration of the fact that volatility to be modelled by h_t in the TV-GJR-GARCH model is much smaller than it is in the GJR-GARCH(1,1) model without the nonstationary component. Even the robust kurtosis estimate in Table 3 shows some decrease, but because its nonrobust value was already small, the decrease has remained rather moderate.

The differences between the two specifications are further illustrated by two figures. Figure 7 contains the autocorrelations of $|\varepsilon_t|$ in Panel (a) and those of $|\varepsilon_t|/\widehat{g}_t^{1/2}$ in Panel (b). It is seen that the increase in the log-likelihood is mainly due to a decrease in the general level of the autocorrelations. In Panel (a), the autocorrelations retain the 'long-memory property', the very slow decay as a function of the lag, which translates into the high value of the persistence measure in the GJR-GARCH model. Panel (b) shows that the autocorrelations of $|\varepsilon_t|/\widehat{g}_t^{1/2}$ are rather small, and only few of them exceed two standard deviations of $|\varepsilon_t|$ under the iid normality assumption, marked by the horizontal line in the figure. A major part of the variation in the daily S&P 500 return series can thus be attributed to the slow-moving component g_t, and only a minor portion of the variation is left to be parameterised by the stationary GJR-GARCH component. It may be argued that the unconditional variance component g_t completely dominates the conditional variance.

Figure 8 contains the estimated conditional standard deviations $h_t^{1/2}$ of $\{\varepsilon_t\}$ for the GJR-GARCH(1,1) model and the ones of $\{\varepsilon_t/\widehat{g}_t^{1/2}\}$. For the GJR-GARCH model, see Panel (a), the graph looks rather 'nonstationary' in that there appears to be a nonlinear 'trend'. From the graph in Panel (b) it is seen that volatility (the conditional standard deviation of $\{\varepsilon_t/\widehat{g}_t^{1/2}\}$) is still changing over time, but the persistent level changes have been absorbed by the deterministic component.

In Figure 9, the estimated news impact curve of the standard GJR-GARCH(1,1) model is compared with corresponding curves of the TV-GJR-GARCH(1,1) model. The news impact curve of the TV-GJR-GARCH model is time-varying because it depends on g_{t-1}. The news impact curve of the GJR-GARCH model is time-invariant, and from the figure it is seen how the curve can vary over time in the TV-GJR-GARCH model. This curve is completely flat for $\varepsilon_{t-1} > 0$ because $\alpha_1 = 0$ in the model. The curves based on the TV-GJR-GARCH model clearly show the obvious fact that when there is plenty of turbulence in the market, the news impact of a particular negative shock is smaller than it is when calm prevails. In the latter case, even a minor piece of 'bad news' (a negative shock) can be 'news', whereas in the former case, even a relatively large negative shock can have a rather small news component. This distinction cannot be made in the standard

GJR-GARCH model. According to our TV-GJR-GARCH model, 'good news' (positive shocks) have no impact on volatility in this application.

7.2 Exchange rate returns

The time series of this section consists of daily returns of the spot SPD/USD exchange rate provided by the Federal Reserve Bank of New York. A graph of the series is shown in Figure 5. It covers the period from May 1, 1997 until July 11, 2005, yielding a total of 2060 observations. At first sight, it appears that one can distinguish two different regimes in the series. A period of high volatility occurs during the East Asian financial crisis due to the significant depreciation of the Singapore dollar relative to the U.S. dollar. After the crisis, the volatility of the currency returns descends to a low level.

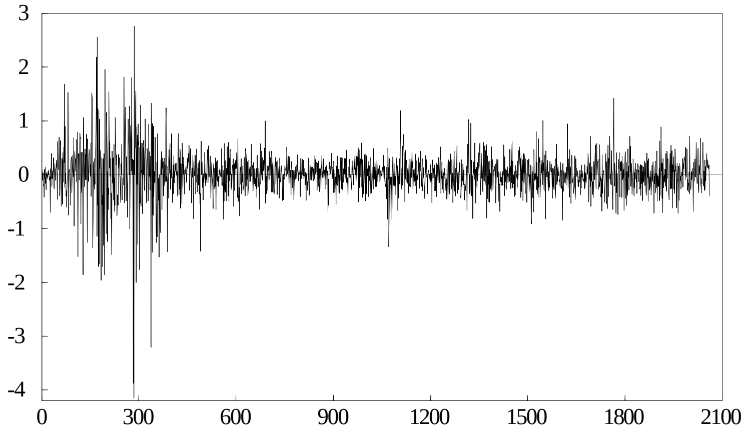


Figure 5. Daily returns of the Singapore Dollar versus US dollar exchange rate from May 1, 1997 until July 11, 2005 (2060 observations).

Descriptive statistics for the SPD/USD exchange rate returns are reported in Table 3. There is plenty of excess kurtosis, and the estimated skewness is strongly negative. These values are due to a limited number of large negative returns early in the series during the so-called Asian crisis. Naturally, the marginal distribution of the returns is far from normal. The robust measure of skewness indicates that there is in fact little skewness and the robust centred kurtosis is substantially smaller than its standard measure. The hypothesis of no ARCH is strongly rejected. The GARCH(1,1) model fitted to this exchange rate return series again shows high persistence of volatility. The estimate of α_1 is larger and that of β_1 smaller than in the S&P 500 model, which is a consequence of the fact that the kurtosis is larger in the exchange rate series than it is in the S&P 500 returns. The misspecification tests in Table 7 that does not contain the test of constant unconditional variance, do not generally indicate misspecification. The 'no ARCH in GARCH' test is the only exception when $r = 5$.

The specification test of constant unconditional variance against a time-varying one in Table 5 has the p-value equal to 0.002. The short test sequence indicates that one should choose $K = 2$, which at first may appear a bit surprising, given Figure 5. A TV-GARCH model with a single transition appears adequate in the sense that the test for another transition has $p = 0.22$.

The diagnostic tests of this model in Table 7 do not indicate misspecification. In particular, the linearity test against the smooth transition GARCH does not indicate remaining nonlinearity, which agrees with the commonly observed fact that a symmetric GARCH component is adequate for exchange rate return series. The model is thus accepted as our final model for the SPD/USD daily return series.

Turning to the properties of the estimated model, it is seen from Table 3 that the excess kurtosis of the rescaled returns $\{\varepsilon_t/\widehat{g}_t^{1/2}\}$ is substantially smaller than that of $\{\varepsilon_t\}$. Furthermore, negative skewness has been reduced from -0.9 to less than -0.3. This can be ascribed to the fact that the original skewness estimate was due to a couple of very large negative returns during the Asian crisis. Their significance has subsequently been reduced in rescaled returns. As may be expected, the robust skewness estimate is not affected by rescaling.

The estimated g_t has the following form:

$$\widehat{g}_t = \{1 - \underset{(0.0062)}{0.8184} G_1(t^*; \widehat{\gamma}_1, \widehat{\mathbf{c}}_1)\}, \tag{28}$$

where

$$G_1(t^*; \widehat{\gamma}_1, \widehat{\mathbf{c}}_1) = (1 + \exp\{-\underset{(-)}{250}(t^* - \underset{(0.0003)}{0.0227})(t^* - \underset{(0.0014)}{0.1888})\})^{-1} \tag{29}$$

The graph of the transition function can be found in Figure 10. It shows how the tranquil period just before the crisis in the cause for selecting $K = 2$, which then gives $\widehat{g}_t$ its hump shape.

The parameter estimates of the GARCH component of the TV-GARCH model appear in Table 4. While $\widehat{\alpha}_1 + \widehat{\beta}_1 = 0.994$ in the GARCH model, after introducing g_t the same sum equals 0.962, which suggests a remarkable drop in persistence. As in the previous example, the characteristics of the model are further illustrated by two figures. Figure 11 shows how the autocorrelations of $|\varepsilon_t|/\widehat{g}_t^{1/2}$ are considerably lower than those of $|\varepsilon_t|$. The first-order autocorrelation that equals 0.304 for $|\varepsilon_t|$ is only equal to 0.117 for $|\varepsilon_t|/\widehat{g}_t^{1/2}$.The decay rate of the autocorrelations of of $|\varepsilon_t|/\widehat{g}_t^{1/2}$ is quite rapid. Again, it seems that the unconditional variance component dominates, and the amount of conditional heteroskedasticity left to be modelled after the changing unconditional variance has been accounted for, is relatively small in comparison.

The graph of the conditional standard deviation $h_t^{1/2}$ in Panel (a) of Figure 12 clearly shows a long period of high volatility that is more than a simple cluster. Panel (b) shows that in the final model this period is explained by the deterministic component g_t, and that the graph of $h_t^{1/2}$ does not show signs of nonstationarity. This is precisely what one would expect after seeing the GARCH parameter estimates in Table 4.

Figure 13 contains the estimated news impact curves of the traditional GARCH(1,1) model and the ones of the TV-GARCH(1,1) model for three regimes. It is seen that symmetry in the response of volatility to news is preserved in the latter model. This is obviously because of certain 'symmetry' of the exchange rates: good news for the US dollar may be bad news for the SPD, and vice versa. As in the previous example, the time-varying news impact curves distinguishes different reaction levels of volatility to news in calm and turbulent times. When there is plenty of turbulence, a piece of news of a given size may drown in the general uncertainty, while the same news can have a large impact during a tranquil period.

8 Concluding remarks

In this paper we introduce two new nonstationary GARCH models whose parameters are allowed to have a smoothly time-varying structure. Time-variation of the (un)conditional variance is incorporated in the model either in an additive or a multiplicative form, of which we focus on the latter. This approach is appealing since most daily financial return series cover a long time period and non-constancy of parameters in models describing them therefore appears quite likely. We also develop a modelling strategy for our multiplicative TV-GARCH model. In order to determine the appropriate number of transitions we propose a procedure consisting of a sequence of Lagrange multiplier tests. The test statistics can be robustified against deviations from the iid assumption. Our simulation experiments suggest that the parameter constancy tests have reasonably good size properties already in samples of moderate size. The modelling strategy appears to work quite well for the data-generating processes that we simulate.

We put our TV-GARCH models to test by applying the modelling strategy to daily stock index and exchange rate returns. We find that parameter constancy against an additive and a multiplicative structure is strongly rejected for both return series. Fitting a traditional GARCH or GJR-GARCH model to these series yields results that are quite different from the ones obtained by our approach and point at the presence of long memory in volatility. Our results show that the long-memory type behaviour of the sample autocorrelation functions of the absolute returns may also be induced by changes in the unconditional variance. Once the model accounts for the time-variation in the unconditional variance, the evidence for long memory is considerably weakened or even vanishes altogether.

An extension to multivariate GARCH models appears possible. The so-called Constant Conditional Correlation (CCC-) GARCH model by Bollerslev (1990) and its extensions typically make use of a standard GARCH(1,1) specification for conditional variances. These GARCH equations may be generalized to account for time-variation in parameters. An interesting question to investigate with our TV-GARCH specifications is how such a generalization would affect estimates of time-varying correlations in a situation in which there are changes in the unconditional variance of the return series. This and other extensions to multivariate models will be left for future work.

References

Amemiya, T. (1985): *Advanced Econometrics.* Blackwell, Oxford.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

Baillie, R. T., and C. Morana (2009): "Modeling Long Memory and Structural Breaks in Conditional Variances: An Adaptive FIGARCH Approach," *Journal of Economic Dynamics and Control*, 33, 1577–1592.

Berkes, I., E. Gombay, L. Horváth, and P. Kokoszka (2004): "Sequential Change-Point Detection in GARCH(p,q) Models," *Econometric Theory*, 20, 1140–1167.

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

Bollerslev, T. (1990): "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *The Review of Economic and Statistics*, 72, 498–505.

Brownlees, C. T., and G. M. Gallo (2010): "Comparison of Volatility Measures: A Risk Management Perspective," *Journal of Financial Econometrics*, 8, 29–56.

Cai, J. (1994): "A Markov Model of Unconditional Variance in ARCH," *Journal of Business & Economic Statistics*, 12, 309–316.

Dahlhaus, R., and S. Subba Rao (2006): "Statistical Inference for Time-Varying ARCH Processes," *Annals of Statistics*, 34, 1075–1114.

Davidson, J. (2000): *Econometric Theory.* Blackwell, Oxford.

Davidson, J., A. Monticini, and D. Peel (2007): "Implementing the Wild Bootstrap Using a Two-Point Distribution," *Economics Letters*, 96, 309–315.

Davidson, R., and J. G. MacKinnon (1998): "Graphical Methods for Investigating the Size and Power of Hypothesis Tests," *The Manchester School*, 66, 1–26.

Diebold, F. X. (1986): "Comment on "Modelling the Persistence of Conditional Variance", by R. Engle and T. Bollerslev," *Econometric Reviews*, 5, 51–56.

Doornik, J. A. (2007): *Object-Oriented Matrix Programming Using Ox.* 3rd ed. London: Timberlake Consultants Press and Oxford: www.doornik.com.

Eitrheim, Ø., and T. Teräsvirta (1996): "Testing the Adequacy of Smooth Transition Autoregressive Models," *Journal of Econometrics*, 74, 59–75.

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

Engle, R. F., and J. Gonzalo Rangel (2008): "The Spline-GARCH Model for Low-Frequency Volatility and its Global Macroeconomic Causes," *Review of Financial Studies*, 21, 1187–1222.

Fan, Y., S. Pastorello, and E. Renault (2007): “Maximization by Parts in Extremum Estimation,” Unpublished paper.

Fiorentini, G., G. Calzolari, and L. Panattoni (1996): “Analytic Derivatives and the Computation of GARCH Estimates,” *Journal of Applied Econometrics*, 11, 399–417.

Gallant, R. (1984): “The Fourier Flexible Form,” *American Journal of Agricultural Economics*, 66, 204–208.

Hamilton, J. D., and R. Susmel (1994): “Autoregressive Conditional Heteroskedasticity and Changes in Regime,” *Journal of Econometrics*, 64, 307–333.

Kim, T.-H., and H. White (2004): “On More Robust Estimation of Skewness and Kurtosis,” *Finance Research Letters*, 1, 56–73.

Lamoureux, C. G., and W. D. Lastrapes (1990): “Persistence in Variance, Structural Change, and the GARCH Model,” *Journal of Business & Economic Statistics*, 8, 225–234.

Li, W. K., and T. K. Mak (1994): “On the Squared Residual Autocorrelations in Non-Linear Time Series With Conditional Heteroskedasticity,” *Journal of Time Series Analysis*, 15, 627–636.

Lundbergh, S., and T. Teräsvirta (2002): “Evaluating GARCH Models,” *Journal of Econometrics*, 110, 417–435.

McLeod, A. I., and W. K. Li (1983): “Diagnostic Checking ARMA Time Series Models Using Squared Residual Autocorrelations,” *Journal of Time Series Analysis*, 4, 269–273.

Meitz, M., and P. Saikkonen (in press): “Parameter Estimation in Nonlinear AR-GARCH Models,” *Econometric Theory.*

Mikosch, T., and C. Stărică (2004): “Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects,” *The Review of Economics and Statistics*, 86, 378–390.

Mishra, S., L. Su, and A. Ullah (2010): “Semiparametric Estimator of Time Series Conditional Variance,” *Journal of Business and Economic Statistics*, 28, 256–274.

Osiewalski, J. (2009): “New Hybrid Models of Multivariate Volatility (a Bayesian Perspective),” *Przeglad Statystycny*, 56, 15–22.

Osiewalski, J., and A. Pajor (2009): “Bayesian Analysis for Hybrid MSF-SBEKK Models of Multivariate Volatility,” *Central European Journal of Economic Modelling and Econometrics*, 1, 179–202.

Song, P. X.-K., Y. Fan, and J. D. Kalbfleisch (2005): “Maximization by Parts in Likelihood Inference,” *Journal of the American Statistical Association*, 100, 1145–1157.

Straumann, D., and T. Mikosch (2006): “Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach,” *Annals of Statistics*, 34, 2449–2495.

Teräsvirta, T. (1998): "Modeling Economic Relationships with Smooth Transition Regressions," in *Handbook of Applied Economic Statistics*, ed. by A. Ullah, and D. Giles, pp. 507–552. New York: Dekker.

——— (2009): "An Introduction to GARCH Models," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch, pp. 17–42, New York. Springer.

——— (2011): "Nonlinear Models for Autoregressive Conditional Heteroskedasticity," in *Wiley Handbook in Financial Engineering and Econometrics: Volatility Models and Their Applications*, ed. by L. Bauwens, C. Hafner, and S. Laurent, New York. Wiley.

Teräsvirta, T., D. Tjøstheim, and C. W. J. Granger (2010): *Modelling Nonlinear Economic Time Series.* Oxford University Press, Oxford.

Teräsvirta, T., and Z. Zhao (in press): "Stylized Facts of Return Series, Robust Estimates, and Three Popular Models of Volatility," *Applied Financial Economics.*

van Bellegem, S. (2011): "Locally Stationary Volatility Models," in *Wiley Handbook in Financial Engineering and Econometrics: Volatility Models and Their Applications*, ed. by L. Bauwens, C. Hafner, and S. Laurent, New York. Wiley.

van Bellegem, S., and R. von Sachs (2004): "Forecasting Economic Time Series with Unconditional Time-Varying Variance," *International Journal of Forecasting*, 20, 611– 627.

Appendix A

Theorem 1 will be proved by verifying the assumptions of Theorem 4.1.6 in Amemiya (1985, p. 114). The quasi log-likelihood function for the model is

$$L_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon}) = \sum_{t=1}^{T} \ell(\boldsymbol{\theta}_1,\varepsilon_t)$$

where

$$\ell(\boldsymbol{\theta}_1,\varepsilon_t) = k - (1/2)\{\ln g_t(\boldsymbol{\theta}_1) + \frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)}\}$$

and $g_t(\boldsymbol{\theta}_1)$ is defined in (12). For notational simplicity, we set $\boldsymbol{\theta}_1 = (\alpha_0, \boldsymbol{\delta}^{*\prime}, \boldsymbol{\gamma}', \mathbf{c}')'$.

We begin by considering the first and the second partial derivatives of the function (15) and show that they are continuous and bounded:

Lemma A.1. The function $\partial\ell(\boldsymbol{\theta}_1,\varepsilon_t)/\partial\boldsymbol{\theta}_1$ has the form

$$\begin{aligned}
\partial\ell(\boldsymbol{\theta}_1,\varepsilon_t)/\partial\boldsymbol{\theta}_1 &= -(1/2)\frac{\partial}{\partial\boldsymbol{\theta}_1}\{\ln g_t(\boldsymbol{\theta}_1) + \frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)}\} \\
&= (1/2)(\frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)} - 1)\frac{1}{g_t(\boldsymbol{\theta}_1)}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}
\end{aligned}$$

where

$$\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1} = (\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\alpha_0}\ \frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\delta}^{*\prime}}\ \frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\gamma}'}\ \frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\mathbf{c}'})'.$$

with

$$\begin{aligned}
\partial g_t/\partial\alpha_0 &= 1 \\
\partial g_t/\partial\boldsymbol{\delta}^* &= (G_{1t}, ..., G_{rt})' && (30) \\
\partial g_t/\partial\boldsymbol{\gamma} &= (g_{\gamma 1t}, ..., g_{\gamma rt})' && (31) \\
\partial g_t/\partial\mathbf{c} &= (g_{c1t}, ..., g_{crt})'. && (32)
\end{aligned}$$

In (31),

$$g_{\gamma jt} = \delta_j^* G_{jt}(1 - G_{jt})(t^* - c_j)$$

and in (32),

$$g_{cjt} = -\gamma_j\delta_j^* G_{jt}(1 - G_{jt})$$

for $j = 1, ..., r$.

The next lemma gives the second partial derivatives of (15).

Lemma A.2.

$$\partial^2 g_t(\boldsymbol{\theta}_1)/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1' = \begin{bmatrix} 0 & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times r} & \mathbf{0}_{1\times r} \\ & \mathbf{0}_{r\times r} & \mathrm{diag}(g_{\delta^*\gamma 1t}, ..., g_{\delta\gamma rt}) & \mathrm{diag}(g_{\delta^* c1t}, ..., g_{\delta crt}) \\ & & \mathrm{diag}(g_{\gamma\gamma 1t}, ..., g_{\gamma\gamma rt}) & \mathrm{diag}(g_{\gamma c1t}, ..., g_{\gamma crt}) \\ & & & \mathrm{diag}(g_{cc1t}, ..., g_{ccrt}) \end{bmatrix}$$

where

$$
\begin{aligned}
g_{\gamma\gamma jt} &= \delta_j^* G_{jt}(1-G_{jt})(1-2G_{jt})(t^*-c_j)^2 \\
g_{ccjt} &= \delta_j^* \gamma_j^2 G_{jt}(1-G_{jt})(1-2G_{jt}) \\
g_{\delta\gamma jt} &= G_{jt}(1-G_{jt})(t^*-c_j) \\
g_{\delta cjt} &= -\gamma_j G_{jt}(1-G_{jt}) \\
g_{\gamma cjt} &= -\delta_j^* \gamma_j G_{jt}(1-G_{jt})(1-2G_{jt})(t^*-c_j)
\end{aligned}
$$

for $j = 1, ..., r$.

Lemma A.3. $\mathsf{E}\sup_{\boldsymbol{\theta}_1\in\Theta} |\ell(\boldsymbol{\theta}_1,\varepsilon_t)| < \infty$.

Proof. From AG1 it follows that Θ is a compact set such that $\boldsymbol{\theta}_1^0$ is an interior point of this set. Then, ignoring the constant k which is irrelevant in this context,

$$
\begin{aligned}
\mathsf{E}\sup_{\boldsymbol{\theta}_1\in\Theta} |\ell(\boldsymbol{\theta}_1,\varepsilon_t)| &= \mathsf{E}\sup_{\boldsymbol{\theta}_1\in\Theta} |\ln g_t(\boldsymbol{\theta}_1) + \frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)}| \leq \mathsf{E}\sup_{\boldsymbol{\theta}_1\in\Theta} \{|\ln g_t(\boldsymbol{\theta}_1)| + \frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)}\} \\
&\leq \mathsf{E}\{\sup_{\boldsymbol{\theta}_1\in\Theta} |\ln g_t(\boldsymbol{\theta}_1)| + \sup_{\boldsymbol{\theta}_1\in\Theta} \frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)}\} \leq \sup_{\boldsymbol{\theta}_1\in\Theta} |\ln g_t(\boldsymbol{\theta}_1)| + \sup_{\boldsymbol{\theta}_1\in\Theta} \frac{g_t(\boldsymbol{\theta}_1^0)}{g_t(\boldsymbol{\theta}_1)} < \infty.
\end{aligned}
$$

because $g_t(\boldsymbol{\theta}_1)$ is bounded away from zero and $\sup_{\boldsymbol{\theta}_1\in\Theta} g_t(\boldsymbol{\theta}_1) < \infty$. ■

Lemma A.4 [Thm 4.1.3, Assumption (A)]. $L_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon})$ continuous in Θ_1 for each $\boldsymbol{\varepsilon}$.

Proof. From Lemma A.1 and the fact that $g_t(\boldsymbol{\theta}_1)$ is continuous for all $\boldsymbol{\theta}_1$ and ε_t, it follows that $\ell(\boldsymbol{\theta}_1,\varepsilon_t)$ is continuous in Θ_1 for each ε_t, and the same is then true for $L_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon})$. ■

Lemma A.5 [Thm 4.1.3, Assumption (B)]. The average Hessian

$$
T^{-1}\mathbf{H}_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon}) = T^{-1}\frac{\partial^2 L_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon})}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'} = T^{-1}\sum_{t=1}^{T}\frac{\partial^2 \ell(\boldsymbol{\theta}_1,\varepsilon_t)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}
$$

converges to a finite nonsingular matrix $\mathbf{A}(\boldsymbol{\theta}_1^0)$ for any sequence $\boldsymbol{\theta}_{1T}^*$ such that $\mathrm{plim}_{T\to\infty}\boldsymbol{\theta}_{1T}^* = \boldsymbol{\theta}_1^0$.

Proof. A straightforward calculation yields

$$
\begin{aligned}
T^{-1}\mathbf{H}_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon}) &= -(1/2T^{-1})\sum_{t=1}^{T}(\frac{2\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)} - 1)\frac{1}{g_t^2(\boldsymbol{\theta}_1)}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1'} \\
&\quad +(1/2T^{-1})\sum_{t=1}^{T}(\frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)} - 1)\frac{1}{g_t(\boldsymbol{\theta}_1)}\frac{\partial^2 g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}. \qquad (33)
\end{aligned}
$$

Since the elements of (33) are continuous and bounded, it can be shown that

$$
\mathsf{E}\sup_{\boldsymbol{\theta}_1\in\Theta} |(\frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)} - 1)\frac{1}{g_t(\boldsymbol{\theta}_1)}[\frac{\partial^2 g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}]_{ij}| < \infty \qquad (34)
$$

for each (i,j) element of $\partial^2 g_t(\boldsymbol{\theta}_1)/(\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1')$. By Kolmogorov's LLN 1,

$$\begin{aligned}
& \text{plim}_{T\to\infty} T^{-1}\frac{\partial^2 L_T(\boldsymbol{\theta}_1,\boldsymbol{\varepsilon})}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'} \\
= \; & -(1/2)\text{lim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}\Big(\frac{2g_t(\boldsymbol{\theta}_1^0)}{g_t(\boldsymbol{\theta}_1)}-1\Big)\frac{1}{g_t^2(\boldsymbol{\theta}_1)}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1'} \\
& +(1/2)\text{lim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}\Big(\frac{g_t(\boldsymbol{\theta}_1^0)}{g_t(\boldsymbol{\theta}_1)}-1\Big)\frac{1}{g_t(\boldsymbol{\theta}_1)}\frac{\partial^2 g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}
\end{aligned} \tag{35}$$

because $\ell(\boldsymbol{\theta}_1,\varepsilon_s)$ and $\ell(\boldsymbol{\theta}_1,\varepsilon_t)$ are independent for $s \neq t$ and because $g_t(\boldsymbol{\theta}_1)$ is twice continuously differentiable for all $\boldsymbol{\theta}_1$ and every ε_t. Since the elements of (33) are continuous in $\boldsymbol{\theta}_1$ and (34) holds, convergence is uniform. Then, applying Theorem 4.1.5 in Amemiya (1985, p. 113) yields

$$\text{plim}_{T\to\infty} T^{-1}\frac{\partial^2 L_T(\boldsymbol{\theta}_1,\boldsymbol{\varepsilon})}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\Big|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_{1T}^*} = \text{plim}_{T\to\infty} T^{-1}\frac{\partial^2 L_T(\boldsymbol{\theta}_1,\boldsymbol{\varepsilon})}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\Big|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0}$$

for any sequence $\{\boldsymbol{\theta}_{1T}^*\}$ such that $\text{plim}_{T\to\infty}\boldsymbol{\theta}_{1T}^* = \boldsymbol{\theta}_1^0$. From (35) it follows that

$$\begin{aligned}
& \text{plim}_{T\to\infty} T^{-1}\mathbf{H}_T(\boldsymbol{\theta}_{1T}^*,\boldsymbol{\varepsilon}) \\
= \; & -(1/2)\text{lim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}\frac{1}{g_t^2(\boldsymbol{\theta}_1^0)}\Big\{\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1'}\Big\}\Big|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0} = \mathbf{A}(\boldsymbol{\theta}_1^0)
\end{aligned}$$

for $\boldsymbol{\theta}_{1T}^* \to \boldsymbol{\theta}_1^0$. ■

Let $\mathbf{s}(\boldsymbol{\theta}_1,\boldsymbol{\varepsilon}) = T^{-1}\partial L_T(\boldsymbol{\theta}_1,\boldsymbol{\varepsilon})/\partial\boldsymbol{\theta}_1$ be the average score of (14). We have

Lemma A.6 [Thm 4.1.3, Assumption C].

$$\begin{aligned}
T^{1/2}\mathbf{s}(\boldsymbol{\theta}_1^0,\boldsymbol{\varepsilon}) \; &= \; (1/2T^{1/2})\sum_{t=1}^{T}\{\partial\ell(\boldsymbol{\theta}_1,\varepsilon_t)/\partial\boldsymbol{\theta}_1\}|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0} \\
& \xrightarrow{D} \mathcal{N}(\mathbf{0},\mathbf{B}(\boldsymbol{\theta}_1^0)).
\end{aligned}$$

Proof. From Lemma A.1 one obtains

$$T^{1/2}\mathbf{s}(\boldsymbol{\theta}_1,\boldsymbol{\varepsilon}) = (1/2T^{1/2})\sum_{t=1}^{T}\Big(\frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)}-1\Big)\frac{1}{g_t(\boldsymbol{\theta}_1)}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}.$$

We have

$$\mathsf{E}\{\partial\ell(\boldsymbol{\theta}_1,\varepsilon_t)/\partial\boldsymbol{\theta}_1|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0}\} = (1/2)\mathsf{E}\Big(\frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1^0)}-1\Big)\frac{1}{g_t(\boldsymbol{\theta}_1^0)}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\Big|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0} = \mathbf{0}$$

and

$$\begin{aligned}
\text{cov}\{\partial\ell(\boldsymbol{\theta}_1,\varepsilon_t)/\partial\boldsymbol{\theta}_1|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0}\} \; &= \; (1/4)\mathsf{E}\Big(\frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1^0)}-1\Big)^2\frac{1}{g_t^2(\boldsymbol{\theta}_1^0)}\Big\{\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1'}\Big\}\Big|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0} \\
&= \; \frac{\mu_4-1}{4}\frac{1}{g_t^2(\boldsymbol{\theta}_1^0)}\Big\{\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1'}\Big\}\Big|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0}
\end{aligned}$$

because $\mathsf{E}\zeta_t^4 = \mu_4$. Let $\boldsymbol{\theta}_1 = (\theta_1, ..., \theta_{3r+1})'$ and $x_{jt} = \partial\ell(\boldsymbol{\theta}_1,\varepsilon_t)/\partial\theta_j|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0}$. Then $\mathsf{E}x_{jt} = 0$ and

$$\mathrm{var}(x_{jt}) = \mathsf{E}x_{jt}^2 = \frac{\mu_4 - 1}{4}\frac{1}{g_t^2(\boldsymbol{\theta}_1^0)}\{\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\theta_j}|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0}\}^2.$$

Applying Kolmogorov's LLN 1 to the sum $T^{-1}\sum_{t=1}^T x_{jt}$ yields

$$\mathrm{plim}_{T\to\infty} T^{-1}\sum_{t=1}^T x_{jt} = 0.$$

Furthermore, since $\mathsf{E}|\zeta_t|^{2(2+\phi)} < \infty$ for some $\phi > 0$, all elements of $\{|x_{jt}|^{2+\phi}\}$ are $O_p(1)$, and so

$$\max_{t=1,...,T} |x_{jt}|^{2+\phi} = O_p(1).$$

This implies that the p-norm

$$||x_{jt}||_{2+\phi} = (\mathsf{E}|x_{jt}|^{2+\phi})^{1/(2+\phi)} = O(1).$$

Also, due to the continuity and differentiability of $g_t(\boldsymbol{\theta}_1)$,

$$\{g_t^{-2}(\boldsymbol{\theta}_1^0)(\partial g_t(\boldsymbol{\theta}_1)/\partial\theta_j|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1^0})\}^2 < \infty$$

for all t, and nearly all the variables in the sequence are positive as $T \to \infty$. It then follows that

$$\{T^{-1}\sum_{t=1}^T \mathrm{var}(x_{jt})\}^{1/2} = O(1)$$

and, consequently, for some $\phi > 0$,

$$\frac{\max_{j=1,...,T} ||x_{jt}||_{2+\phi}}{\{T^{-1}\sum_{t=1}^T \mathrm{var}(x_{jt})\}^{1/2}} \leq M < \infty,\ T \geq 1.$$

Thus x_{jt} satisfies the assumptions of Theorem 3.3.2 in Davidson (2000, p. 44), which proves that $T^{-1/2}\sum_{t=1}^T x_{jt} \xrightarrow{D} x_j \sim \mathcal{N}(0, \lim_{T\to\infty} T^{-1}\sum_{t=1}^T \mathrm{var}(x_{jt}))$, $j = 1, ..., 3r+1$. It follows for all linear combinations $\boldsymbol{\lambda}'\mathbf{x}_t$ with $\boldsymbol{\lambda} \neq \mathbf{0}$ that $\boldsymbol{\lambda}'\mathbf{x}_t \xrightarrow{D} \boldsymbol{\lambda}'\mathbf{x}$, where $\mathbf{x} = (x_1, ..., x_{3r+1})'$. From Theorems 3.3.3 and 3.3.4 in Davidson (2000, p. 46) one concludes that $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{B}(\boldsymbol{\theta}_1^0))$, where

$$\mathbf{B}(\boldsymbol{\theta}_1^0) = \frac{\mu_4 - 1}{4}\lim_{T\to\infty} T^{-1}\sum_{t=1}^T \frac{1}{g_t^2(\boldsymbol{\theta}_1^0)}\{\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1'}\}|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0}.\blacksquare \tag{36}$$

Lemma A.7 [Thm 4.1.6, Assumption (A)]. Function $T^{-1}L_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon})$ converges to a nonstochastic function $L(\boldsymbol{\theta}_1)$ in probability uniformly in $\boldsymbol{\theta}_1$ (in a neighbourhood of $\boldsymbol{\theta}_1^0$).

Proof. We have

$$\begin{aligned} T^{-1}L_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon}) &= T^{-1}\sum_{t=1}^T \ell(\boldsymbol{\theta}_1,\varepsilon_t) = T^{-1}\sum_{t=1}^T [k - (1/2)\{\ln g_t(\boldsymbol{\theta}_1) + \frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1)}\}] \\ &= T^{-1}\sum_{t=1}^T [k - (1/2)\{\ln g_t(\boldsymbol{\theta}_1) + (\frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1^0)} - 1)\frac{g_t(\boldsymbol{\theta}_1^0)}{g_t(\boldsymbol{\theta}_1)} + \frac{g_t(\boldsymbol{\theta}_1^0)}{g_t(\boldsymbol{\theta}_1)}\}] \\ &\to L(\boldsymbol{\theta}_1) = k - (1/2)\lim_{T\to\infty} T^{-1}\sum_{t=1}^T \{\ln g_t(\boldsymbol{\theta}_1) + \frac{g_t(\boldsymbol{\theta}_1^0)}{g_t(\boldsymbol{\theta}_1)}\} \end{aligned}$$

as $T \to \infty$. This follows from the Kolmogorov LLN 1 applied to the sequence of independent variables

$$u_t = (\frac{\varepsilon_t^2}{g_t(\boldsymbol{\theta}_1^0)} - 1)\frac{g_t(\boldsymbol{\theta}_1^0)}{g_t(\boldsymbol{\theta}_1)}.$$

That the convergence is uniform is a consequence of Lemmata A.3 and A.4. ■

Lemma A.8. [Thm 4.1.6, Assumption (C)] The probability limit

$$\mathrm{plim}_{T\to\infty} T^{-1}\mathbf{H}_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon}) = \mathrm{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} \frac{\partial^2}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\ell_T(\boldsymbol{\theta}_1, \boldsymbol{\varepsilon})$$

exists and is continuous in a neighbourhood of $\boldsymbol{\theta}_1^0$.

Proof. The probability limit of the average Hessian is given in (35). It is continuous for all $\boldsymbol{\theta}_1$. The continuity follows from the fact that $g_t(\boldsymbol{\theta}_1)$ is bounded and infinitely many times differentiable in Θ_1. ■

When $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$, (35) becomes

$$\begin{aligned} & \mathrm{plim}_{T\to\infty} T^{-1}\mathbf{H}_T(\boldsymbol{\theta}_1^0, \boldsymbol{\varepsilon}) \\ = & -(1/2)\mathrm{lim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} \frac{1}{g_t^2(\boldsymbol{\theta}_1^0)}\{\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\frac{\partial g_t(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1'}\}|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^0} = \mathbf{A}(\boldsymbol{\theta}_1^0) \end{aligned}$$

which is a negative definite matrix.

Proof of Theorem 1. The result follows from the fact that $\mathbf{A}(\boldsymbol{\theta}_1^0)$ is negative definite and from Lemmata A.3–A.8. ■

Appendix B: Tables and Figures

Table 1. Model selection frequencies for the sequential testing procedure ($r = 0$)

DGPs	Number of transitions	$T = 1000$ LM$_1$	$T = 1000$ LM$_3$	$T = 2500$ LM$_1$	$T = 2500$ LM$_3$	$T = 5000$ LM$_1$	$T = 5000$ LM$_3$
DGP (i)	$r = 0$	**95.56**	**93.46**	**95.16**	**94.66**	**95.88**	**95.12**
	$r = 1$	3.54	5.50	4.78	4.12	4.02	4.24
	$r \geq 2$	0.90	1.04	0.06	1.22	0.10	0.64
DGP (ii)	$r = 0$	**94.08**	**91.82**	**94.74**	**93.88**	**95.52**	**94.36**
	$r = 1$	4.24	5.02	3.50	3.74	2.96	3.34
	$r \geq 2$	1.68	3.16	1.76	2.38	1.52	2.30
DGP (iii)	$r = 0$	**93.56**	**90.48**	**93.82**	**93.20**	**94.92**	**93.98**
	$r = 1$	5.12	7.24	6.02	5.06	5.04	5.08
	$r \geq 2$	1.32	2.28	0.16	1.74	0.04	0.94
DGP (iv)	$r = 0$	**94.44**	**93.30**	**94.82**	**94.44**	**95.12**	**95.54**
	$r = 1$	3.14	4.46	2.86	3.70	2.72	3.34
	$r \geq 2$	2.42	2.24	2.32	1.86	2.16	2.12

Notes: Selection frequencies in percentage of the standard LM parameter constancy test based on the DGPs described in Section 6.1. The number of replications is 5000 for each simulation. The initial nominal significance level equals 5%. The columns 'LM$_1$' and 'LM$_3$' correspond to the test procedure based on the first-order and third-order Taylor expansions, respectively.

Table 2. Model selection frequencies for the sequential testing procedure ($r = 1$)

δ_1	γ_1	Number of transitions	$T = 1000$ LM$_1$	$T = 1000$ LM$_3$	$T = 2500$ LM$_1$	$T = 2500$ LM$_3$	$T = 5000$ LM$_1$	$T = 5000$ LM$_3$
-0.05	5	$r = 0$	71.95	80.10	39.55	55.75	10.15	22.75
		$r = 1$	**27.65**	**18.25**	**59.20**	**43.00**	**84.25**	**73.55**
		$r \geq 2$	0.40	1.65	1.25	1.25	5.60	3.70
	10	$r = 0$	53.00	65.90	14.30	24.70	0.60	2.20
		$r = 1$	**46.15**	**32.00**	**83.40**	**73.05**	**92.90**	**94.05**
		$r \geq 2$	0.85	2.10	2.30	2.25	6.50	3.75
0.05	5	$r = 0$	57.95	69.85	24.20	42.70	3.65	11.40
		$r = 1$	**41.35**	**28.05**	**73.60**	**54.15**	**86.30**	**81.27**
		$r \geq 2$	0.70	2.10	2.20	3.15	10.05	7.33
	10	$r = 0$	39.00	52.35	5.40	13.30	0.10	0.53
		$r = 1$	**59.85**	**44.35**	**90.65**	**82.25**	**89.65**	**93.37**
		$r \geq 2$	1.15	3.30	3.95	4.45	10.25	6.10

Notes: Selection frequencies in percentage of the standard LM parameter constancy test based on the DGPs described in Section 6.1. The number of replications is 2000 for each simulation. The initial nominal significance level equals 5%. The columns 'LM$_1$' and 'LM$_3$' correspond to the test procedure based on the first-order and third-order Taylor expansions, respectively.

Table 3. Descriptive statistics and diagnostics for the daily returns

	S&P 500 returns			SPD/USD returns		
	S&P 500	$\varepsilon_t/\hat{g}_t^{1/2}$	$\varepsilon_t/(\hat{h}_t\hat{g}_t)^{1/2}$	SPD/USD	$\varepsilon_t/\hat{g}_t^{1/2}$	$\varepsilon_t/(\hat{h}_t\hat{g}_t)^{1/2}$
Minimum	−7.1127	−4.4083	−6.1887	−4.1444	−1.9206	−6.2395
Maximum	4.9887	3.0918	3.5742	2.7618	1.4231	3.9959
Skewness	−0.3678	−0.3427	−0.3922	−0.9045	−0.2908	−0.2450
Robust SK	0.0325	0.0216	0.0280	−0.0044	−0.0265	−0.0185
Ex.kurtosis	5.2867	2.7575	1.8776	14.593	3.2007	2.2351
Robust KR	0.2541	0.1713	0.1342	0.1662	0.0969	0.0983
Std. dev.	0.8912	0.6093	0.9533	0.4150	0.2913	0.9975
Mean	0.0538	0.0407	0.0594	0.0077	0.0037	0.0144
LJB	3004 (0.0000)	851 (0.0000)	437 (0.0000)	18600 (0.0000)	908 (0.0000)	449 (0.0000)
ARCH(4)	154 (3×10^{-32})	–	–	340 (3×10^{-72})	–	–
T	2531	2531	2531	2060	2060	2060

Notes: LJB denotes the Lomnicki-Jarque-Bera normality test. ARCH(4) is the fourth-order ARCH LM test statistic described in Engle (1982). Robust SK denotes the robust measure for skewness based on quantiles proposed by Bowley (see Kim and White (2004)) and the robust KR denotes the robust centred coefficient for kurtosis proposed by Moors (see Kim and White (2004)). The numbers in parentheses are p-values.

Table 4. Estimation results for the GJR-GARCH and GARCH models in the two applications

	GJR-GARCH model				GARCH model			
	S&P 500		$\varepsilon_t/\hat{g}_{t_{S\&P500}}^{1/2}$		SPD/USD		$\varepsilon_t/\hat{g}_{t_{SPD/USD}}^{1/2}$	
	Estimate	(Std. error)	Estimate	(Std. error)	Estimate	(Std. error)	Estimate	(Std. error)
$\hat{\alpha}_0$	0.009	(0.006)	0.033	(0.011)	0.001	(7×10^{-4})	0.003	(0.001)
$\hat{\alpha}_1$	0.014	(0.008)	–	(–)	0.057	(0.008)	0.061	(0.021)
$\hat{\beta}_1$	0.939	(0.029)	0.848	(0.042)	0.937	(0.020)	0.901	(0.034)
$\hat{\lambda}_1$	0.079	(0.042)	0.137	(0.043)	–	(–)	–	(–)
$\hat{\alpha}_1 + \hat{\beta}_1 + \frac{\hat{\lambda}_1}{2}$	0.992		0.917		0.994		0.962	
Log-Lik	−3034.98		−2261.39		−635.66		−308.26	

Table 5. Results of the sequence of tests of constant unconditional variance against a time-varying GARCH model with multiplicative structure

Transitions in the alternative model	Parameter constancy test		Decision rule for selecting K					
			H_{03}		H_{02}		H_{01}	
	LM	p-value	LM_3	p-value	LM_2	p-value	LM_1	p-value
Single transition								
S&P 500	17.18	7×10^{-4}	2.30	0.1293	11.77	6×10^{-4}	3.13	0.0769
SPD/USD	14.93	0.0019	4.01	0.0453	7.82	0.0052	3.14	0.0766
Double transition								
S&P 500	9.54	0.0229	1.34	0.2471	0.42	0.5192	7.79	0.0053
SPD/USD	4.43	0.2188	2.17	0.1405	0.33	0.5655	1.93	0.1650
Triple transition								
S&P 500	2.38	0.4970	0.05	0.8270	0.43	0.5107	1.90	0.1678
SPD/USD	–	–	–	–	–	–	–	–

Table 6. Misspecification tests for the GJR-GARCH model in the stock returns application

	No ARCH-in-GARCH			GARCH(1,1) vs. GARCH(1,2)	GARCH(1,1) vs GARCH(2,1)	No ST-GARCH
	$r=1$	$r=5$	$r=10$			
S&P 500						
LM test	0.519	2.096	3.064	0.039	8.537	13.48
p-value	0.471	0.836	0.980	0.844	0.003	0.001
$\varepsilon_t/\hat{g}_{t_{S\&P500}}^{1/2}$						
LM test	1.955	4.432	7.202	0.001	4.617	4.933
p-value	0.162	0.489	0.706	0.977	0.032	0.085

Notes: The tests are those against remaining ARCH in the standardised residuals, GARCH(1,2) and GARCH(2,1) models, additional transition in the function g_t, and ST-GARCH(1,1) model.

Table 7. Misspecification tests for the GARCH model in the exchange rate returns application

	No ARCH-in-GARCH			GARCH(1,1) vs. GARCH(1,2)	GARCH(1,1) vs GARCH(2,1)	No ST-GARCH
	$r = 1$	$r = 5$	$r = 10$			
SPD/USD						
LM test	0.349	12.89	15.97	0.498	2.837	2.729
p-value	0.555	0.024	0.101	0.480	0.092	0.256
$\varepsilon_t/\hat{g}_{t_{SPD/USD}}^{1/2}$						
LM test	0.033	7.452	10.96	0.043	0.006	2.333
p-value	0.857	0.189	0.361	0.835	0.938	0.312

Notes: The tests are those against remaining ARCH in the standardised residuals, GARCH(1,2) and GARCH(2,1) models, additional transition in the function g_t, and ST-GARCH(1,1) model.

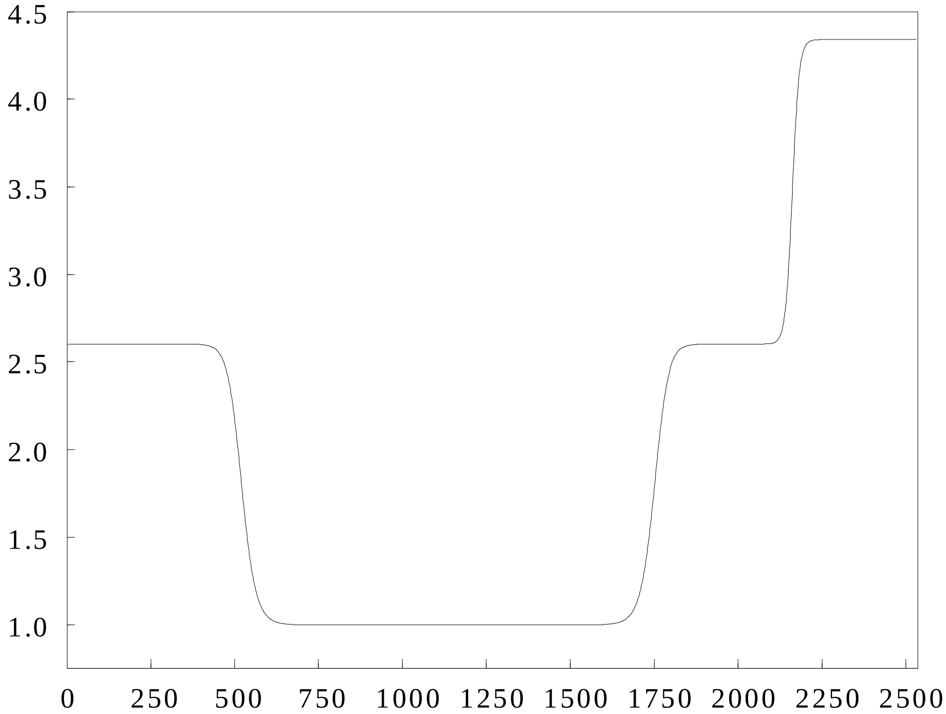
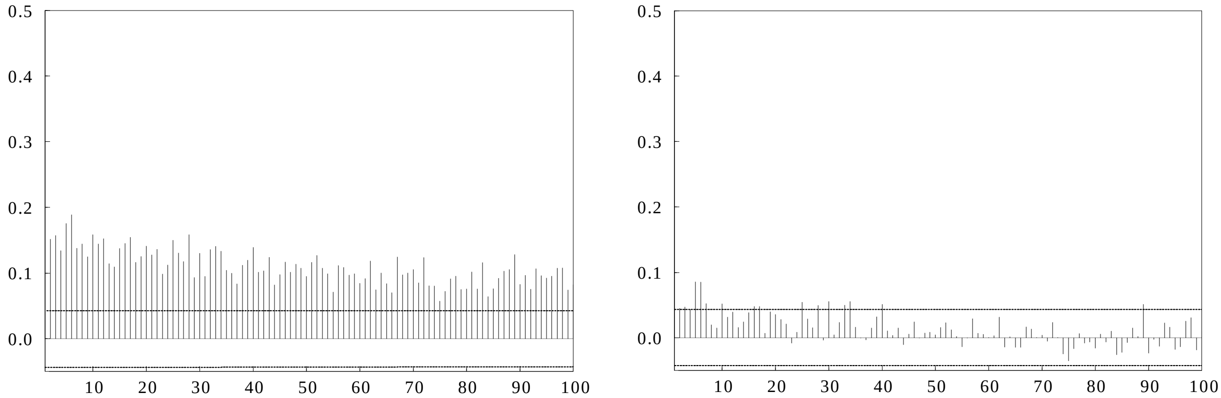
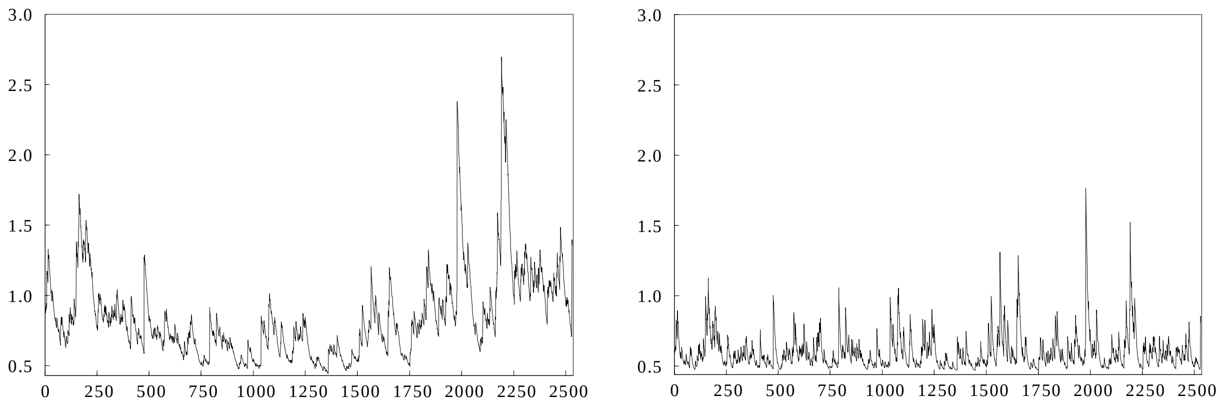


Figure 6. Graph of the final estimated function g_t for the S&P 500 returns model as a smooth function of the rescaled time variable t^* as given in (25)-(27).



(a) S&P 500 returns (b) 16th iteration

Figure 7. Sample autocorrelations of absolute log returns of the S&P 500 returns and the standardized variable $|\varepsilon_t|/\hat{g}_{t_{S\&P500}}^{1/2}$ for the first and the final iterations with the 95% confidence bounds.



(a) S&P 500 returns (b) 16th iteration

Figure 8. Conditional standard deviation of the GJR-GARCH(1,1) model for the S&P 500 returns and the standardized variable $\varepsilon_t/\hat{g}_{t_{S\&P500}}^{1/2}$ for the first and the final iterations.

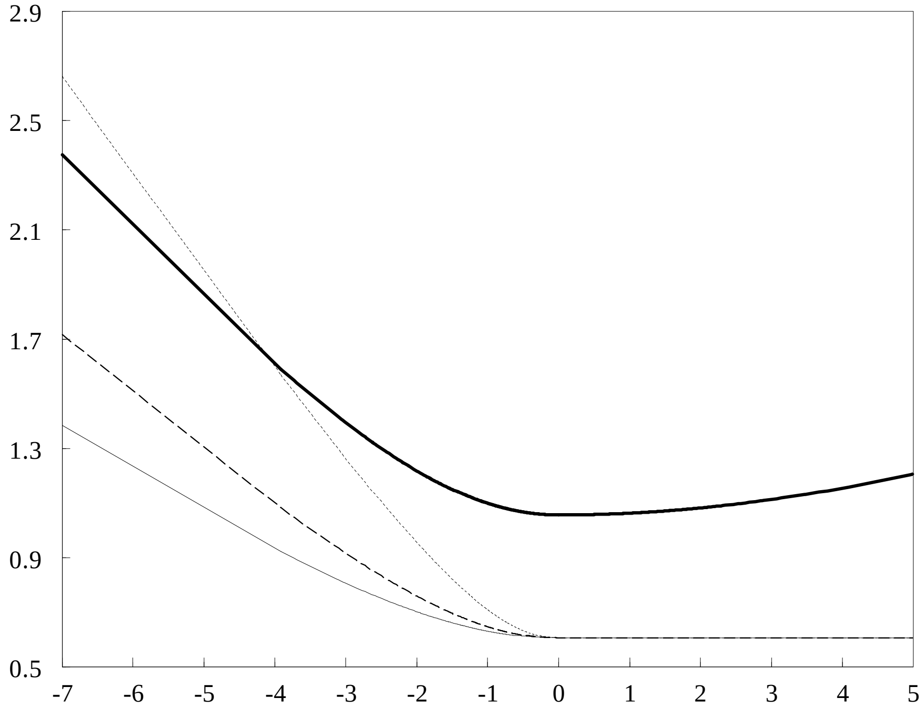


Figure 9. News impact curves of the GJR-GARCH(1,1) (solid line in boldface) and the TV-GJR-GARCH(1,1) models for several regimes in the stock returns application. The time-varying news impact curves are plotted for the lower regime, i.e. $G_1(t^*) = G_2(t^*) = 0$ (dotted line), for an intermediate regime, i.e. $G_1(t^*) = 1$ and $G_2(t^*) = 0$ (dashed line) and for the higher regime, i.e. $G_1(t^*) = G_2(t^*) = 1$ (solid line).

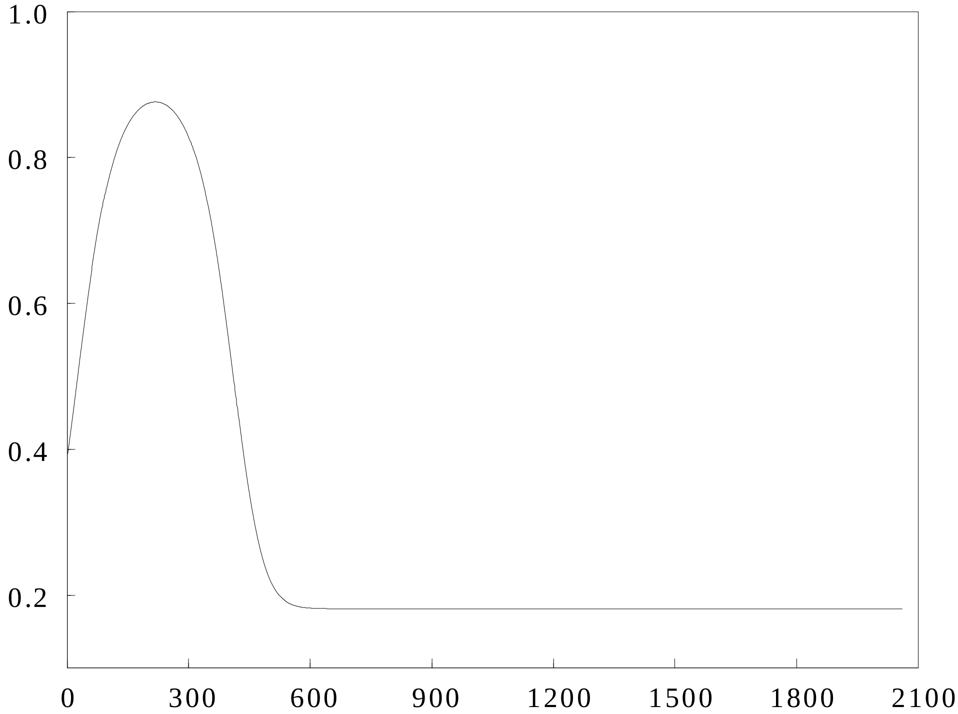
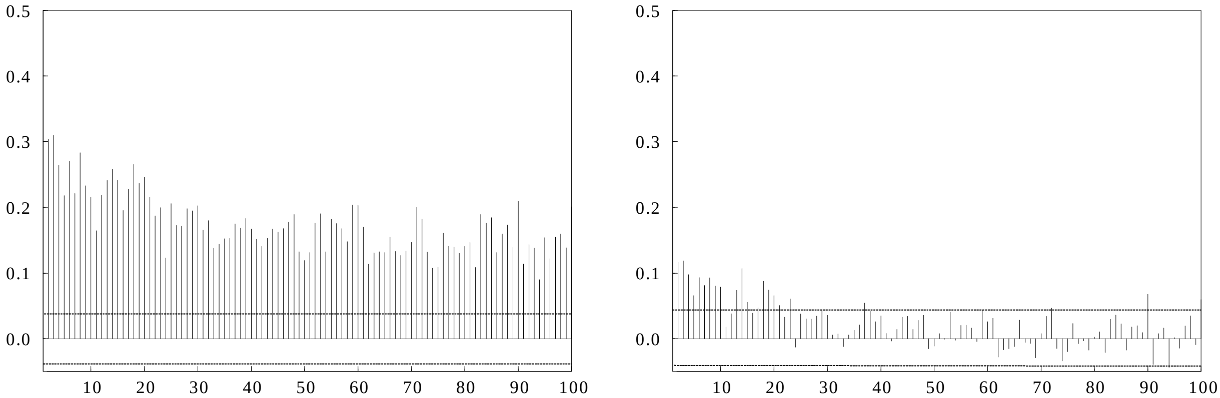
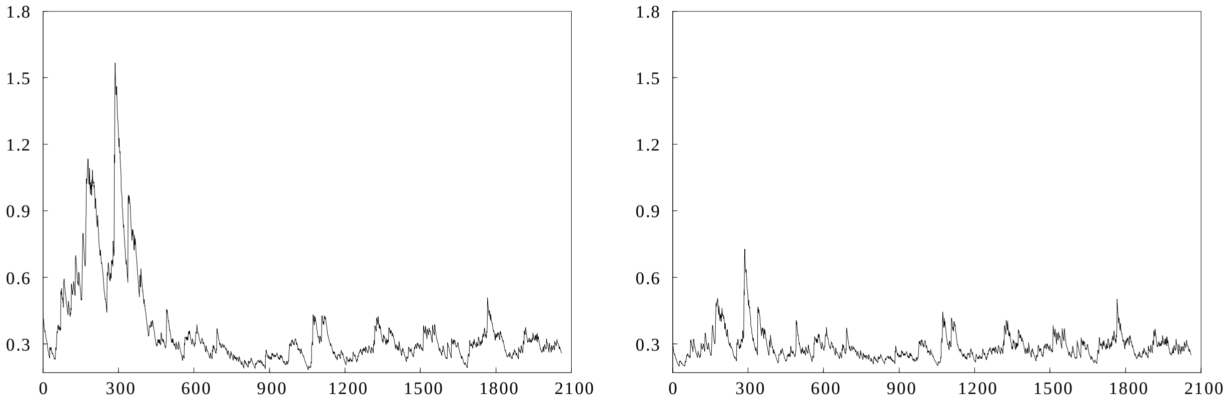


Figure 10. Graph of the final estimated function g_t for the SPD/USD returns model as a smooth function of the rescaled time variable t^* as given in (28)-(29).



(a) SPD/USD returns　　　(b) 82th iteration

Figure 11. Sample autocorrelations of absolute log returns of the SPD/USD returns and for the standardized variable $|\varepsilon_t|/\hat{g}^{1/2}_{t_{SPD/USD}}$ for the first and the final iterations with the 95% confidence bounds.



(a) SPD/USD returns　　　(b) 82th iteration

Figure 12. Conditional standard deviation of the GARCH(1,1) model for the SPD/USD returns and for the standardized variable $\varepsilon_t/\hat{g}^{1/2}_{t_{SPD/USD}}$ for the first and the final iterations.

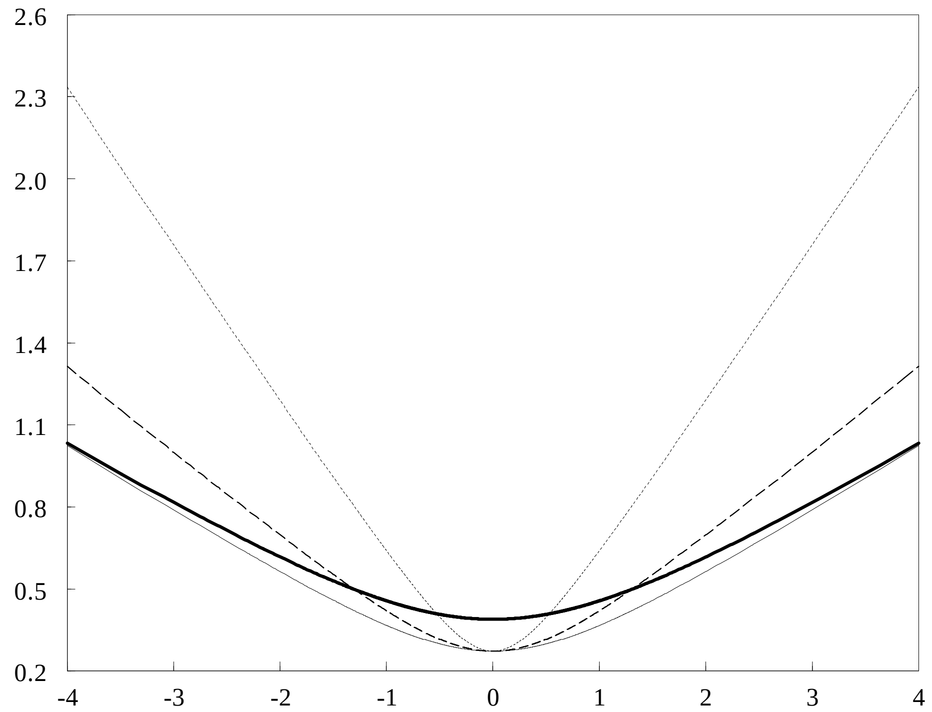


Figure 13. News impact curves of the GARCH(1,1) (solid line in boldface) and the TV-GARCH(1,1) models for several regimes in the exchange rate returns application. The time-varying news impact curves are plotted for the lower regime, i.e. $G_1(t^*) = 0$ (dotted line), for an intermediate regime, i.e. $G_1(t^*) = 0.5$ (dashed line) and for the higher regime, i.e. $G_1(t^*) = 1$ (solid line).

Research Papers 2011

2010-64: Tim Bollerslev and Viktor Todorov: Jump Tails, Extreme Dependencies, and the Distribution of Stock Returns

2010-65: Almut E. D. Veraart: How precise is the finite sample approximation of the asymptotic distribution of realised variation measures in the presence of jumps?

2010-66: Ole E. Barndorff-Nielsen, David G. Pollard and Neil Shephard: Integer-valued Lévy processes and low latency financial econometrics

2010-67: Shin Kanaya and Dennis Kristensen: Estimation of Stochastic Volatility Models by Nonparametric Filtering

2010-68: Dennis Kristensen and Anders Rahbek: Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models

2010-69: Søren Johansen: The analysis of nonstationary time series using regression, correlation and cointegration –with an application to annual mean temperature and sea level

2010-70: Søren Johansen and Morten Ørregaard Nielsen: A necessary moment condition for the fractional functional central limit theorem

2010-71: Nektarios Aslanidis and Isabel Casas : Modelling asset correlations during the recent financial crisis: A semiparametric approach

2010-72: Søren Johansen and Katarina Juselius: An invariance property of the common trends under linear transformations of the data

2010-73: Peter Sandholt Jensen and Allan H. Würtz: Estimating the effect of a variable in a high-dimensional regression model

2010-74: Peter R. Hansen, Asger Lunde and Valeri Voev: Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility and CoVolatility

2010-75: Laurent A.F. Callot: A Bootstrap Cointegration Rank Test for Panels of VAR Models

2010-76: Peter R. Hansen, Asger Lunde and James M. Nason: The Model Confidence Set

2011-01: Cristina Amado and Timo Teräsvirta: Modelling Volatility by Variance Decomposition