Modelling Conditional and Unconditional Heteroskedasticity with Smoothly Time-Varying Structure

Cristina Amado* and Timo Teräsvirta†

†*CREATES, School of Economics and Management, University of Aarhus*
Building 1322, DK-8000 Aarhus, Denmark

†**Department of Economic Statistics, Stockholm School of Economics*
P.O. Box 6501, SE-113 83 Stockholm, Sweden

**University of Minho and NIPE*
Campus de Gualtar, 4710-057 Braga, Portugal

Stockholm School of Economics
SSE/EFI Working Paper Series in Economics and Finance
No. 691

January 2008

*E-mail: cristina.amado@hhs.se

†E-mail: tterasvirta@econ.au.dk

Acknowledgements: 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; and at the seminars at Banca d'Italia, Rome, European University Institute, Florence, Humboldt University, Berlin, University of Minho, Braga, Stockholm School of Economics, Leonard N. Stern School of Business at New York University, and University of Vilnius. We would like to thank the participants at these occasions for their comments, and Stefan Lundbergh, Mika Meitz, Anders Rahbek and Esther Ruiz for useful discussions and suggestions. 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 conditional variance to have a smooth time-varying structure of either additive or multiplicative type. The suggested parameterizations describe both nonlinearity and structural change in the conditional and unconditional variances where the transition between regimes over time is smooth. A modelling strategy for these new time-varying parameter GARCH models is developed. It relies on a sequence of Lagrange multiplier tests, and the adequacy of the estimated models is investigated by Lagrange multiplier type misspecification tests. Finite-sample properties of these procedures 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

The modelling of 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). The increasing popularity of the class 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 estimation towards 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 (2007) 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. Engle and Gonzalo Rangel (2005) assumed that the variance of the process of interest can be decomposed into two components, a stationary and a nonstationary one. The nonstationary component is described by using splines, and the stationary component follows a GARCH process. The parameters of the latter are estimated conditionally on the spline component.

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 Engle and Gonzalo Rangel (2005). 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 standard GARCH model is first tested against these time-varying alternatives. If the null hypothesis is rejected, the structure of the time-varying component of the model is determined using the data. This is done by testing a sequence of hypotheses, and these tests are presented in the paper. 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 model and discuss some of its properties. In Section 3 we derive LM parameter constancy tests against an additive and a multiplicative alternative. In Section 4 we present a modelling strategy for both specifications. Details regarding the estimation are discussed in Section 5, and diagnostic tests for the TV-GARCH model are given in Section 6. Section 7 contains simulation results on the empirical performance of the tests and the specification strategy. In Section 8 we apply our modelling cycle to both stock and exchange rate returns. Finally, Section 9 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 $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = \mu_t$ 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 sigma-field generated by the available information until $t-1$. We assume that $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, because our focus will be on the conditional variance σ_t^2. More precisely, define

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

where $\{\zeta_t\}$ is a sequence of independent standard normal variables. 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. Thus, we assume that h_t follows the standard GARCH(p,q) model of Bollerslev (1986):

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}. \tag{4}$$

Then the GARCH(p,q) model is nested in (2) when $g_t \equiv 0$ and in (3) when $g_t \equiv 1$. More generally, when (3) holds, ε_{t-i}^2 is replaced by $\varepsilon_{t-i}^2/g_{t-i}$, $i = 1,...,q$, in (4). Both parameterizations (2) and (3) 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 by defining the function g_t in (2) as follows:

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

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, $t^* = t/T$, where T is the number of observations. A suitable choice for $G(t^*;\gamma,\mathbf{c})$ is the general logistic smooth transition function defined as follows:

$$G(t^*;\gamma,\mathbf{c}) = \left(1+\exp\left\{-\gamma\prod_{k=1}^{K}(t^*-c_k)\right\}\right)^{-1}, \ \gamma>0, \ c_1\leq c_2\leq ...\leq c_K. \tag{6}$$

This transition function is such that the parameters of the GARCH model (1)-(2) fluctuate smoothly over time between (α_i, β_j) and $(\alpha_i+\alpha_i^*, \beta_j+\beta_j^*)$, $i=0,1,...,q$, $j=1,...,p$. 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. One can observe that large values of γ increase the velocity of transition from 0 to 1 as a function of t^*. When $\gamma\longrightarrow\infty$, a smooth parameter change approaches a structural break because then the process switches instantaneously over time from one regime to another. 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 first change and eventually move back to their original values.

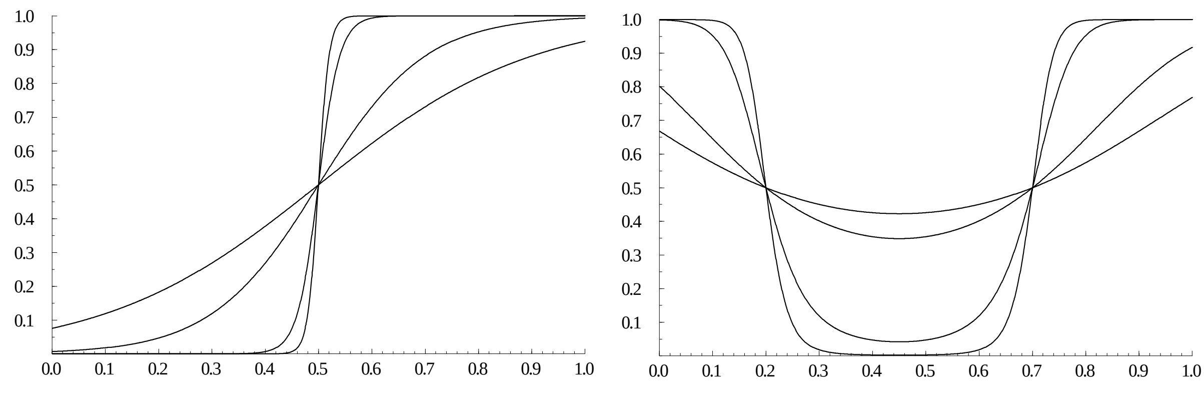


Figure 1. Plots of the logistic transition function (6) for: (a) $K=1$ with location parameter $c_1=0.5$; and (b) $K=2$ with location parameters $c_1=0.2$ and $c_2=0.7$ for $\gamma=5,10,50$, and 100 where the lowest value of γ corresponds to the smoothest function.

More generally, one can define an extended version of the additive TV-GARCH model allowing for more than one transition function. A multiple TV-GARCH model can be obtained by adding r transition functions as follows

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

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 $\alpha_i+\sum_{l=1}^{r}\alpha_{il}>0$, $i=0,...,q$, $i=0,...,q$; $\forall j=1,...,r$ and $\beta_i+\sum_{l=1}^{j}\beta_{il}\geq 0$, $i=1,...,p$; $\forall j=1,...,r$. These conditions are sufficient to guarantee strictly positive conditional variances.

The model (2), (4) and (7) is an additive TV-GARCH model whose intercept, ARCH and GARCH parameters are 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 and probably makes the model difficult to estimate in practice. A more applicable but still flexible model is obtained by only letting the "baseline volatility" or the intercept to 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 (2007) 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 model in the sense that it is in other respects a FIGARCH model, and the authors called it the Adaptive FIGARCH model.

In the GARCH(p, q) model, the unconditional variance of the returns is constant over time, that is, $\mathsf{E}(\varepsilon_t^2) = \alpha_0/(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j)$ if and only if $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$. However, this assumption is not consistent with the behaviour of the volatilities of the stock market returns if the dynamic behaviour of volatility changes in the long run. The additive TV-GARCH model with a time-varying intercept is capable of generating changes in the dynamics of the unconditional variance over time. 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 between α_0 and $\alpha_0 + \sum_{l=1}^{r} \alpha_{0l} G_l(t^*; \gamma_l, \mathbf{c}_l)$. Therefore, it can generate smooth changes over time in the "baseline volatility". Hence, such parameterization can explain the systematic movements of the conditional variance as in the GARCH model but relaxing the assumption of constancy of the unconditional volatility.

Consider again the model (2), (4) and (7) and assume that $\alpha_{0l} = \alpha_0 \delta_l$, $\alpha_{il} = \alpha_i \delta_l$, $i = 1, ..., q$; $\beta_{jl} = \beta_j \delta_l$, $j = 1, ..., p$. Furthermore, assume $\delta_l > 0$, $l = 1, ..., r$, if the transition function $G_l(t^*; \gamma_l, \mathbf{c}_l)$ is increasing over time. For the case $G_l(t^*; \gamma_l, \mathbf{c}_l)$ is a decreasing function assume $\sum_{l=1}^{r} \delta_l < 1$ for $l = 1, ..., r$. Imposing these restrictions on (7) and rewriting (2) yields

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

Setting $g_t = 1 + \sum_{l=1}^{r} \delta_l G_l(t^*; \gamma_l, \mathbf{c}_l)$ in (9) gives the multiplicative representation (3). It is thus seen to be a special case of the additive TV-GARCH model (2), (4) and (7). The multiplicative model has a straightforward interpretation. Writing it in terms of (1) as

$$\phi_t = \varepsilon_t / g_t^{1/2} = \zeta_t {h_t}^{1/2} \tag{10}$$

it is seen that ϕ_t has a constant unconditional variance $\mathsf{E}h_t$ and, moreover, that ϕ_t has a standard stationary GARCH(p, q) representation h_t. Turning (10) around, one obtains that $\psi_t = \varepsilon_t / h_t^{1/2}$, $t = 1, ..., T$, form a sequence of independent but not identically distributed observations, as the unconditional variance of ψ_t changes smoothly as a function of time.

We consider properties of both time-varying GARCH specifications by generating 1000 replications with Gaussian errors each with 5000 observations. Figure 2 illustrates the relation of the average excess kurtosis of the two models given the persistence and the time-varying constants α_{01} and δ_1. The degree of persistence, measured by the sum $\alpha_1+\beta_1$, varies between 0.90 and 0.99. The range of parameters α_{01} and δ_1 varies between 0 and 0.1 while $\alpha_0 = 0.01$. Interestingly, simply by assuming normality the proposed models are capable of generating higher kurtosis than the standard GARCH model. Larger values of the time-varying constants generate larger values of the excess kurtosis for both time-varying parameterizations. A high degree of persistence is also able to reproduce heavy-tailed marginal distributions that are often observed in financial return series.

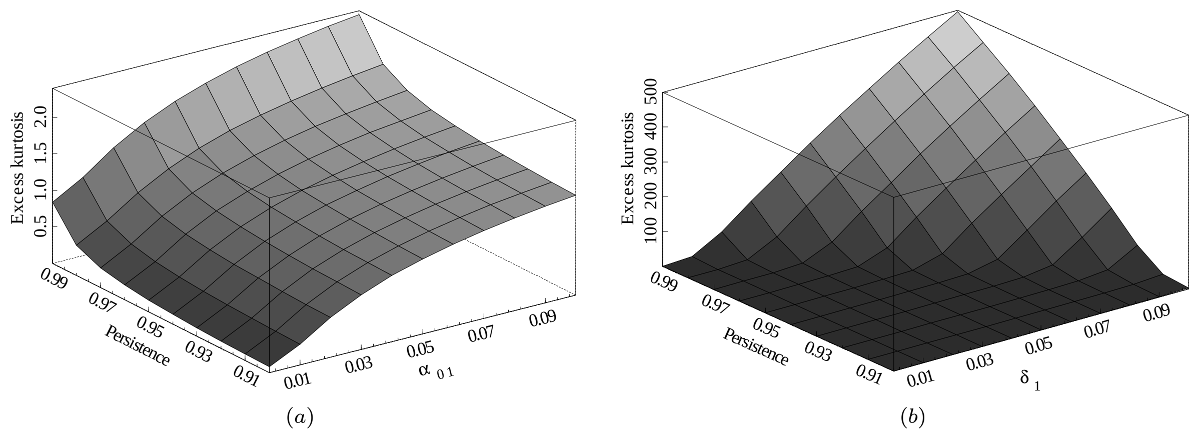


Figure 2. Plots of the excess kurtosis, persistence and the constants α_{01} and δ_1 for: (a) an additive TV-GARCH model with a time-varying constant; and (b) a multiplicative TV-GARCH model.

The level of persistence generated by the TV-GARCH models is another property of interest. Figure 3 depicts the first 100 autocorrelations of absolute returns of two simulated TV-GARCH processes. The autocorrelations for the additive and multiplicative form are plotted in Figure 3(a) and Figure 3(b), respectively. The sample length in both cases is 5000 observations. The artificial series are generated with $\alpha_0 = 0.01$, $\alpha_1 = 0.05$, $\beta_1 = 0.90$, $\alpha_{01} = 0.03$, $\delta_1 = 0.04$, $\gamma_1 = 10$ and $c_1 = 0.50$. The dotted horizontal lines represent the 95% confidence bounds corresponding to the ACF of an iid Gaussian process. A visual

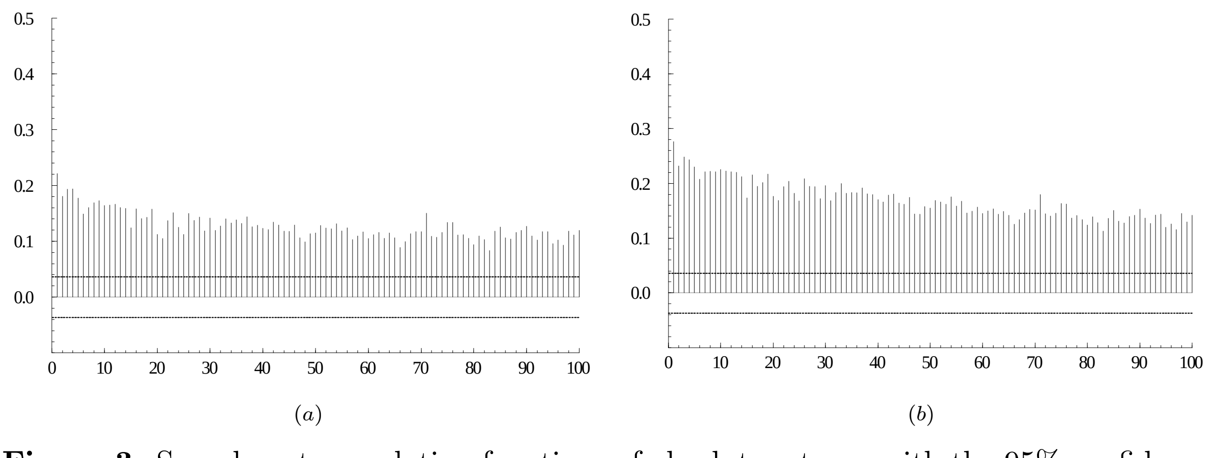


Figure 3. Sample autocorrelation functions of absolute returns with the 95% confidence bounds for: (a) an additive TV-GARCH model with a time-varying constant; and (b) a multiplicative TV-GARCH model.

inspection of Figure 3 shows that both time-varying specifications can generate long-range dependence looking behaviour.

The dependence structure of each model is also illustrated by the empirical distribution of the GPH estimates of the long-memory parameter d; see Geweke and Porter-Hudak (1983). The results obtained by using absolute values of the returns are displayed in Figure 4. The standard GARCH model is known to have a short memory in the sense that the theoretical autocorrelation function decays to zero at an exponential rate. The exponential decay turns out to be too fast if one wants to adequately describe the high persistence observed in financial data. This may be seen from Figure 4(a). If the data are generated by the standard GARCH model, the estimates of the long memory parameter are rather close to zero. However, when the intercept of the GARCH model changes smoothly over time, the degree of the long-memory dependence in the data increases. This is seen from the fact that the empirical distribution for the GPH estimates in Figure 4(b) has shifted to the right. As Figure 4(c) shows, this effect is even more evident for the TV-GARCH with a multiplicative time-varying structure as more than one half of the probability mass of the empirical distribution of the long-memory parameter is located in the nonstationary area, $d > 0.5$.

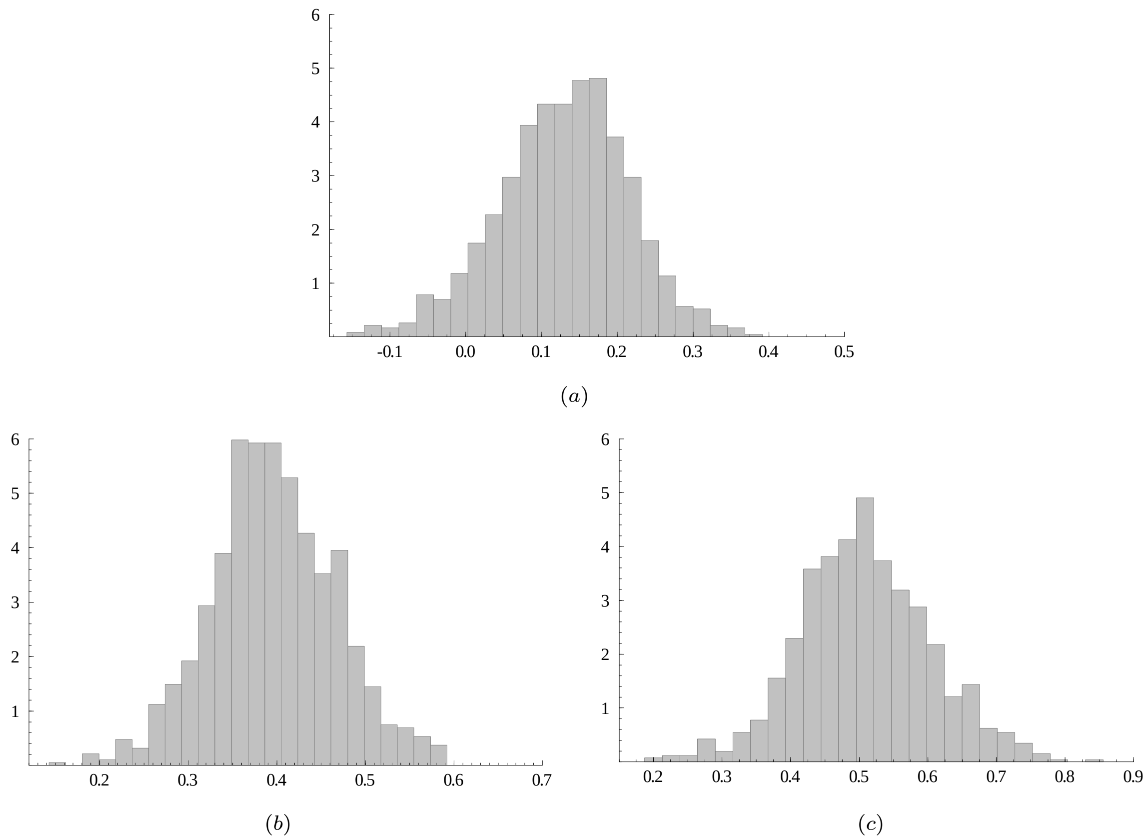


Figure 4. Histograms of the GPH long memory parameter estimates for: (a) a GARCH model; (b) an additive TV-GARCH model with a time-varying constant; and (c) a multiplicative TV-GARCH model. The artificial series are generated with $\alpha_0 = 0.01$, $\alpha_1 = 0.05$, $\beta_1 = 0.90$, $\alpha_{01} = 0.03$, $\delta_1 = 0.04$, $\gamma_1 = 10$ and $c_1 = 0.50$ for a sample of 5000 observations based on 1000 replications.

3 Testing parameter constancy

3.1 Testing against an additive alternative

Against the background discussed above, testing parameter constancy is an important tool for checking the adequacy of a GARCH model. If one rejects parameter constancy against a GARCH model with time-varying parameters one may conclude that the structure of the dynamics of volatility is changing over time. Other interpretations cannot be excluded, however, because a rejection of a null hypothesis does not imply that the alternative hypothesis is true. In this section, we propose two parameter constancy tests that allow the parameters to change smoothly over time under the alternative. The first one tests parameter constancy of the GARCH model against an additive TV-GARCH specification. This idea has previously been considered by Lundbergh and Teräsvirta (2002). The second one is a test of constant unconditional variance against the alternative that the variance changes smoothly over time.

We shall first look at the additive alternative where the nonstationary component g_t is defined in (5). In order to derive the test statistic rewrite the model as

$$\begin{aligned}
\varepsilon_t &= \zeta_t {h_t}^{1/2}, \qquad \varepsilon_t|\mathcal{F}_{t-1} \sim N(0,h_t) \\
h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} + (\alpha_{01} + \sum_{i=1}^{q} \alpha_{i1} \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{j1} h_{t-j}) G(t^*;\gamma,\mathbf{c}) \quad (11)
\end{aligned}$$

where, for simplicity, $r = 1$ and $\mathcal{F}_{t-1}$ is the information set containing all information until $t-1$. The null hypothesis of parameter constancy corresponds to testing $\mathrm{H}_0 : \gamma = 0$ against $\mathrm{H}_1 : \gamma > 0$ in (11). Under the null hypothesis, $g_t \equiv 1/2$. One can see that model (11) is only identified under the alternative. In particular, when $\gamma = 0$, the parameters α_{i1}, $i = 0, ..., q$, and β_{j1}, $j = 1, ..., p$, as well as $\mathbf{c}$ are not identified. This makes the standard asymptotic inference invalid as the test statistics have a nonstandard asymptotic null distribution. This identification problem was first considered in Davies (1977) and more recently, among others, in Hansen (1996).

In this paper, we circumvent the identification problem following Luukkonen, Saikkonen, and Teräsvirta (1988). Thus we replace the transition function by its first-order Taylor approximation around $\gamma = 0$. Without losing generality, we replace $G(t^*;\gamma,\mathbf{c})$ by $\widetilde{G}(t^*;\gamma,\mathbf{c}) = G(t^*;\gamma,\mathbf{c}) - 1/2$ for notational convenience. From Taylor's theorem one obtains

$$\begin{aligned}
\widetilde{G}(t^*;\gamma,\mathbf{c}) &= \widetilde{G}(t^*;0,\mathbf{c}) + \frac{\partial \widetilde{G}(t^*;0,\mathbf{c})}{\partial \gamma}\gamma + R(t^*;\gamma,\mathbf{c}) \\
&= \frac{1}{4}\gamma \prod_{k=1}^{K} (t^* - c_k) + R(t^*;\gamma,\mathbf{c}) \\
&= \sum_{k=0}^{K} \gamma \tilde{c}_k (t^*)^k + R(t^*;\gamma,\mathbf{c}) \qquad (12)
\end{aligned}$$

where $R(t^*;\gamma,\mathbf{c})$ is the remainder term. Replacing $G(t^*;\gamma,\mathbf{c})$ in (11) by (12) and rearranging terms gives

$$h_t = \alpha_0^* + \sum_{i=1}^{q} \alpha_i^* \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j^* h_{t-j} + \sum_{k=1}^{K} \left(\omega_k (t^*)^k + \sum_{i=1}^{q} \varphi_{ik} (t^*)^k \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \lambda_{jk} (t^*)^k h_{t-j} \right) + R_1^* \qquad (13)$$

where $\alpha_s^* = \alpha_s + \gamma\alpha_{s1}\tilde{c}_0$, $s = 0,...,q$, $\beta_j^* = \beta_j + \gamma\beta_{j1}\tilde{c}_0$, $j = 1,...,p$, $\omega_k = \gamma\alpha_{01}\tilde{c}_k$, $\varphi_{ik} = \gamma\alpha_{i1}\tilde{c}_k$, $i = 1,...,q$, and $\lambda_{jk} = \gamma\beta_{j1}\tilde{c}_k$, $k = 1,...,K$. The parameters $\tilde{c}_k$, $k = 0,...,K$, are functions of the original location parameters c_k. In particular, $\tilde{c}_0 = \frac{1}{4}\prod_{k=1}^K c_k$ and $\tilde{c}_K = \frac{1}{4}$. Under H_0, the remainder $R_1^* \equiv 0$, so it does not affect the asymptotic null distribution of the test statistic. Using the reparameterization (13) it follows that the null hypothesis of parameter constancy becomes

$$\mathrm{H}_0' : \omega_k = \varphi_{ik} = \lambda_{jk} = 0,\ k = 1,...,K,\ i = 1,...,q,\ j = 1,...,p. \tag{14}$$

This hypothesis can be tested by a standard LM test. One can also test constancy of a subset of parameters. For example, it may be assumed that $\alpha_{i1} = 0$, $i = 1,...,q$, and $\beta_{j1} = 0$, $j = 1,...,p$, which means that only the intercept is time-varying under the alternative. In this case the null hypothesis reduces to $\mathrm{H}_0' : \omega_k = 0$, $k = 1,...,K$.

In Theorem 1 we present the LM-type statistic for testing parameter constancy against the additive TV-GARCH specification. Under the null hypothesis, the "hats" indicate maximum likelihood estimators and $\hat{h}_t^0$ denotes the conditional variance at time t estimated under H_0.

Theorem 1 *Consider the model (13) and let* $\boldsymbol{\theta}_1 = (\alpha_0^*, \alpha_1^*, ..., \alpha_q^*, \beta_1^*, ..., \beta_p^*)'$ *and* $\boldsymbol{\theta}_2 = (\boldsymbol{\omega}', \boldsymbol{\varphi}_i', \boldsymbol{\lambda}_j')'$ *where* $\boldsymbol{\omega} = (\omega_1, ..., \omega_K)'$, $\boldsymbol{\varphi}_i = (\varphi_{i1}, ..., \varphi_{iK})'$ *and* $\boldsymbol{\lambda}_j = (\lambda_{j1}, ..., \lambda_{jK})'$ *for* $i = 1, ..., q$ *and* $j = 1, ..., p$. *In addition, denote* $\mathbf{z}_t = (1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, h_{t-1}, ..., h_{t-p})'$, $\mathbf{Z}_{1t} = [t^{*k}\varepsilon_{t-i}^2]$ $(k = 1, ..., K,\ i = 1, ..., q)$ *and* $\mathbf{Z}_{2t} = [t^{*k}h_{t-j}]$ $(k = 1, ..., K,\ j = 1, ..., p)$. *Furthermore, assume that the maximum likelihood estimator of* $\boldsymbol{\theta}_1$ *is asymptotically normal. Under* $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$, *the LM type statistic*

$$\xi_{LM} = \frac{1}{2}\sum_{t=1}^T \hat{u}_t\hat{\mathbf{x}}_{2t}' \left\{ \sum_{t=1}^T \hat{\mathbf{x}}_{2t}\hat{\mathbf{x}}_{2t}' - \sum_{t=1}^T \hat{\mathbf{x}}_{2t}\hat{\mathbf{x}}_{1t}' \left(\sum_{t=1}^T \hat{\mathbf{x}}_{1t}\hat{\mathbf{x}}_{1t}' \right)^{-1} \sum_{t=1}^T \hat{\mathbf{x}}_{1t}\hat{\mathbf{x}}_{2t}' \right\}^{-1} \sum_{t=1}^T \hat{u}_t\hat{\mathbf{x}}_{2t} \tag{15}$$

is asymptotically χ^2*-distributed with* $\dim(\boldsymbol{\theta}_2)$ *degrees of freedom, where* $\hat{u}_t = \hat{\varepsilon}_t^2/\hat{h}_t^0 - 1$,

$$\hat{\mathbf{x}}_{1t} = \frac{1}{\hat{h}_t^0}\frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_1}\bigg|_{H_0} = (\hat{h}_t^0)^{-1}(\hat{\mathbf{z}}_t + \sum_{j=1}^p \hat{\beta}_j^* \frac{\partial \hat{h}_{t-j}}{\partial \boldsymbol{\theta}_1}\bigg|_{H_0}) \tag{16}$$

and

$$\hat{\mathbf{x}}_{2t} = \frac{1}{\hat{h}_t^0}\frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_2}\bigg|_{H_0} = (\hat{h}_t^0)^{-1}((t^*, ..., t^{*K}, (\mathrm{vec}\,\mathbf{Z}_{1t})', (\mathrm{vec}\,\mathbf{Z}_{2t})')' + \sum_{j=1}^p \hat{\beta}_j^* \frac{\partial \hat{h}_{t-j}}{\partial \boldsymbol{\theta}_2}\bigg|_{H_0}) \tag{17}$$

Proof. *See Appendix A.* ■

In practice, the test of Theorem 1 may be carried out in a straightforward way using an auxiliary least squares regression. Thus:

1. Estimate consistently the parameters of the conditional variance under the null hypothesis, and compute $\hat{u}_t = \hat{\varepsilon}_t^2/\hat{h}_t^0 - 1$, $t = 1, ..., T$, and the residual sum of squares, $SSR_0 = \sum_{t=1}^T \hat{u}_t^2$.

2. Regress $\hat{u}_t$ on $\hat{\mathbf{x}}_{1t}'$ and $\hat{\mathbf{x}}_{2t}'$, $t = 1, ..., T$, and compute the sum of the squared residuals, SSR_1.

3. Compute the χ^2 test statistic as

$$\xi_{LM} = \frac{T(SSR_0 - SSR_1)}{SSR_0}.$$

As a computational detail, note that $\partial\hat{h}_t/\partial\boldsymbol{\theta}_1|_{H_0}$ and $\partial\hat{h}_t/\partial\boldsymbol{\theta}_2|_{H_0}$ in (16) and (17) are obtained recursively in connection with the parameter estimation, where it is assumed that $\partial\hat{h}_t/\partial\boldsymbol{\theta}_1|_{H_0} = \mathbf{0}$ and $\partial\hat{h}_t/\partial\boldsymbol{\theta}_2|_{H_0} = \mathbf{0}$ for $t = 0, -1, \ldots$. We shall call our LM test statistic LM_K, where K indicates the order of the polynomial in the exponent of the transition function and the tests carried out by means of an auxiliary regression are called LM-type tests.

It should also be mentioned that a robust version of the test statistics (15) can be derived when ζ_t are not identically distributed. One can construct a robust version using the procedure by Wooldridge (1990,1991). This test can be carried out as follows:

1. Estimate by quasi maximum likelihood the conditional variance under H_0, compute $\hat{\varepsilon}_t^2/\hat{h}_t^0 - 1$, $\hat{\mathbf{x}}_{1t}'$ and $\hat{\mathbf{x}}_{2t}'$, $t = 1, \ldots, T$.

2. Regress $\hat{\mathbf{x}}_{2t}$ on $\hat{\mathbf{x}}_{1t}$, and compute the $(\dim \boldsymbol{\theta}_2 \times 1)$ residual vectors $\mathbf{r}_t$, $t = 1, \ldots, T$.

3. Regress 1 on $\left(\hat{\varepsilon}_t^2/\hat{h}_t^0 - 1\right)\mathbf{r}_t$ and compute the residual sum of squares SSR_0 from this regression. Under the null hypothesis, the test statistic $\xi_{LM_R} = T - SSR_0$ has an asymptotic χ^2 distribution with $\dim \boldsymbol{\theta}_2$ degrees of freedom.

One may extend Theorem 1 to the case where the model has been estimated with $r-1$ transition functions and one wants to test $r-1$ against r transitions. For that purpose, consider the model

$$\begin{aligned} \varepsilon_t &= \zeta_t {h_t}^{1/2}, \qquad \varepsilon_t|\mathcal{F}_{t-1} \sim N(0, h_t) \\ h_t &= (\boldsymbol{\theta}_0 + \sum_{l=1}^{r-1} \boldsymbol{\theta}_{1l} G_l(t^*; \gamma_l, \mathbf{c}_l))'\mathbf{z}_t + \boldsymbol{\theta}_{1r}' \widetilde{G}_r(t^*; \gamma_r, \mathbf{c}_r)\mathbf{z}_t \end{aligned} \tag{18}$$

where $\boldsymbol{\theta}_0 = (\alpha_0, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)'$, $\boldsymbol{\theta}_{1l} = (\alpha_{0l}, \alpha_{1l}, \ldots, \alpha_{ql}, \beta_{1l}, \ldots, \beta_{pl})'$, $l = 1, \ldots, r-1, r$, and $\mathbf{z}_t = (1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-q}^2, h_{t-1}, \ldots, h_{t-p})'$. The null hypothesis is then $\text{H}_0 : \gamma_r = 0$. Again, model (18) is not identified under the null hypothesis. To circumvent the problem we proceed as before and expand the logistic function $G_r(t^*; \gamma_r, \mathbf{c}_r)$ into a first-order Taylor approximation around $\gamma_r = 0$. After rearranging terms we have

$$h_t = (\boldsymbol{\eta} + \sum_{l=1}^{r-1} \boldsymbol{\theta}_{1l} G_l(t^*; \gamma_l, \mathbf{c}_l))'\mathbf{z}_t + \sum_{k=1}^{K} \boldsymbol{\mu}_k'(t^*)^k \mathbf{z}_t + R_2^* \tag{19}$$

where $\boldsymbol{\eta} = \boldsymbol{\theta}_0 + \gamma_r \boldsymbol{\theta}_{1r} \tilde{c}_0$, $\boldsymbol{\mu}_k = \gamma_r \boldsymbol{\theta}_{1r} \tilde{c}_k$, $k = 1, \ldots, K$. The test statistic is based on the following corollary of Theorem 1.

Corollary 2 *Consider the model (19) and let* $\boldsymbol{\theta}_1 = (\boldsymbol{\eta}', \boldsymbol{\theta}_{1l}', \gamma_l, \mathbf{c}_l')'$ *and* $\boldsymbol{\theta}_2 = (\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_K')'$. *In addition, denote* $\mathbf{z}_t = (1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-q}^2, h_{t-1}, \ldots, h_{t-p})'$, $\mathbf{Z}_{1t} = [t^{*k} \varepsilon_{t-i}^2]$ $(k = 1, \ldots, K,$ $i = 1, \ldots, q)$, $\mathbf{Z}_{2t} = [t^{*k} h_{t-j}]$ $(k = 1, \ldots, K,$ $j = 1, \ldots, p)$ *and* $G_l(t^*) \equiv G_l(t^*; \gamma_l, \mathbf{c}_l)$. *Assume* *that the maximum likelihood estimator of* $(\boldsymbol{\theta}_0', \boldsymbol{\theta}_{11}', ..., \boldsymbol{\theta}_{1,r-1}', \gamma_1, ..., \gamma_{r-1}, c_1, ..., c_{r-1})'$ *is asymptotically normal. Under* $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$, *the LM type statistic (15) with* $\hat{u}_t = \hat{\varepsilon}_t^2/\hat{h}_t^0 - 1$,

$$\hat{\mathbf{x}}_{1t} = \frac{1}{\hat{h}_t^0}\frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_1}\bigg|_{H_0} = (\hat{h}_t^0)^{-1}(\hat{\mathbf{z}}_t + \sum_{l=1}^{r-1}\hat{\mathbf{z}}_t\hat{G}_l(t^*) + \sum_{l=1}^{r-1}\hat{\boldsymbol{\theta}}_{1l}'\hat{\mathbf{z}}_t\frac{\partial \hat{G}_l(t^*)}{\partial \boldsymbol{\theta}_1} + \sum_{j=1}^{p}(\hat{\beta}_j + \sum_{l=1}^{r-1}\hat{\beta}_{jl}^*\hat{G}_l(t^*))\,\frac{\partial \hat{h}_{t-j}}{\partial \boldsymbol{\theta}_1}\bigg|_{H_0})$$

and

$$\hat{\mathbf{x}}_{2t} = \frac{1}{\hat{h}_t^0}\frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_2}\bigg|_{H_0} = (\hat{h}_t^0)^{-1}((t^*, ..., t^{*K}, (\mathrm{vec}\,\mathbf{Z}_{1t})', (\mathrm{vec}\,\mathbf{Z}_{2t})')' + \sum_{j=1}^{p}(\hat{\beta}_j + \sum_{l=1}^{r-1}\hat{\beta}_{jl}^*\hat{G}_l(t^*))\,\frac{\partial \hat{h}_{t-j}}{\partial \boldsymbol{\theta}_2}\bigg|_{H_0})$$

has an asymptotic χ^2-*distribution with* $\dim(\boldsymbol{\theta}_2)$ *degrees of freedom.*

Remark 3 *The assumption of asymptotic normality in this corollary remains unverified. The existing asymptotic theory of nonlinear GARCH models does not cover the case where the transition function is a function of time. Besides, Meitz and Saikkonen (in press) who have worked out asymptotic theory for smooth transition GARCH models, have only obtained results on ergodicity and stationarity. Asymptotic normality of maximum likelihood estimators has not even been proven for 'standard' smooth transition GARCH models in which the transition variable is a stochastic variable. For these reasons, showing asymptotic normality of* $\boldsymbol{\theta}_1$ *in (19) is beyond the scope of this paper. Two things should be emphasized in this context. First, sequential testing to find* r *is just a model selection device analogous to model selection criteria such as AIC or BIC. The p-values of the tests are simply indicators helping the modeller to choose the number of transitions. Second, our simulation results do not contradict the assumption that the asymptotic null distribution of the test statistic is a* χ^2*-distribution.*

3.2 Testing against a multiplicative alternative

In order to consider the problem of testing parameter constancy in the unconditional variance assume that the error term is parameterized as

$$\varepsilon_t = \zeta_t {h_t}^{1/2}$$

where h_t is a GARCH(p, q) model as in (4) and ζ_t is a time-varying random variable satisfying

$$\zeta_t = z_t g_t^{1/2}$$

such that $\{z_t\}$ is a sequence of independent standard normal variables and $g_t = 1 + \sum_{l=1}^{r}\delta_l G_l(t^*; \gamma_l, \mathbf{c}_l)$. This formulation allows the unconditional variance of ζ_t and thus ε_t to change smoothly over time. As already mentioned, $\{\zeta_t\}$ is a sequence of independent variables. The null hypothesis of constant unconditional variance is then $\mathrm{H}_0 : \delta_l = 0$, $l = 1, ..., r$. For the purpose of deriving the test statistic consider $r = 1$ and rewrite the model as follows:

$$\begin{aligned}
\varepsilon_t &= z_t(h_t g_t)^{1/2}, \qquad \varepsilon_t|\mathcal{F}_{t-1} \sim N(0, h_t g_t) \\
h_t g_t &= (\alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j h_{t-j})(1 + \delta_1\widetilde{G}(t^*; \gamma, \mathbf{c})). \qquad (20)
\end{aligned}$$

The null hypothesis of constant unconditional variance equals $\mathrm{H}_0 : \gamma = 0$ against $\mathrm{H}_1 : \gamma > 0$. In testing this hypothesis we encounter the same identification problem as the one present in testing parameter constancy against an additive TV-GARCH process. Even here, our solution consists of approximating the transition function with a Taylor expansion around $\gamma = 0$. Proceeding as before, we reparameterize equation (20) as follows:

$$h_t g_t = (\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j})(\tilde{\delta}_0 + \sum_{k=1}^{K} \omega_k (t^*)^k + R_3^*) \tag{21}$$

where $\tilde{\delta}_0 = 1 + \gamma\delta_1\tilde{c}_0$ and $\omega_k = \gamma\delta_1\tilde{c}_k$, $k = 1, ..., K$. Under the null hypothesis, the remainder $R_3^* \equiv 0$ and does not affect the distribution theory. The null hypothesis of parameter constancy for the multiplicative structure becomes

$$\mathrm{H}_0' : \omega_k = 0, \;\; k = 1, ..., K.$$

The following corollary of Theorem 1 defines the LM-type test statistic for testing parameter constancy in the unconditional variance. The notation $\hat{g}_t^0$ denotes the estimated g_t evaluated under H_0.

Corollary 4 *Consider the model (21) and let* $\boldsymbol{\theta}_1 = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)'$ *and* $\boldsymbol{\theta}_2 = (\omega_1, ..., \omega_K)'$. *In addition, denote* $\mathbf{z}_t = (1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, h_{t-1}, ..., h_{t-p})'$ *and* $g_t = 1 + \delta_1 G(t^*; \gamma, \mathbf{c})$. *Under* $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$, *the LM type statistic (15) with* $\hat{u}_t = \hat{\varepsilon}_t^2/\hat{h}_t^0 - 1$,

$$\hat{\mathbf{x}}_{1t} = \frac{1}{\hat{h}_t^0} \frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_1}\bigg|_{H_0} = (\hat{h}_t^0)^{-1}(\hat{\mathbf{z}}_t + \sum_{j=1}^{p} \hat{\beta}_j^* \frac{\partial \hat{h}_{t-j}}{\partial \boldsymbol{\theta}_1}\bigg|_{H_0})$$

and

$$\hat{\mathbf{x}}_{2t} = \frac{1}{\hat{g}_t^0} \frac{\partial \hat{g}_t}{\partial \boldsymbol{\theta}_2}\bigg|_{H_0} = (t^*, t^{*2}, ..., t^{*K})'$$

has an asymptotic χ^2-*distribution with* $\dim(\boldsymbol{\theta}_2)$ *degrees of freedom.*

Once the TV-GARCH model with a single transition has been estimated we may want to investigate the possibility of remaining parameter nonconstancy in the unconditional variance. This is important from the model specification point of view. Thus, similarly to the additive structure, the previous corollary may be extended to the case where we want to test $r = 1$ against $r \geq 2$. To derive the test, consider the model

$$\begin{aligned} \varepsilon_t &= z_t (h_t g_t)^{1/2}, \qquad \varepsilon_t | \mathcal{F}_{t-1} \sim N(0, h_t g_t) \\ h_t g_t &= (\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j})(1 + \textstyle\sum_{l=1}^{2} \delta_l G_l(t^*; \gamma_l, \mathbf{c}_l)). \end{aligned} \tag{22}$$

The null hypothesis is $\mathrm{H}_0 : \gamma_2 = 0$. Again, model (22) is only identified under the alternative. The solution to the identification problem consists of replacing the transition function $G_2(t^*; \gamma_2, \mathbf{c}_2)$ by a Taylor approximation around $\gamma_2 = 0$. After a reparameterization, the resulting model is

$$h_t g_t = (\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j})(\tilde{\delta}_0 + \delta_1 G_1(t^*; \gamma_1, \mathbf{c}_1) + \sum_{k=1}^{K} \omega_k (t^*)^k + R_4^*) \tag{23}$$

where $\tilde{\delta}_0 = 1 + \gamma_2\delta_2\tilde{c}_0$ and $\omega_k = \gamma_2\delta_2\tilde{c}_k$, $k = 1, ..., K$. Under the null, the remainder $R_4^* \equiv 0$.

The next corollary to Theorem 1 gives the test statistic.

Corollary 5 *Consider the model (23) and let* $\boldsymbol{\theta}_1 = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p, \delta_1, \gamma_1, \mathbf{c}_1')'$ *and* $\boldsymbol{\theta}_2 = (\omega_1, ..., \omega_K)'$. *In addition, denote* $\mathbf{z}_t = (1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, h_{t-1}, ..., h_{t-p})'$ *and* $g_t = 1 + \sum_{l=1}^{2} \delta_l G_l(t^*; \gamma_l, \mathbf{c}_l)$. *Under* $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$, *the LM type statistic (15) with* $\hat{u}_t = \hat{\varepsilon}_t^2 / \hat{h}_t^0 \hat{g}_t^0 - 1$,

$$\hat{\mathbf{x}}_{1t} = \frac{1}{\hat{h}_t^0} \frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_1}\bigg|_{H_0} = (\hat{h}_t^0)^{-1} (\hat{\mathbf{z}}_t \hat{g}_t^0 + \hat{h}_t^0 \frac{\partial \hat{g}_t^0}{\partial \boldsymbol{\theta}_1} + \sum_{j=1}^{p} \hat{\beta}_j \hat{g}_t^0 \frac{\partial \hat{h}_{t-j}}{\partial \boldsymbol{\theta}_1}\bigg|_{H_0})$$

and

$$\hat{\mathbf{x}}_{2t} = \frac{1}{\hat{g}_t^0} \frac{\partial \hat{g}_t}{\partial \boldsymbol{\theta}_2}\bigg|_{H_0} = (\hat{g}_t^0)^{-1} (t^*, t^{*2}, ..., t^{*K})'$$

has an asymptotic χ^2-*distribution with* $\dim(\boldsymbol{\theta}_2)$ *degrees of freedom.*

Remark 6 *The previous remark is valid even here.*

A special case of this test, in which $h_t \equiv \alpha_0$, will be used in the specification of multiplicative TV-GARCH models in Section 4.2.

4 Model specification

We propose a model-building cycle for TV-GARCH models identical to the specific-to-general strategy for nonlinear models recommended by Granger (1993) or Teräsvirta (1998), among others. The idea is to begin with a parsimonious model and proceed to more complicated ones until the evaluation techniques indicate that an adequate model has been obtained. Adapting this approach to the present situation means determining the number of smooth transitions sequentially by LM-type tests discussed in Section 3. These tests can be used to build a GARCH model with time-varying parameters using either the additional or the multiplicative structure. We start off with a restricted specification and gradually increase the number of transition functions as long as the hypothesis of parameter constancy is rejected. The final model is estimated after the first non-rejection of the null hypothesis and evaluated through a sequence of misspecification tests.

4.1 Specification of additive TV-GARCH models

In order to describe the specification procedure for TV-GARCH models with an additional time-varying structure, we consider the function g_t defined in (7) such that all parameters are changing smoothly over time. However, the strategy may also be applied to a more restrictive functions such as g_t in (8). The time-varying conditional variance equals

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} + \sum_{l=1}^{r} (\alpha_{0l} + \sum_{i=1}^{q} \alpha_{il} \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{jl} h_{t-j}) G_l(t^*; \gamma_l, \mathbf{c}_l), \quad (24)$$

where the transition function $G_l(t^*; \gamma_l, \mathbf{c}_l)$ is defined in (6).

Our specification procedure for building additive TV-GARCH models contains the following stages:

1. Check for the presence of conditional heteroskedasticity by testing the null hypothesis of no ARCH against high-order ARCH. When the order of the ARCH process is sufficiently high, the standard LM test has adequate power against GARCH. If

the null hypothesis is rejected, model the conditional variance by a GARCH(1,1) model. Evaluate the estimated GARCH(1,1) model by misspecification tests and, if necessary, expand it to a higher-order model. The squared standardized errors of the selected GARCH model should be free of serial correlation. Neglected autocorrelation may bias tests of parameter constancy.

2. Test the final GARCH model against the alternative of smoothly changing parameters over time using the LM-type statistic described in Theorem 1. If parameter constancy is rejected at a predetermined significance level α, estimate the TV-GARCH model (24) with a single transition function. If the null hypothesis of parameter constancy in (14) is rejected, the problem of choosing the order of the polynomial of the transition function arises. For the specification of K, we propose a model selection rule based on a sequence of nested tests as in Teräsvirta (1994) and Lin and Teräsvirta (1994). Assume $K = 3$ to ensure a parameterization sufficiently flexible for $G(t^*; \gamma, \mathbf{c})$. If parameter constancy is rejected, test the following sequence of hypotheses:

$$
\begin{aligned}
\mathrm{H}_{03} &: \omega_3 = 0,\ \varphi_{i3} = 0,\ \lambda_{j3} = 0, \\
\mathrm{H}_{02} &: \omega_2 = 0,\ \varphi_{i2} = 0,\ \lambda_{j2} = 0 \ \mid\ \omega_3 = 0,\ \varphi_{i3} = 0,\ \lambda_{j3} = 0, \\
\mathrm{H}_{01} &: \omega_1 = 0,\ \varphi_{i1} = 0,\ \lambda_{j1} = 0 \ \mid\ \omega_2 = \omega_3 = 0,\ \varphi_{i2} = \varphi_{i3} = 0,\ \lambda_{j2} = \lambda_{j3} = 0,
\end{aligned}
$$

where $i = 1, ..., q,\ j = 1, ..., p$, in (13), by means of LM-type tests. The results of this test sequence may be used as follows. If H_{01} and H_{03} are rejected more strongly, measured by p-values, than H_{02}, then either $K = 1$ or $K = 3$. If testing H_{02} yields the strongest rejection, the choice is $K = 2$. Furthermore, if only H_{01} is rejected at the appropriate significance level or is rejected clearly more strongly than the other two null hypotheses, then the modeller should choose $K = 1$. Visual inspection of the return series is also helpful in making a decision about K. The rules or suggestions based on p-values are based on expressions of the parameters ω_k, φ_{ik} and λ_{jk} in the auxiliary regression as functions of the original parameters at different values of K. The test sequence is analogous to that proposed in Teräsvirta (1994) for specifying the type of the smooth transition autoregressive model, where the choice is between $K = 1$ and $K = 2$.

3. Test the TV-GARCH model with one transition function against the TV-GARCH model with two transition functions at the significance level $\alpha\tau$, $0 < \tau < 1$. The significance level is decreased giving a preference for parsimonious models. The overall significance level of the sequence of tests may be approximated by the Bonferroni upper bound. The user can choose the value for τ. In our simulations we set $\tau = 1/2$. If the null hypothesis is rejected, specify K for the next transition and estimate the TV-GARCH model (12) with two transition functions.

4. Proceed sequentially by testing the TV-GARCH model with $r - 1$ transition functions against the TV-GARCH model with r transitions at the significance level $\alpha\tau^{r-1}$ until the first non-rejection of the null hypothesis. Evaluate the selected model by misspecification tests and once it passes them accept it as the final model. In the opposite case, modify the specification of the model or try another family of models.

4.2 Specification of multiplicative TV-GARCH models

The specific-to-general approach for specifying TV-GARCH models with a multiplicative time-varying component consists in first modelling the unconditional variance as follows:

1. Use the LM-type statistic developed in Section 3.2 to test the null hypothesis of constant variance against a time-varying unconditional variance with a single transition function at the significance level α. First, assume $h_t = \alpha_0$ and test $\mathrm{H}_{10} : g_t \equiv 1$ against $\mathrm{H}_{11} : g_t = 1 + \delta_1 G_1(t^*; \gamma_1, \mathbf{c}_1)$. In case of a rejection, 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. The significance level is reduced at each step of the testing procedure and converging to zero for reasons previously mentioned.

2. After specifying g_t, test the null hypothesis of no conditional heteroskedasticity in $\{\zeta_t\}$. If it is rejected, model the conditional variance h_t of the standardized variable $\varepsilon_t / g_t^{1/2}$ in the standard fashion, such that

$$h_t = \alpha_0^* + \sum_{i=1}^{q} \alpha_i \left(\frac{\varepsilon_{t-i}^2}{g_{t-i}} \right) + \sum_{j=1}^{p} \beta_j h_{t-j}. \tag{25}$$

3. The estimated model is evaluated by means of LM-type diagnostic tests proposed by Lundbergh and Teräsvirta (2002). If the model passes all the misspecification tests, tentatively accept it. Otherwise, modify it or consider another family of volatility models.

5 Estimation of the TV-GARCH model

Suppose that ε_t is generated by a GARCH model with a time-varying structure described in Section 2. Let $h_t = h_t(\boldsymbol{\theta}_1)$ and $g_t = g_t(\boldsymbol{\theta}_2)$ where $\boldsymbol{\theta}_1 = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)'$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\delta}', \boldsymbol{\alpha}_1', ..., \boldsymbol{\alpha}_r', \boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_r', \gamma_1, ..., \gamma_r, \mathbf{c}_1, ..., \mathbf{c}_r)'$ with $\boldsymbol{\delta} = (\delta_1, ..., \delta_r)'$, $\boldsymbol{\alpha}_i = (\alpha_{1i}, ..., \alpha_{qi})'$ and $\boldsymbol{\beta}_i = (\beta_{1i}, ..., \beta_{pi})'$, $i = 1, ..., r$. For the additive parameterization, $\boldsymbol{\delta} = \mathbf{0}$ and for the multiplicative one, $\boldsymbol{\alpha}_i = \mathbf{0}$ and $\boldsymbol{\beta}_i = \mathbf{0}$. The quasi maximum likelihood (QML) estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1', \widehat{\boldsymbol{\theta}}_2')'$ is obtained maximizing $\sum_{t=1}^{T} \ell_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ where the log-likelihood for observation t equals

$$\ell_t(\boldsymbol{\theta}) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln\{h_t(\boldsymbol{\theta}_1) + g_t(\boldsymbol{\theta}_2)\} - \frac{1}{2} \frac{\varepsilon_t^2}{h_t(\boldsymbol{\theta}_1) + g_t(\boldsymbol{\theta}_2)} \tag{26}$$

for the additive TV-GARCH model or

$$\ell_t(\boldsymbol{\theta}) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \{\ln h_t(\boldsymbol{\theta}_1) + \ln g_t(\boldsymbol{\theta}_2)\} - \frac{1}{2} \frac{\varepsilon_t^2}{h_t(\boldsymbol{\theta}_1) g_t(\boldsymbol{\theta}_2)} \tag{27}$$

for the multiplicative TV-GARCH model.

The asymptotic properties of the QML estimators for the GARCH(p, q) process have been studied, among others, by Ling and Li (1997). They showed that the QML estimators are consistent and asymptotic normal provided that $\mathsf{E}\varepsilon_t^4 < \infty$. Ling and McAleer (2003) established consistency for the global maximum of QML estimators under the condition

$\mathsf{E}\varepsilon_t^2 < \infty$. Berkes, Horváth, and Kokoszka (2003) obtained consistency of the QML estimators assuming $\mathsf{E}\varepsilon_t^2 < \infty$ and asymptotic normality by assuming $\mathsf{E}\varepsilon_t^4 < \infty$. These results have in common the assumption that the process y_t is stationary and ergodic such that the laws of large numbers apply. More recently, Jensen and Rahbek (2004) relaxed this assumption and allowed the parameters to lie in the region where the process is nonstationary. They showed that for the GARCH(1,1) case, under a finite conditional variance for ζ_t^2, consistency and asymptotic normality still hold independently of whether the process y_t is stationary or not. As already mentioned, asymptotic normality for the parameter estimators of the TV-GARCH models has not yet been proven.

Three 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 $\mathbf{c}_l$ 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 $\mathbf{c}_l$. 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 smoothness 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. This can be done without significantly affecting the value of the test statistic.

The second comment 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 is common practice to use numerical derivates that are relatively fast to compute and reliable, and the derivation of exact analytic derivatives is avoided. 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.

The third remark is related to the manner in which the parameter estimates are obtained. The parameters in the additive TV-GARCH model are estimated simultaneously by full conditional maximum likelihood. In this context, care is required in the estimation. Since the log-likelihood (27) may contain several local maxima, it is advisable to initiate the estimation from different sets of starting-values before settling for the final parameter estimates. Numerical problems in the estimation of the multiplicative TV-GARCH model can be alleviated by concentrating the likelihood iteratively. This considerably reduces the dimensionality problem and is computationally much easier than maximizing the log-likelihood with respect to all parameters simultaneously. The estimation of the TV-GARCH model with multiplicative structure can be simplified since the log-likelihood can be decomposed into two separate sets of parameters: the GARCH and the time-varying parameter vectors. The estimation is divided into two steps which are then repeated one after the other. The iterations start by first estimating $\boldsymbol{\theta}_2$, assuming h_t to be a positive constant, for instance $h_t = \hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\varepsilon_t^2$, and continue by estimating $\boldsymbol{\theta}_1$, given the estimates of $\boldsymbol{\theta}_2$. The estimate of $\boldsymbol{\theta}_1$ will then be used for re-estimating $\boldsymbol{\theta}_2$, and so on. The iterative two-stage estimation procedure is terminated when a local maximum of the log-likelihood has been reached.

6 Misspecification testing of TV-GARCH models

The final step of the modelling strategy consists of evaluating the adequacy of the estimated TV-GARCH model by means of a sequence of misspecification tests. We shall assume that the true process of either the additive or the multiplicative time-varying variance is misspecified. The general idea is to construct an augmented version of the TV-GARCH model by introducing a new component $f_t = f(\mathbf{v}_t; \boldsymbol{\theta}_3)$ into the original model. This component is a function that is at least twice continuously differentiable with respect to the elements of $\boldsymbol{\theta}_3$, vector of additional parameters. The vector $\mathbf{v}_t$ is a vector of omitted random variables, and its definition varies from one test to the next.

6.1 Misspecification tests for the multiplicative model

The misspecification tests considered here may be divided into three categories. The first two correspond to additive and the third one to multiplicative misspecification. Let $h_t = h_t(\boldsymbol{\theta}_1)$ and $g_t = g_t(\boldsymbol{\theta}_2)$, such that the parameter vectors $\boldsymbol{\theta}_i$, $i = 1, 2$, represent the parameters belonging to $h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j h_{t-j}$ and $g_t = 1 + \sum_{l=1}^{r}\delta_l G_l(t^*; \gamma_l, \mathbf{c}_l)$. Under $\mathrm{H}_0 : \boldsymbol{\theta}_3 = \mathbf{0}$, the augmented model reduces to the multiplicative TV-GARCH model.

6.1.1 *Additive misspecification - case 1*

The first category of tests assumes that, under the alternative hypothesis, the original TV-GARCH model may be extended by assuming

$$\varepsilon_t = \zeta_t (h_t + f_t)^{1/2} g_t^{1/2}. \tag{28}$$

Under the null hypothesis, $f_t \equiv 0$, which is equivalent to $\boldsymbol{\theta}_3 = \mathbf{0}$. If $g_t \equiv 1$, the test collapses into the additive misspecification test in Lundbergh and Teräsvirta (2002). At least three types of alternative hypotheses can be considered within this family of tests. The test of the GARCH(p, q) component against higher-order alternatives as well as the test against a smooth transition GARCH (ST-GARCH) and, furthermore, the test against an asymmetric component (GJR-GARCH) belong to the additive class (28).

The log-likelihood function for observation t of model (28) is

$$\ell_t = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\{\ln(h_t + f_t) + \ln g_t\} - \frac{\varepsilon_t^2}{2(h_t + f_t)g_t}. \tag{29}$$

When the estimated multiplicative TV-GARCH model is tested against the different types of alternatives, the first component of the score corresponding to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, evaluated under H_0, is equal to

$$\left.\frac{\partial \ell_t}{\partial \boldsymbol{\theta}}\right|_{H_0} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t g_t} - 1\right)\mathbf{x}_{1t}$$

where $\mathbf{x}_{1t} = \left(\frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}_1'}, \frac{1}{g_t}\frac{\partial g_t}{\partial \boldsymbol{\theta}_2'}\right)'$ and the parameter vector $\boldsymbol{\theta}$ is partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$. The estimated quantities for $\frac{\partial h_t}{\partial \boldsymbol{\theta}_1}|_{H_0}$ and $\frac{\partial g_t}{\partial \boldsymbol{\theta}_2}|_{H_0}$ are defined as

$$\left.\frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_1}\right|_{H_0} = \hat{\mathbf{z}}_t + \sum_{j=1}^{p} \hat{\beta}_j \left.\frac{\partial \hat{h}_{t-j}}{\partial \boldsymbol{\theta}_1}\right|_{H_0} \tag{30}$$

$$\left.\frac{\partial \hat{g}_t}{\partial \boldsymbol{\theta}_2}\right|_{H_0} = \sum_{l=1}^{r} G_l(t^*; \hat{\gamma}_l, \hat{\mathbf{c}}_l) + \sum_{l=1}^{r} \hat{\delta}_l \frac{\partial G_l(t^*; \hat{\gamma}_l, \hat{\mathbf{c}}_l)}{\partial \boldsymbol{\theta}_2}. \tag{31}$$

The differences show up in the partial derivatives of (29) with respect to $\boldsymbol{\theta}_3$. It follows that the additional block of the score for observation t due to $\boldsymbol{\theta}_3$ has the form

$$\frac{\partial \ell_t}{\partial \boldsymbol{\theta}_3} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{(h_t + f_t)g_t} - 1\right)\frac{1}{h_t}\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}$$

so that, under H_0,

$$\left.\frac{\partial \ell_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t g_t} - 1\right)\frac{1}{h_t}\left.\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0}$$

where $\frac{\partial f_t}{\partial \boldsymbol{\theta}_3} = \mathbf{v}_t$. The resulting LM test may be easily performed using an auxiliary regression as in Section 3. In terms of previous notation, we have

$$\hat{\mathbf{x}}_{1t} = \left(\frac{1}{\hat{h}_t^0}\left.\frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_1'}\right|_{H_0}, \frac{1}{\hat{g}_t^0}\left.\frac{\partial \hat{g}_t}{\partial \boldsymbol{\theta}_2'}\right|_{H_0}\right)' \tag{32}$$

$$\hat{\mathbf{x}}_{2t} = \frac{1}{\hat{h}_t^0}\left.\frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0} = \frac{\hat{\mathbf{v}}_t}{\hat{h}_t^0} \tag{33}$$

where $\frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_1}|_{H_0}$ and $\frac{\partial \hat{g}_t}{\partial \boldsymbol{\theta}_2}|_{H_0}$ are as in (30) and (31), respectively. We shall now concentrate our attention on tests against higher-order alternatives and a smooth transition GARCH model.

Testing the GARCH(p, q) component against higher-order alternatives

An evident source of misspecification is to select too low an order in the GARCH(p, q) component. A similar testing procedure to the one proposed by Bollerslev (1986) for testing a GARCH(p, q) model against higher-order alternatives is presented. Under the alternative GARCH$(p,\ q + r)$, the additional component equals

$$f_t = \sum_{i=q+1}^{q+r} \alpha_i \varepsilon_{t-i}^2 \tag{34}$$

or

$$f_t = \sum_{j=p+1}^{p+r} \beta_j h_{t-j} \tag{35}$$

if we take the GARCH$(p + r, q)$ as alternative. The identification problem discussed in Bollerslev (1986) prevents us from considering the alternative GARCH$(p + r, q + s)$, $r, s > 0$. Under the null hypothesis $\mathrm{H}_0 : \boldsymbol{\theta}_3 = \mathbf{0}$, i.e. $\alpha_{q+1} = ... = \alpha_{q+r} = 0$ for the former case and $\beta_{p+1} = ... = \beta_{p+r} = 0$ for the latter case, the models reduce to the GARCH(p,q) model.

Corollary 7 defines the test statistic for testing $\alpha_{q+1} = ... = \alpha_{q+r} = 0$. A similar result holds for testing $\beta_{p+1} = ... = \beta_{p+r} = 0$ in (35) and can be stated by replacing $\boldsymbol{\theta}_3 = (\alpha_{q+1}, ..., \alpha_{q+r})'$ and $\hat{\mathbf{v}}_t = (\varepsilon^2_{t-(q+1)}, ..., \varepsilon^2_{t-(q+r)})'$ in Corollary 7 by $\boldsymbol{\theta}_3 = (\beta_{p+1}, ..., \beta_{p+r})'$ and $\hat{\mathbf{v}}_t = (h_{t-(p+1)}, ..., h_{t-(p+r)})'$.

Corollary 7 *Consider the model (28) where $\{\zeta_t\}$ is a sequence of independent standard normal variables. Let $\boldsymbol{\theta}_1 = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)'$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\delta}', \gamma_1, ..., \gamma_r, \mathbf{c}_1, ..., \mathbf{c}_r)'$ with $\boldsymbol{\delta} = (\delta_1, ..., \delta_r)'$. Furthermore, f_t is defined by (34) such that $\boldsymbol{\theta}_3 = (\alpha_{q+1}, ..., \alpha_{q+r})'$ and $\hat{\mathbf{v}}_t = (\varepsilon^2_{t-(q+1)}, ..., \varepsilon^2_{t-(q+r)})'$. Assume that the maximum likelihood estimators of the parameters of (28) are asymptotically normal when $H_0 : \boldsymbol{\theta}_3 = \mathbf{0}$ is valid. Thus, under this null hypothesis, the LM statistic (15), with $\hat{u}_t = \hat{\varepsilon}^2_t/\hat{h}^0_t \hat{g}^0_t - 1$, $\hat{\mathbf{x}}_{1t}$ as in (32) and $\hat{\mathbf{x}}_{2t}$ as in (33) is asymptotically χ^2-distributed with r degrees of freedom.*

Remark 8 *Note that the result stated in Corollary 7 depend on an assumption of asymptotic normality which so far remains unproven. Asymptotic normality has, however, been proven in the special case $\boldsymbol{\theta}_2 = \mathbf{0}$ when the null model (28) is a standard GARCH(p,q) model. A similar remark will hold for Corollaries 9, 10, 11 and 12.*

Testing the GARCH(p, q) component against a nonlinear specification

It is possible that responses of volatility in financial series to negative and positive shocks are not symmetric around zero (or some other value). The GARCH literature offers a variety of parameterizations for describing asymmetric effects of shocks on the conditional variance. The ST-GARCH model, discussed in Hagerud (1997), González-Rivera (1998) and Anderson, Nam, and Vahid (1999), is one of them. Symmetry of the estimated TV-GARCH can be tested against asymmetry or, more generally, against nonlinearity, using these models as alternatives. To this end, let

$$f_t = \sum_{i=1}^{q} (\alpha^*_{1i} + \alpha^*_{2i}\varepsilon^2_{t-i}) G(\varepsilon_{t-i}; \gamma, \mathbf{c}) \tag{36}$$

where $G(\varepsilon_{t-i}; \gamma, \mathbf{c})$ is the transition function given in (6) with ε_{t-i} as the transition variable. With the purpose of simplifying the derivation of the test we replace $G(\varepsilon_{t-i}; \gamma, \mathbf{c})$ by $\widetilde{G}(\varepsilon_{t-i}; \gamma, \mathbf{c}) = G(\varepsilon_{t-i}; \gamma, \mathbf{c}) - 1/2$. The null hypothesis of linearity is $\mathrm{H}_0 : \gamma = 0$ under which $G(\varepsilon_{t-i}; \gamma, \mathbf{c}) \equiv 1/2$. However, the remaining parameters in (36) are not identified under the null hypothesis. Again the identification problem may be circumvented using a Taylor series approximation of the transition function around $\gamma = 0$. After rearranging terms, one obtains

$$h_t + f_t = \alpha^*_0 + \sum_{i=1}^{q} \alpha^*_i \varepsilon^2_{t-i} + \sum_{j=1}^{p} \beta_j h_{t-j} + \sum_{i=1}^{q} \sum_{k=1}^{K} (\varpi_{ik} \varepsilon^k_{t-i} + \pi_{ik} \varepsilon^{k+2}_{t-i}) + R^*_5 \tag{37}$$

where $\alpha^*_0 = \alpha_0 + \sum_{i=1}^{q} \gamma \alpha^*_{1i} \tilde{c}_0$, $\alpha^*_i = \alpha_i + \gamma \alpha^*_{2i} \tilde{c}_0$, $\varpi_{ik} = \gamma \alpha^*_{1i} \tilde{c}_k$ and $\pi_{ik} = \gamma \alpha^*_{2i} \tilde{c}_k$. The component given in (36) can be rewritten as

$$f_t = \sum_{i=1}^{q} \sum_{k=1}^{K} (\varpi_{ik} \varepsilon^k_{t-i} + \pi_{ik} \varepsilon^{k+2}_{t-i}) + R^*_5 \tag{38}$$

When the null hypothesis holds, the remainder R_5^* vanishes, and so does not affect the distributional properties of the test. Using this notation, the hypothesis of no additional nonlinear structure becomes $\mathrm{H}_0' : \varpi_{ik} = \pi_{ik} = \mathbf{0}$, $i = 1, ..., q$, $k = 1, ..., K$. The next corollary gives the test statistic.

Corollary 9 *Consider the model (28) where $\{\zeta_t\}$ is a sequence of independent standard normal variables. Let $\boldsymbol{\theta}_1 = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)'$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\delta}', \gamma_1, ..., \gamma_r, \mathbf{c}_1, ..., \mathbf{c}_r)'$ with $\boldsymbol{\delta} = (\delta_1, ..., \delta_r)'$. Furthermore, f_t is defined by (38) such that $\boldsymbol{\theta}_3 = (\boldsymbol{\varpi}_i', \boldsymbol{\pi}_i')'$, where $\boldsymbol{\varpi}_i = (\varpi_{i1}, ..., \varpi_{iK})'$ and $\boldsymbol{\pi}_i = (\pi_{i1}, ..., \pi_{iK})'$, $i = 1, ..., q$. In addition, let $\hat{\mathbf{v}}_t = (\hat{\mathbf{v}}_{1,t}', ..., \hat{\mathbf{v}}_{K+2,t}')'$ with $\mathbf{v}_{it} = (\varepsilon_{t-1}^i, ..., \varepsilon_{t-q}^i)'$, $i = 1, ..., K+2$. Assume that the maximum likelihood estimators of the parameters of (28) are asymptotically normal when $H_0 : \boldsymbol{\theta}_3 = \mathbf{0}$ is valid. Thus, under this null hypothesis, the LM statistic (15), with $\hat{u}_t = \hat{\varepsilon}_t^2/\hat{h}_t^0\hat{g}_t^0 - 1$, $\hat{\mathbf{x}}_{1t}$ as in (32) and $\hat{\mathbf{x}}_{2t}$ as in (33) is asymptotically χ^2-distributed with* $\dim(\boldsymbol{\theta}_3)$ *degrees of freedom.*

6.1.2 *Additive misspecification - case 2*

We shall now consider the case in which the true model has the following form:

$$\varepsilon_t = \zeta_t h_t^{1/2}(g_t + f_t)^{1/2}. \tag{39}$$

Under the null hypothesis, $f_t \equiv 0$, which is again equivalent to $\boldsymbol{\theta}_3 = \mathbf{0}$. The model again reduces to (1) and (3). The log-likelihood for the observation t equals

$$\ell_t = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\{\ln h_t + \ln(g_t + f_t)\} - \frac{\varepsilon_t^2}{2h_t(g_t + f_t)}.$$

The block of the score containing the first partial derivatives with respect to $\boldsymbol{\theta}_3$ is

$$\frac{\partial \ell_t}{\partial \boldsymbol{\theta}_3} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t(g_t + f_t)} - 1\right)\frac{1}{g_t}\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}$$

which, under H_0, is equal to

$$\left.\frac{\partial \ell_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t g_t} - 1\right)\frac{1}{g_t}\left.\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0}.$$

For this alternative, the quantity $\hat{\mathbf{x}}_{1t}$ is defined as in (32) and

$$\hat{\mathbf{x}}_{2t} = \frac{1}{\hat{g}_t^0}\left.\frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0} = \frac{\hat{\mathbf{v}}_t}{\hat{g}_t^0}. \tag{40}$$

Testing the hypothesis of no additional transitions

Once the TV-GARCH model has been estimated, one may use this set-up, for example, to re-check the need for another transition function in g_t. Taking the multiplicative TV-GARCH model with $r + s$ transitions as the alternative, it follows that

$$f_t = \sum_{l=r+1}^{r+s} \delta_l G_l(t^*; \gamma_l, \mathbf{c}_l) \tag{41}$$

The hypothesis of no additional transitions is $\mathrm{H}_0 : \gamma_{r+1} = ... = \gamma_{r+s} = 0$. Under this hypothesis, the parameters $(\delta_l, \mathbf{c}_l')'$ are not identified. To circumvent this problem, we replace the transition function $G_l(t^*;\gamma_l,\mathbf{c}_l)$ by its first-order Taylor expansion around $\gamma_l = 0$, $l = r+1,...,r+s$. After merging terms, we obtain

$$\begin{aligned} g_t + f_t &= 1 + \sum_{l=1}^{r} \delta_l G_l(t^*;\gamma_l,\mathbf{c}_l) + \sum_{l=r+1}^{r+s} \delta_l(\gamma_l \tilde{c}_0 + \sum_{k=1}^{K} \gamma_l \tilde{c}_k (t^*)^k) + R_6^* \\ &= \delta_l^* + \sum_{l=1}^{r} \delta_l G_l(t^*;\gamma_l,\mathbf{c}_l) + \sum_{l=r+1}^{r+s} \sum_{k=1}^{K} \psi_{lk}(t^*)^k + R_6^* \end{aligned} \tag{42}$$

where $\delta_l^* = 1 + \sum_{l=r+1}^{r+s} \gamma_l \delta_l \tilde{c}_0$ and $\psi_{lk} = \gamma_l \delta_l \tilde{c}_k$, $l = r+1,...,r+s$, $k = 1,...,K$. It is convenient to reparameterize (41) as follows:

$$f_t = \sum_{l=r+1}^{r+s} \sum_{k=1}^{K} \psi_{lk}(t^*)^k + R_6^* \tag{43}$$

Under the null hypothesis, the remainder R_6^* vanishes. It seems that the coefficients ψ_{lk}, $l = r+1,...,s$, for a fixed k, are not identified because they are all related to the same variable $(t^*)^k$. They have to be merged, which leads to

$$f_t = \sum_{k=1}^{K} \psi_k^*(t^*)^k + R_6^*.$$

In other words, the test statistic is the same, independent of whether we would be testing against including G_{r+1} or including $G_{r+1},...,G_{r+s}$, $s \geq 2$. Compare this with Corollary 5, which is a special case. In fact, Corollary 5 contains another example of a misspecification test of the multiplicative model in which the misspecification is of the type $h_t(g_t + f_t)$.

6.1.3 *Multiplicative misspecification*

Under multiplicative misspecification, the parametric alternative to the TV-GARCH model is formulated as

$$\varepsilon_t = \zeta_t (h_t g_t f_t)^{1/2}. \tag{44}$$

In this framework, $\mathrm{H}_0 : f_t \equiv 1$, which is equivalent to $\boldsymbol{\theta}_3 = \mathbf{0}$. Under the null hypothesis, the model reduces to the multiplicative TV-GARCH model. For this specification, the log-likelihood function for observation t may be written

$$\ell_t = -\frac{1}{2}\ln 2\pi - \frac{1}{2}(\ln h_t + \ln g_t + \ln f_t) - \frac{\varepsilon_t^2}{2h_t g_t f_t}.$$

The additional block of the score has the form

$$\frac{\partial \ell_t}{\partial \boldsymbol{\theta}_3} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t g_t f_t} - 1\right)\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}$$

which, under H_0, reduces to

$$\left.\frac{\partial \ell_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t g_t} - 1\right)\left.\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0}.$$

Taking (44) as the alternative, the vector $\hat{\mathbf{x}}_{1t}$ is given in (32) and

$$\hat{\mathbf{x}}_{2t} = \left.\frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0} = \hat{\mathbf{v}}_t. \tag{45}$$

This category includes general misspecification tests of adequacy of the estimated specification. After the estimation of the TV-GARCH model, one may want to check whether the estimated standardized errors still contain some structure. In the GARCH context, Lundbergh and Teräsvirta (2002) proposed a Lagrange multiplier statistic for testing the hypothesis of no remaining ARCH which is asymptotically equivalent to the portmanteau statistic introduced by Li and Mak (1994). A similar test statistic can be obtained for the multiplicative TV-GARCH model.

Testing the hypothesis of no remaining ARCH

An important misspecification test for the multiplicative TV-GARCH specification is the so-called 'ARCH-in-GARCH' test. The original model

$$\varepsilon_t = \zeta_t h_t^{1/2} g_t^{1/2}, \qquad \zeta_t \sim \text{nid}(0,1)$$

is extended by assuming that, under the alternative, $\zeta_t = \xi_t f_t^{1/2}$, where $\xi_t \sim \text{nid}(0,1)$, and

$$f_t = 1 + \sum_{j=1}^{s} \phi_j \zeta_{t-j}^2. \tag{46}$$

The hypothesis of interest is $\text{H}_0 : \phi_1 = ... = \phi_s = 0$ and $\frac{\partial \widehat{f}_t}{\partial \boldsymbol{\theta}_3}|_{H_0} = (\widehat{\zeta}_1^2, ..., \widehat{\zeta}_s^2)'$. Some special cases may be mentioned. If $g_t \equiv 1$, the test collapses into the test of 'no ARCH-in-GARCH' in Lundbergh and Teräsvirta (2002). If $h_t \equiv 1$ as well, the test coincides with the Engle's test of no ARCH. Setting only $h_t \equiv 1$, it reduces to the test of no ARCH in $\varepsilon_t / \widehat{g}_t^{1/2}$. The test is presented in the next corollary.

Corollary 10 *Consider the model (44) where $\{\zeta_t\}$ is a sequence of independent standard normal variables. Let $\boldsymbol{\theta}_1 = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)'$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\delta}', \gamma_1, ..., \gamma_r, \mathbf{c}_1, ..., \mathbf{c}_r)'$ with $\boldsymbol{\delta} = (\delta_1, ..., \delta_r)'$. Furthermore, f_t is defined by (46) such that $\boldsymbol{\theta}_3 = (\phi_1, ..., \phi_s)'$ and $\hat{\mathbf{v}}_t = (\widehat{\zeta}_1^2, ..., \widehat{\zeta}_s^2)'$. Assume that the maximum likelihood estimators of the parameters of (44) are asymptotically normal when $H_0 : \boldsymbol{\theta}_3 = \mathbf{0}$ is valid. Thus, under this null hypothesis, the LM statistic (15), with $\hat{u}_t = \hat{\varepsilon}_t^2 / \hat{h}_t^0 \hat{g}_t^0 - 1$, $\hat{\mathbf{x}}_{1t}$ as in (32) and $\hat{\mathbf{x}}_{2t} = \hat{\mathbf{v}}_t$ is asymptotically χ^2-distributed with s degrees of freedom.*

6.2 Misspecification tests for the additive model

In this section we shall consider the additive TV-GARCH model and assume that it is either additively or multiplicatively misspecified. The former possibility may include, for example, tests against remaining nonlinearity and additional transitions, whereas the test of the adequacy of the estimated model belongs to the latter one. To this end, let $h_t = h_t(\boldsymbol{\theta}_1)$ and $g_t = g_t(\boldsymbol{\theta}_2)$, such that $\boldsymbol{\theta}_i$, $i = 1, 2$, represent the parameters belonging to $h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$ and $g_t = \sum_{l=1}^{r} (\alpha_{0l} + \sum_{i=1}^{q} \alpha_{il} \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{jl} h_{t-j}) G_l(t^*; \gamma_l, \mathbf{c}_l)$. Under the null hypothesis of no misspecification, the extended model reduces to the additive TV-GARCH parameterization.

6.2.1 *Additive misspecification*

In order to define the set of alternative models for this class, consider a general alternative written as

$$\varepsilon_t = \zeta_t (h_t + g_t + f_t)^{1/2}. \tag{47}$$

Under the null hypothesis, $f_t \equiv 0$. If $g_t \equiv 0$, the test coincides to the additive test developed in Lundbergh and Teräsvirta (2002). In the case of the additive parameterization, the diagnostic tests mentioned in Sections 6.1.1 and 6.1.2 belong to the class (47). Such tests can be easily adapted into the present context, where the quantities $\hat{u}_t$, $\hat{\mathbf{x}}_{it}$, $i = 1, 2$, and $\hat{\mathbf{v}}_t$ have to be modified accordingly. We shall therefore be concerned with a general alternative hypothesis rather than describing individual situations.

The log-likelihood function for observation t is

$$\ell_t = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\{\ln(h_t + g_t + f_t)\} - \frac{\varepsilon_t^2}{2(h_t + g_t + f_t)}$$

and the vector of the first partial derivatives with respect to $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ under H_0 equals

$$\left.\frac{\partial \ell_t}{\partial \boldsymbol{\theta}}\right|_{H_0} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t + g_t} - 1\right)\mathbf{x}_{1t}$$

where $\mathbf{x}_{1t} = \left(\frac{1}{h_t+g_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}_1'}, \frac{1}{h_t+g_t}\frac{\partial g_t}{\partial \boldsymbol{\theta}_2'}\right)'$. The appropriate estimates of $\frac{\partial h_t}{\partial \boldsymbol{\theta}_1}|_{H_0}$ and $\frac{\partial g_t}{\partial \boldsymbol{\theta}_2}|_{H_0}$ are

$$\left.\frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_1}\right|_{H_0} = \hat{\mathbf{z}}_t + \sum_{j=1}^{p}\hat{\beta}_j \left.\frac{\partial \hat{h}_{t-j}}{\partial \boldsymbol{\theta}_1}\right|_{H_0} \tag{48}$$

$$\left.\frac{\partial \hat{g}_t}{\partial \boldsymbol{\theta}_2}\right|_{H_0} = \sum_{l=1}^{r}\hat{\mathbf{z}}_t \hat{G}_l(t^*) + \sum_{l=1}^{r}\hat{\boldsymbol{\theta}}_{2l}'\hat{\mathbf{z}}_t \frac{\partial \hat{G}_l(t^*)}{\partial \boldsymbol{\theta}_2} + \sum_{j=1}^{p}\sum_{l=1}^{r}\hat{\beta}_{jl}\hat{G}_l(t^*)\left.\frac{\partial \hat{g}_{t-j}}{\partial \boldsymbol{\theta}_2}\right|_{H_0} \tag{49}$$

where $\mathbf{z}_t = (1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, h_{t-1}, ..., h_{t-p})'$, $\boldsymbol{\theta}_{2l} = (\alpha_{0l}, \alpha_{1l}, ..., \alpha_{ql}, \beta_{1l}, ..., \beta_{pl})'$, $l = 1, ..., r$, and $G_l(t^*) \equiv G_l(t^*; \gamma_l, \mathbf{c}_l)$. The additional block of the score for observation t, under H_0, equals

$$\left.\frac{\partial \ell_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t + g_t} - 1\right)\frac{1}{h_t + g_t}\left.\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0}$$

where $\frac{\partial f_t}{\partial \boldsymbol{\theta}_3} = \mathbf{v}_t$. To define the LM statistic, set

$$\hat{\mathbf{x}}_{1t} = \left(\frac{1}{\hat{h}_t^0 + \hat{g}_t^0}\left.\frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_1'}\right|_{H_0}, \frac{1}{\hat{h}_t^0 + \hat{g}_t^0}\left.\frac{\partial \hat{g}_t}{\partial \boldsymbol{\theta}_2'}\right|_{H_0}\right)' \tag{50}$$

$$\hat{\mathbf{x}}_{2t} = \frac{1}{\hat{h}_t^0 + \hat{g}_t^0}\left.\frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0} = \frac{\hat{\mathbf{v}}_t}{\hat{h}_t^0 + \hat{g}_t^0} \tag{51}$$

where $\frac{\partial \hat{h}_t}{\partial \boldsymbol{\theta}_1}|_{H_0}$ and $\frac{\partial \hat{g}_t}{\partial \boldsymbol{\theta}_2}|_{H_0}$ are given in (48) and (49), respectively. These results apply to the test against remaining nonlinearity. The test will be presented in the following corollary.

Corollary 11 *Consider the model (47) where* $\{\zeta_t\}$ *is a sequence of independent standard normal variables. Let* $\boldsymbol{\theta}_1 = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)'$, $\boldsymbol{\theta}_2 = (\boldsymbol{\delta}', \gamma_1, ..., \gamma_r, \mathbf{c}_1, ..., \mathbf{c}_r)'$ *with* $\boldsymbol{\delta} = (\delta_1, ..., \delta_r)'$ *and* $\boldsymbol{\theta}_3 = (\boldsymbol{\varpi}_i', \boldsymbol{\pi}_i')'$, *where* $\boldsymbol{\varpi}_i = (\varpi_{i1}, ..., \varpi_{iK})'$ *and* $\boldsymbol{\pi}_i = (\pi_{i1}, ..., \pi_{iK})'$, $i = 1, ..., q$. *Assume that the maximum likelihood estimators of the parameters of (47) are asymptotically normal when* $H_0 : \boldsymbol{\theta}_3 = \mathbf{0}$ *is valid. Thus, under this null hypothesis, the LM statistic (15), with* $\hat{u}_t = \hat{\varepsilon}_t^2/(\hat{h}_t^0 + \hat{g}_t^0) - 1$, $\hat{\mathbf{x}}_{1t}$ *as in (50) and* $\hat{\mathbf{x}}_{2t}$ *as in (51) with* $\hat{\mathbf{v}}_t = (\hat{\mathbf{v}}_{1,t}', ..., \hat{\mathbf{v}}_{K+2,t}')'$ *where* $\mathbf{v}_{it} = (\varepsilon_{t-1}^i, ..., \varepsilon_{t-q}^i)'$, $i = 1, ..., K+2$, *is asymptotically* χ^2-*distributed with* $\dim(\boldsymbol{\theta}_3)$ *degrees of freedom.*

6.2.2 *Multiplicative misspecification*

Consider the following extended TV-GARCH model

$$\varepsilon_t = \zeta_t (h_t + g_t)^{1/2} f_t^{1/2}. \tag{52}$$

Under the null hypothesis, $f_t \equiv 1$. This category entails the test for assessing the adequacy of the functional form of the estimated model. This test was already discussed when the TV-GARCH model was in the multiplicative form and the same considerations apply here.

The log-likelihood function for a single observation on (52) is

$$\ell_t = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \{\ln(h_t + g_t) + \ln f_t)\} - \frac{\varepsilon_t^2}{2(h_t + g_t) f_t}$$

and the relevant block of the score due to $\boldsymbol{\theta}_3$, under H_0, has the form

$$\left.\frac{\partial \ell_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0} = \frac{1}{2} \left(\frac{\varepsilon_t^2}{h_t + g_t} - 1 \right) \left.\frac{\partial f_t}{\partial \boldsymbol{\theta}_3}\right|_{H_0}.$$

The hypothesis of interest is that the squared standardized error sequence is iid. Under the alternative, f_t is defined in (46). In this framework, the vector $\hat{\mathbf{x}}_{1t}$ is given as in (50) and $\hat{\mathbf{x}}_{2t} = \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}_3}|_{H_0} = \hat{\mathbf{v}}_t$. The following Corollary defines the test statistic.

Corollary 12 *Consider the model (52) where* $\{\zeta_t\}$ *is a sequence of independent standard normal variables. Let* $\boldsymbol{\theta}_1 = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)'$ *and* $\boldsymbol{\theta}_2 = (\boldsymbol{\delta}', \gamma_1, ..., \gamma_r, \mathbf{c}_1, ..., \mathbf{c}_r)'$ *with* $\boldsymbol{\delta} = (\delta_1, ..., \delta_r)'$. *Furthermore,* f_t *is defined by (46) such that* $\boldsymbol{\theta}_3 = (\phi_1, ..., \phi_s)'$ *and* $\hat{\mathbf{v}}_t = (\widehat{\zeta_1^2}, ..., \widehat{\zeta_s^2})'$. *Assume that the maximum likelihood estimators of the parameters of (52) are asymptotically normal when* $H_0 : \boldsymbol{\theta}_3 = \mathbf{0}$ *is valid. Thus, under this null hypothesis, the LM statistic (15), with* $\hat{u}_t = \hat{\varepsilon}_t^2/(\hat{h}_t^0 + \hat{g}_t^0) - 1$, $\hat{\mathbf{x}}_{1t}$ *as in (50) and* $\hat{\mathbf{x}}_{2t} = \hat{\mathbf{v}}_t$ *is asymptotically* χ^2-*distributed with* s *degrees of freedom.*

7 Simulation study

7.1 Monte Carlo design

In this section, we conduct a small simulation experiment to evaluate the finite-sample properties of the proposed parameter constancy tests. These are the tests against an additive and a multiplicative TV-GARCH specifications. Specifically, we shall investigate the size and power properties of the LM-type tests involved in the modelling strategies as well as the success rate of the specification procedures. Sample lengths of 1000, 2500 and 5000 observations have been used in all simulations. For each design, the total number of replications equals 2000. 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 3.30 (see Doornik (2002)). The behaviour of the test statistics is examined for several data generating processes (DGP's) that can be nested in the following TV-GARCH specification:

$$\begin{aligned} y_t &= \varepsilon_t, \qquad \varepsilon_t|\mathcal{F}_{t-1} \sim N(0, h_t) \\ h_t &= \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1} + (\alpha_{01} + \alpha_{11}\varepsilon_{t-1}^2 + \beta_{11}h_{t-1})G_1(t^*; \gamma_1, c_1). \end{aligned} \tag{53}$$

The data generating processes are as following:

DGP (i) $\quad h_t = 0.10 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$

$\alpha_1 = \{0.05, 0.09, 0.10\}$ and $\beta_1 = \{0.80, 0.85, 0.90\}$

DGP (ii) $\quad h_t = 0.10 + \alpha_{01}G_1(t^*; \gamma_1, c_1) + 0.10\varepsilon_{t-1}^2 + 0.80h_{t-1}$

$\alpha_{01} = \{0.10, 0.30\}$

DGP (iii) $\quad h_t = 0.10 + (0.10 + \alpha_{11}G_1(t^*; \gamma_1, c_1))\varepsilon_{t-1}^2 + 0.80h_{t-1}$

$\alpha_{11} = \{0.05, 0.09\}$

DGP (iv) $\quad h_t = (0.10 + \alpha_{01}G_1(t^*; \gamma_1, c_1)) + (0.10 + \alpha_{11}G_1(t^*; \gamma_1, c_1))\varepsilon_{t-1}^2 + 0.80h_{t-1}$

$\alpha_{01} = \{0.10, 0.30\}$ and $\alpha_{11} = \{0.05, 0.09\}$

DGP (v) $\quad h_t = 0.10 + 0.10\varepsilon_{t-1}^2 + (0.80 + \beta_{11}G_1(t^*; \gamma_1, c_1))h_{t-1}$

$\beta_{11} = \{0.05, 0.09\}$

DGP (vi) $\quad h_t = 0.10 + \alpha_{01}G_1(t^*; \gamma_1, c_1) + 0.10\varepsilon_{t-1}^2 + (0.80 + \beta_{11}G_1(t^*; \gamma_1, c_1))h_{t-1}$

$\alpha_{01} = \{0.10, 0.30\}$ and $\beta_{11} = \{0.05, 0.09\}$

DGP (vii) $\quad h_t = (0.10 + 0.10\varepsilon_{t-1}^2 + 0.85h_{t-1})(1 + \delta_1 G_1(t^*; \gamma_1, c_1))$

$\delta_1 = \{0.05, 0.08\}$

The first six designs concern the additive TV-GARCH model, whereas the remaining one relates to the multiplicative model. In all these seven experiments, 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 $t \leq 0$.

7.2 Finite sample properties

In this section we shall look at the small-sample properties of the modelling strategy for the TV-GARCH model. We first report results on the size and power properties of our parameter constancy tests. Then we turn to the specification of TV-GARCH models.

Size and power simulations

The size and the power results of the tests are presented in graphs following the recommendation by Davidson and MacKinnon (1998). Both the ordinary and the robustified versions of each test are computed using auxiliary regressions. Results of the size simulations appear in the form of p-value discrepancy plots in Figure 5. In these graphs, the difference between the empirical size and the nominal size is plotted against the nominal size. The upper panel of Figure 5 presents the results for the size simulations for the test against an additive alternative, whereas the bottom panel shows the empirical size results of the test against a multiplicative alternative. For each test we calculate the actual rejection frequencies for the three sample sizes at the following nominal levels: 0.1%, 0.3%, 0.5%, 0.7%, 0.9%, 1%, ..., 10%. The series are generated from the GARCH model given by the DGP (i) where $\alpha_0 = 0.10$, $\alpha_1 = 0.10$ and $\beta_1 = 0.85$.

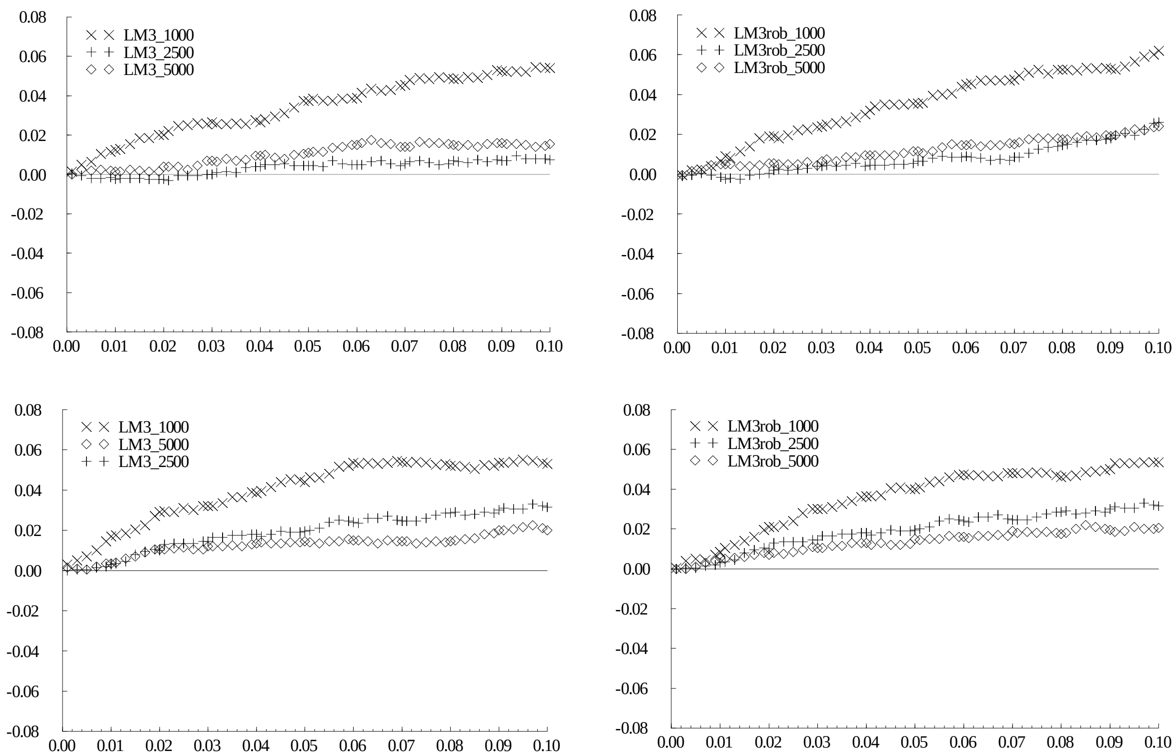


Figure 5. Size discrepancy plots of the additive (upper panel) and multiplicative (lower panel) parameter constancy tests. Both the ordinary (left) and the robust (right) versions of the tests are plotted.

Both tests are somewhat size-distorted at the sample size $T = 1000$, but the results become 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, the size distortions in the robust version of the tests do not differ too much from those in the non-robust test. 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 $T \geq 2500$. Employing a robust test even when the errors are normal does not seem to lead to a large loss of power.

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.

Because of this, in Figure 6 we only report power results for our tests. In these graphs the rejection frequencies are plotted against the nominal significance levels 0.1%, 0.3%, 0.5%, 0.7%, 0.9%, 1%, ..., 10%. Instead of the size-adjusted power-size curves suggested by Davidson and MacKinnon (1998), we simply report power curves as the tests have good size properties.

The power results in Figure 6 have been obtained by generating artificial data from the DGP (ii) where the coefficient $\alpha_{01} = 0.10$, the slope parameter $\gamma_1 = 5$ and the location parameter $c_1 = 0.5$. The rejection frequencies of the additive LM test statistics shown in the top panel are moderate when $T = 1000$ and increase with the sample size. The pattern of the power results for the robustified version of the test is very similar to the non-robust one.

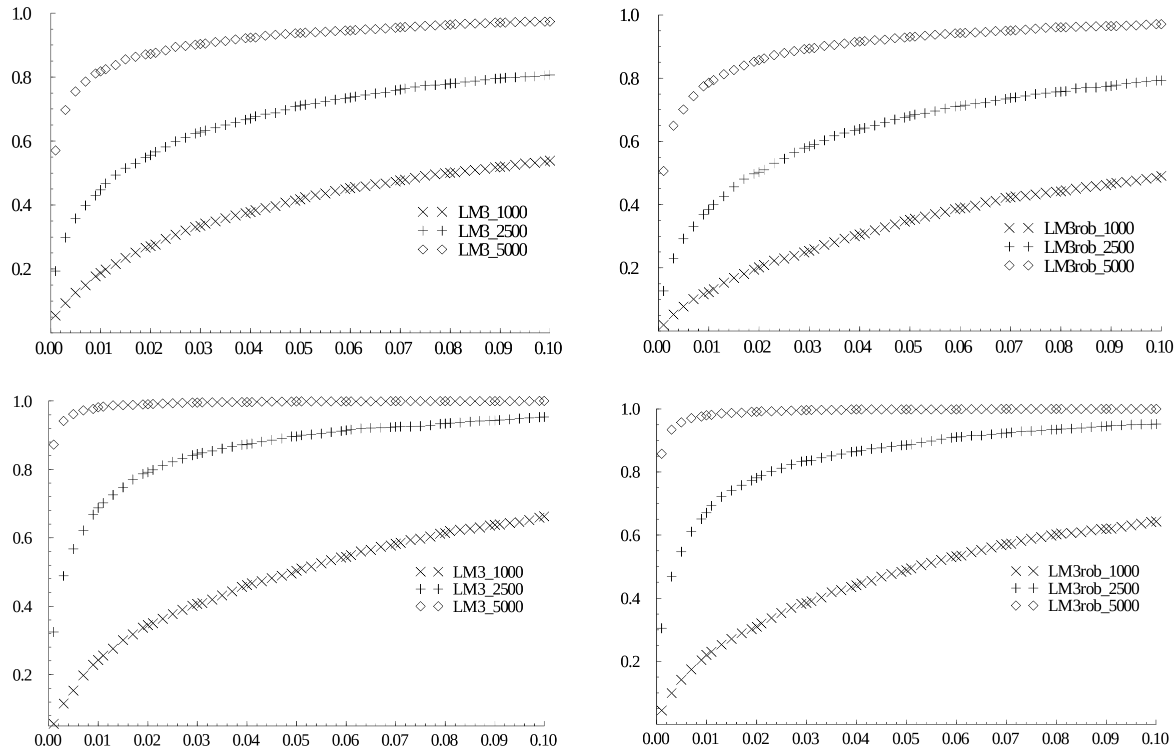


Figure 6. Power curves of the additive (upper panel) and multiplicative (lower panel) parameter constancy tests. Both the ordinary (left) and robust (right) versions of the tests are plotted.

Rejection frequencies for the LM-type test against a multiplicative alternative are shown in the lower panel of Figure 6. The results refer to power simulations when the data generating process is a multiplicative TV-GARCH model (DGP vii). The coefficient $\delta_1 = 0.05$ and $\gamma_1 = 5$ as before. As expected, the rejection frequencies are an increasing function of the sample size and of the parameter δ_1 (as well as of the parameter α_{01} in the additive case). Moreover, the LM-type test statistic turns out to be very powerful even for short time series. Again, the behaviour of the robust version of the test in the power simulations is quite similar to that of the non-robust version.

Simulating the model selection strategy

In this section we consider the performance of the specific-to-general specification strategy for TV-GARCH models with an additive time-varying structure. This is done by studying the selection frequencies of various models. The specification procedure has been discussed in Section 4.1. A total of 2000 replications are carried out for each DGP and all three sample sizes. The first 1000 observations of each generated series are discarded to avoid the initialization effects. Throughout, we set $\alpha = 0.05$ for both the LM_1 and LM_3 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.

Results for DGP (i) are reported in Table 1 (see Appendix B). The frequencies of the correct number of transitions are shown in boldface. The column labelled 'choice' 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 can be found in Table 2. We report separately an additive time-varying structure in each parameter of the GARCH model when $c_1 = 0.50$. This corresponds to the DGP's (ii), (iii) and (v). For all the cases, the parameters of the linear GARCH are $\alpha_0 = 0.10$, $\alpha_1 = 0.10$ and $\beta_1 = 0.80$. Clearly, the constant-parameter GARCH model is chosen too often for parameterizations with smoothest changes and shortest series. For large sample sizes, the selection frequencies of the true model become quite high even for very smooth changes. Again, the LM_1-test has higher power than LM_3. As expected, the correct model is selected more frequently for high than for low values of α_{01}, α_{11} or β_{11}. Moreover, the correct model is selected slightly more often when the change only occurs either in the constant α_0 or in the GARCH parameter β_1 than when it does in the ARCH parameter α_1.

The model selection frequencies when the series are generated from DGP (iv) are given in Table 3. The correct model is chosen more frequently when the change in α_{01} and α_{11} becomes large. It also becomes easier to identify a single transition when the slope parameter γ increases. Again, the results concern the case when the change occurs in the middle of the sample. Finally, Table 4 contains the frequencies of the selected models for the DGP (vi). In this case, the power of our procedure turns out to be very similar to that shown in Table 3. This may be explained by the fact that either changes in α_{01} and α_{11} or the ones in α_{01} and β_{11} simultaneously change the amplitude of clusters as well as the unconditional variance. We also carried out simulations for the DGP (vii) which are not reported in the paper. The results are almost identical to what is reported for the additive TV-GARCH model. Overall, the sequential procedure seems to work relatively well for all combinations of parameters considered and for sample sizes $T \geq 1000$.

8 Applications

In this section we shall present two empirical examples involving two financial time series, a stock index and an exchange rate return series. The former is the Standard and Poor 500 composite index (S&P 500) and the latter 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 the continuously compounded rates of return.

8.1 Stock index returns

The daily S&P 500 return series was provided by 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 7. It contains periods of large volatility both in the beginning and at the end of the sample period, whereas the average volatility in the middle of the sample is somewhat lower than in both ends.

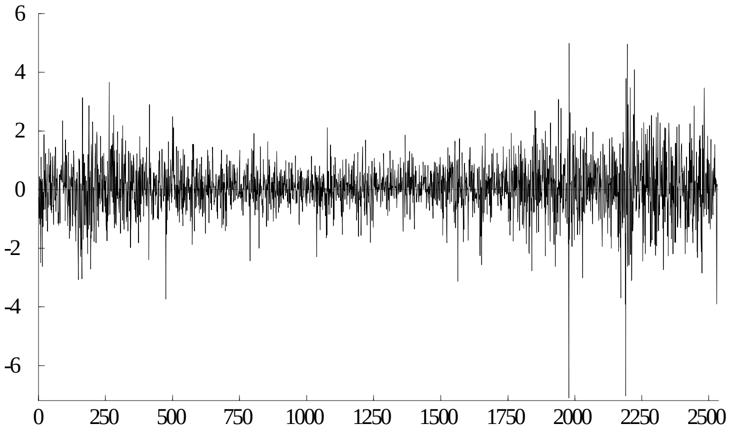


Figure 7. Daily returns of the S&P 500 composite index from January 2, 1990 until December 31, 1999 (2531 observations).

Summary statistics for the series can be found in the second column of Table 5. 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 (2007)) 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.

We first estimate a standard GARCH(1,1) model to this series. In order to save space, the results are not shown here. Results of the parameter constancy test against an additive time-varying structure are reported in Table 7. The test of parameter constancy against an additive TV-GARCH model, when several parameters are assumed to change under the alternative, rejects the null hypothesis. The tests against alternatives in which some parameters remain constant, suggest that the the intercept may be the main source of nonconstancy.

Instead of specifying and estimating an additive TV-GARCH model with a time-varying intercept, we test the iid hypothesis of our stochastic sequence $\{\varepsilon_t\}$ against deterministic change. This is Step 1 in the specification of multiplicative TV-GARCH models outlined in Section 4.2. The results can be found in Table 8. The null hypothesis is rejected very strongly as the p-value of the test equals 3×10^{-23}. The test sequence for specifying the structure of the deterministic function g_t points towards $K = 2$. Fitting the TV-GARCH model with a single transition function and $K = 2$ to the series and testing for another transition still leads to rejecting the null hypothesis. The p-value, however, is now considerably larger, equalling 0.0028, and the specification test sequence now clearly suggests $K = 1$. Accepting this outcome, fitting the corresponding TV-GARCH model to the series and testing for yet another transition yields the p-value 0.0623. If the null hypothesis is tested directly against a standard logistic transition function, the p-value equals 0.0197. Given the relatively large number of observations, this is not a small value, and the model with two transitions is tentatively accepted as the final model.

In this model, the estimate of g_t has the following form:

$$\widehat{g}_t = \{1 + \underset{(0.4265)}{1.7041} G_1(t^*; \widehat{\gamma}_1, \widehat{\mathbf{c}}_1) + \underset{(0.5455)}{1.7335} G_2(t^*; \widehat{\gamma}_2, \widehat{c}_2)\} \tag{54}$$

with

$$G_1(t^*; \widehat{\gamma}_1, \widehat{\mathbf{c}}_1) = (1 + \exp\{-\underset{(-)}{100}(t^* - \underset{(0.0100)}{0.1643})(t^* - \underset{(0.0831)}{0.6950})\})^{-1} \tag{55}$$

and

$$G_2(t^*; \widehat{\gamma}_2, \widehat{c}_2) = (1 + \exp\{-\underset{(-)}{100}(t^* - \underset{(0.0043)}{0.8534})\})^{-1}. \tag{56}$$

The graph of the deterministic component $\widehat{g}_t$ is depicted in Figure 9. The two transitions are clearly visible and illustrate how volatility first decreases and then increases over time. A GARCH model is fitted to the standardized residuals $\varepsilon_t/\widehat{g}_t^{1/2}$, and the estimated model is subjected to misspecification tests described in Section 6. Table 9 contains the test results. The hypothesis of 'no ARCH in GARCH' is not rejected for any lag length considered. As may be expected, the hypothesis of no additional transitions is not rejected either. There is, however, some indication of nonlinearity in the conditional variance as the GARCH(1,1) component is strongly rejected against a STGARCH(1,1) one for $K = 1$. In order to remedy this problem, we specify a GJR-GARCH(1,1) model for h_t.

The parameter estimates of the GJR-GARCH model can be found in Table 6. It is seen that the persistence factor equals $\widehat{\alpha}_1 + \widehat{\beta}_1 + \widehat{\gamma}_1/2 = 0.993$, so that the estimated model is practically an integrated GJR-GARCH model. For illustration, Table 6 also contains the parameter estimates at the point where the parameters in h_t have been estimated for the first time. It is seen that there is already a large change in the value of the log-likelihood compared to the maximum found for the GJR-GARCH(1,1) model. The persistence, however, has not yet decreased very much. Figure 10 contains the autocorrelations of $|\varepsilon_t|$ (Panel (a)) and those of $|\varepsilon_t|/\widehat{g}_t^{1/2}$ after a single iteration (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. At the same time, the autocorrelations retain the 'long-memory property', the very slow decay as a function of the lag, that is obvious in the autocorrelations of $|\varepsilon_t|$.

The log-likelihood considerably increases with further iterations, and the final persistence indicator has the remarkably low value $\widehat{\alpha}_1 + \widehat{\beta}_1 + \widehat{\gamma}_1/2 = 0.918$,. A clear trade-off is observed here. When it is assumed that the process is stationary there is only one level (unconditional variance) to which the conditional variance converges when it is assumed that $z_t = 0$ for $t > t_0$. This convergence then takes a very long time ($\widehat{\alpha}_1 + \widehat{\beta}_1 + \widehat{\gamma}_1/2 = 0.993$ is very close to unity). In the TV-GJR-GARCH model this level is time-varying, and the rate of convergence to a particular level can thus be much more rapid than it is in the standard GJR-GARCH model. Panel (c) of Figure 10 now shows that the autocorrelations of $|\varepsilon_t|/\widehat{g}_t^{1/2}$ have decreased even further, and only few of them exceed two standard deviations of $|\varepsilon_t|$ under the iid normality assumption, marked by the straight 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 surprisingly little remains to be explained by the traditional GJR-GARCH component.

Table 10 contains the misspecification test results for this model. Even if the GJR-GARCH model is a rather crude representation of asymmetry compared to the smooth transition GARCH specification, it manages to capture most of the asymmetry. The p-value of the test of no additional nonlinearity, when applied to the TV-GJR-GARCH model, equals 0.035, which is much larger than 1×10^{-10} obtained when the test was applied to the estimated TV-GARCH(1,1) model. Applying the 1% significance level, the other misspecification tests do not reject the model either, and the TV-GJR-GARCH model is thus accepted to be our final model.

Figure 11 that contains the estimated conditional standard deviations $h_t^{1/2}$ of $\{\varepsilon_t\}$ for the GJR-GARCH(1,1) model and the ones of $\{\varepsilon_t/\widehat{g}_t^{1/2}\}$ illustrates the situation as well. For the GJR-GARCH model, see Panel (a), the graph looks rather 'nonstationary'. Some nonstationarity remains after a single iteration, as the autocorrelations of $\{\varepsilon_t/\widehat{g}_t^{1/2}\}$ in Panel (b) also demonstrate. From the graph in Panel (c) (the final model) it is seen that volatility is still changing over time, but there no longer seem to be persistent level changes. They have been absorbed by the deterministic component.

Column 4 in Table 5 contains the skewness and kurtosis estimates for $\varepsilon_t/\widehat{g}_t^{1/2}$. The negative skewness remains 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 (2.2) than the original number (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 5 shows some decrease, but because its nonrobust value was already small, the decrease has remained rather modest.

In Figure 12, 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. Its estimate was originally slightly negative but statistically insignificant, and the model was re-estimated after restricting α_1 to zero. 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.

8.2 Exchange rate data

The data of this section consist of daily returns of the spot SPD/USD exchange rate provided by the Federal Reserve Bank of New York. The time series is shown in Figure 8. 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.

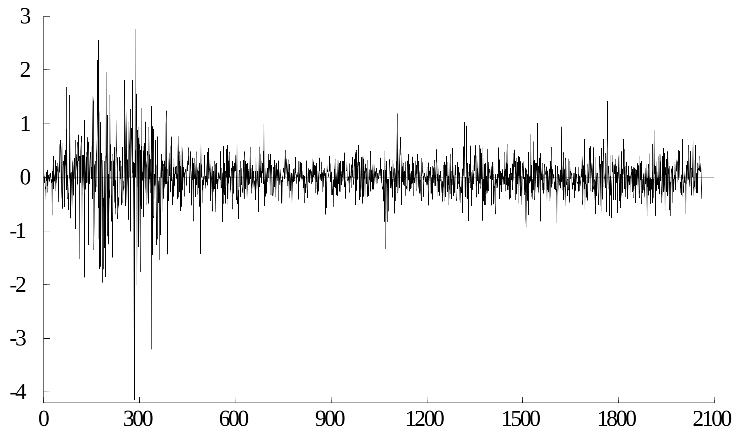


Figure 8. Daily returns of the Singapore Dollar versus US dollar exchange rate from May 1, 1997 until July 11, 2005 (2060 observations).

Descriptive statistics for the SPD/USD exchange rate returns are reported in Table 5. 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, as can be expected. 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.

Parameter constancy of the GARCH(1,1) model is rejected against an additive TV-GARCH model. These test results are presented in Table 7. In this case, however, the rejection is not due to the intercept but rather to the other two parameters. As in the previous application, we shall not fit any additive TV-GARCH models to our return series but choose to work with the multiplicative model. The test of constant unconditional variance against a time-varying one has the p-value equal to 1×10^{-20}. Table 8 contains the outcomes of the sequence of specification tests. The results indicate that one should choose $K = 1$, that is, have a monotonically increasing transition function. A multiplicative TV-GARCH model with a single transition appears adequate in the sense that the test for another transition has $p = 0.14$. The diagnostic tests of this model in Table 9 do not reject the model. There is no remaining ARCH in the standardized errors, no evidence of higher-order structure in the GARCH component, and nothing suggests the existence of additional transitions. Finally, the linearity test against the smooth transition GARCH does not indicate remaining nonlinearity. Judging from these statistics, the model seems to be adequately specified. It is thus tentatively accepted as our final model for the SPD/USD daily return series.

The final estimates for the function g_t are as follows:

$$\widehat{g}_t = \{1 - \underset{(0.0074)}{0.7890} G_1(t^*; \widehat{\gamma}_1, \widehat{c}_1)\}, \tag{57}$$

where

$$G_1(t^*; \widehat{\gamma}_1, \widehat{c}_1) = (1 + \exp\{-\underset{(-)}{100}(t^* - \underset{(0.0014)}{0.2101})\})^{-1} \tag{58}$$

The graph of the transition function can be found in Figure 13. Figure 8 already shows that the volatility is high in the beginning and settles down to a lower level after about 500 observations (two years). From Table 5 it is seen that the excess kurtosis has decreased substantially from its value for $\{\varepsilon_t\}$ and, furthermore, that the skewness has been reduced from -0.9 to less than -0.3. This large reduction can be ascribed to the fact that the original skewness was due to a couple of very large negative returns during the Asian crisis. Their significance has subsequently been reduced in $\{\varepsilon_t/\widehat{g}_t^{1/2}\}$ where the conditional heteroskedasticity component has been standardized by the underlying nonstationary volatility component. Besides, according to the robust estimates the skewness has not been affected, which is in line with this conclusion as well.

The parameter estimates of the model appear in Table 6. It can be seen that even for the exchange rate series, the first iteration already has a large effect on the value of the log-likelihood. Figure 14 shows that at that stage, the autocorrelations of $|\varepsilon_t|/\widehat{g}_t^{1/2}$ are considerably lower than those of $|\varepsilon_t|$, although their decay as a function of the lag length is still slow. The final estimates indicate more persistence than there is in the S&P 500 case, but the decrease is still large compared to the GARCH(1,1) model. The decay rate of the autocorrelations of $|\varepsilon_t|/\widehat{g}_t^{1/2}$ in Figure 14 is quite rapid and looks more or less exponential. The first-order autocorrelation that was about 0.304 for $|\varepsilon_t|$ equals 0.121 for $|\varepsilon_t|/\widehat{g}_t^{1/2}$. The graph of the conditional variance h_t in Panel (a) of Figure 15 clearly shows the period of high volatility, which is the cause of the high persistence suggested by the GARCH(1,1) model. Panel (c) shows that in the final model this high-volatility period is explained by the deterministic component g_t, and that the graph of h_t does not show signs of nonstationarity. This is precisely what one would expect after a look at the parameter estimates in Table 6.

Figure 16 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. An additional result, similarly to the previous application, is the ability of the time-varying news impact curves to distinguish different reaction levels of volatility to news in calm and turbulent times. In general, the impact of news on volatility tends to be high in expansions and low in recessions.

9 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. 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 TV-GARCH specifications. 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 reasonable good 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 model to these series yields results that are quite different from the ones obtained by our approach and suggest 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 baseline volatility or 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 could 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

ANDERSON, H. M., K. NAM, AND F. VAHID (1999): "Asymmetric Nonlinear Smooth Transition GARCH Models," in *Nonlinear Time Series Analysis of Economic and Financial Data*, ed. by P. Rothman, pp. 191–207. Boston: Kluwer.

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 (2007): "Modeling Long Memory and Structural Breaks in Conditional Variances: An Adaptive FIGARCH Approach," Working Paper No. 593, Queen Mary University of London, Department of Economics.

BERKES, I., E. GOMBAY, L. HORVÁTH, AND P. KOKOSZKA (2004): "Sequential Change-Point Detection in GARCH(p,q) Models," *Econometric Theory*, 20, 1140–1167.

BERKES, I., L. HORVÁTH, AND P. KOKOSZKA (2003): "GARCH Processes: Structure and Estimation," *Bernoulli*, 9, 201–207.

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.

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

DAVIES, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 64, 247–254.

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. (2002): *Object-Oriented Matrix Programming Using Ox*. London: Timberlake Consultants Press and Oxford: www.doornik.com, 3rd edn.

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 United Kingdom Inflation," *Econometrica*, 50, 987–1007.

ENGLE, R. F., AND J. GONZALO RANGEL (2005): "The Spline GARCH Model for Unconditional Volatility and its Global Macroeconomics Causes," *Mimeo, NYU and UCSD.*

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.

Geweke, J., and S. Porter-Hudak (1983): "The Estimation and Application of Long-Memory Time-Series Models," *Journal of Time Series Analysis*, 4, 221–238.

González-Rivera, G. (1998): "Smooth Transition GARCH Models," *Studies in Nonlinear Dynamics and Econometrics*, 3, 61–78.

Granger, C. W. J. (1993): "Strategies for Modelling Nonlinear Time-Series Relationships," *Economic Record*, 69, 233–239.

Hagerud, G. E. (1997): "A New Non-Linear GARCH Model," Ph.D. thesis, EFI, Stockholm School of Economics.

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

Hansen, B. E. (1996): "Inference When a Nuisance Parameter is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

Jensen, S. T., and A. Rahbek (2004): "Asymptotic Inference for Nonstationary GARCH," *Econometric Theory*, 20, 1203–1226.

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.

Lin, C.-F. J., and T. Teräsvirta (1994): "Testing the Constancy of Regression Parameters Against Continuous Structural Change," *Journal of Econometrics*, 62, 211–228.

Ling, S., and W. Li (1997): "Diagnostic Checking of Nonlinear Multivariate Time Series with Multivariate ARCH Errors," *Journal of Time Series Analysis*, 18, 447–464.

Ling, S., and M. McAleer (2003): "Asymptotic Theory for a New Vector ARMA-GARCH Model," *Econometric Theory*, 19, 280–310.

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

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988): "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika*, 75, 491–499.

Meitz, M., and P. Saikkonen (in press): "Ergodicity, Mixing, and Existence of Moments of a Class of Markov Models with Applications to GARCH and ACD Models," *Econometric Theory.*

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

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

——— (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.

Teräsvirta, T., and Z. Zhao (2007): "Stylized Facts of Return Series, Robust Estimates, and Three Popular Models of Volatility," SSE/EFI Working Paper Series in Economics and Finance No. 662, Stockholm School of Economics.

Wooldridge, J. M. (1990): "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory*, 6, 17–43.

——— (1991): "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics*, 47, 5–46.

Appendix A

Proof of Theorem 1. Assuming the independent innovations to be normally distributed, it follows that for model (11), the conditional log-likelihood function is given by

$$L_T(\boldsymbol{\theta}) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln h_t - \frac{1}{2}\sum_{t=1}^{T}\frac{\varepsilon_t^2}{h_t}.$$

Let $\boldsymbol{\theta}$ be a parameter vector partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$. The null hypothesis is $\boldsymbol{\theta}_2 = \mathbf{0}$. The corresponding partition of the average score vector $\mathbf{q}_{(T)}(\boldsymbol{\theta})$ is $\mathbf{q}_{(T)}(\boldsymbol{\theta}) = (\mathbf{q}_{1(T)}(\boldsymbol{\theta}_1)'$, $\mathbf{q}_{2(T)}(\boldsymbol{\theta}_2)')'$. Let h_t^0 denote the conditional variance under the null hypothesis and let the true parameter vector under H_0 be $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_1^{0\prime}, \mathbf{0}')'$. The Lagrange multiplier statistic is defined as follows:

$$\xi_{LM} = T\mathbf{q}_{(T)}(\hat{\boldsymbol{\theta}})'\mathbf{I}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{q}_{(T)}(\hat{\boldsymbol{\theta}})$$

where T is the sample size, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1', \mathbf{0}')'$ is the constrained maximum likelihood estimator of $\boldsymbol{\theta}$,

$$\mathbf{q}_{(T)}(\hat{\boldsymbol{\theta}}) = (\mathbf{0}', \mathbf{q}_{2(T)}(\mathbf{0})')' = (\mathbf{0}', \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}|_{\mathrm{H}_0})'$$

is the average score vector and $\mathbf{I}(\hat{\boldsymbol{\theta}})$ the information matrix, both evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. In this case, the partial derivatives with respect to $\boldsymbol{\theta}$ have the form

$$\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t} - 1\right)\frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}} = \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t} - 1\right)\mathbf{x}_t$$

where $\mathbf{x}_t = (\mathbf{x}_{1t}', \mathbf{x}_{2t}')'$, with $\mathbf{x}_{1t} = \frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}_1}$ and $\mathbf{x}_{2t} = \frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}_2}$. Accordingly,

$$\mathbf{q}_{(T)}(\hat{\boldsymbol{\theta}}) = \frac{1}{2T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t^2}{\hat{h}_t^0} - 1\right)\hat{\mathbf{x}}_t = \left(\mathbf{0}', \frac{1}{2T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t^2}{\hat{h}_t^0} - 1\right)\hat{\mathbf{x}}_{2t}'\right)'$$

where $\hat{h}_t^0$ and $\hat{\mathbf{x}}_t = (\hat{\mathbf{x}}_{1t}', \hat{\mathbf{x}}_{2t}')'$ denote h_t^0 and $\mathbf{x}_t$, respectively, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Under normality, the population information matrix equals the negative expected value of the average Hessian matrix:

$$\mathbf{I}(\boldsymbol{\theta}) = -\mathsf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right].$$

The Hessian of the log-likelihood equals

$$\sum_{t=1}^{T}\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = -\frac{1}{2}\sum_{t=1}^{T}\left[\frac{\varepsilon_t^2}{h_t^3}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\theta}'} + \frac{1}{h_t}\left(\frac{\varepsilon_t^2}{h_t} - 1\right)\left(\frac{\partial^2 h_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} - \frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\theta}'}\right)\right]$$

so the information matrix becomes

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{2T}\sum_{t=1}^{T}\mathsf{E}\left[\frac{\varepsilon_t^2}{h_t^3}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\theta}'}\right] = \frac{1}{2T}\sum_{t=1}^{T}\mathsf{E}\mathbf{x}_t\mathbf{x}_t'.$$

As the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}_0$,

$$\mathbf{I}(\hat{\boldsymbol{\theta}}) = \frac{1}{2T}\sum_{t=1}^{T}\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t'$$

is consistent for $\mathbf{I}(\boldsymbol{\theta})$. Then the Lagrange multiplier type test statistic for testing parameter constancy has the standard form:

$$\begin{aligned}
\xi_{LM} &= \frac{1}{2}\sum_{t=1}^{T}\hat{u}_t\hat{\mathbf{x}}_t'\left(\sum_{t=1}^{T}\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t'\right)^{-1}\sum_{t=1}^{T}\hat{\mathbf{x}}_t\hat{u}_t \\
&= \frac{1}{2}\sum_{t=1}^{T}\hat{u}_t\hat{\mathbf{x}}_{2t}'\left\{\sum_{t=1}^{T}\hat{\mathbf{x}}_{2t}\hat{\mathbf{x}}_{2t}'-\sum_{t=1}^{T}\hat{\mathbf{x}}_{2t}\hat{\mathbf{x}}_{1t}'\left(\sum_{t=1}^{T}\hat{\mathbf{x}}_{1t}\hat{\mathbf{x}}_{1t}'\right)^{-1}\sum_{t=1}^{T}\hat{\mathbf{x}}_{1t}\hat{\mathbf{x}}_{2t}'\right\}^{-1}\sum_{t=1}^{T}\hat{\mathbf{x}}_{2t}\hat{u}_t.
\end{aligned}$$

where $\hat{u}_t = \varepsilon_t^2/\hat{h}_t^0 - 1$. Under H_0 and standard regularity conditions, the statistic ξ_{LM} has an asymptotic χ^2-distribution with $\dim(\boldsymbol{\theta}_2)$ degrees of freedom. ■

Appendix B

Table 1. Model selection frequencies based on the additive sequential procedure

α_1	β_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$_3$
DGP (i): GARCH model with $\alpha_0 = 0.10$								
0.10	0.80	$r = 0$	**95.45**	**94.15**	**95.00**	**95.35**	**95.50**	**94.35**
		$r = 1$	3.36	4.10	3.90	3.05	3.30	3.90
		$r \geq 2$	1.20	1.75	1.10	1.60	1.20	1.75
0.10	0.85	$r = 0$	**94.70**	**91.90**	**94.30**	**94.10**	**95.10**	**93.70**
		$r = 1$	3.65	5.20	4.60	3.90	3.40	4.45
		$r \geq 2$	1.65	2.90	1.05	2.00	1.50	1.85
0.05	0.90	$r = 0$	**94.45**	**90.45**	**94.20**	**93.75**	**94.40**	**93.25**
		$r = 1$	4.00	6.35	4.50	4.20	4.00	4.40
		$r \geq 2$	1.55	3.20	1.30	2.05	1.60	2.35
0.09	0.90	$r = 0$	**90.45**	**76.95**	**92.80**	**88.10**	**94.30**	**90.85**
		$r = 1$	8.20	14.10	4.95	7.10	4.10	5.50
		$r \geq 2$	1.35	8.95	2.25	4.80	1.60	3.65

Notes: Selection frequencies in percentage of the standard LM parameter constancy test based on 2000 replications. 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 based on the additive sequential procedure

Parameters		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 (ii): Change only in the constant								
α_{01}=0.10	γ_1=5	$r=0$	64.75	72.65	31.75	50.55	6.70	19.25
		$r=1$	**33.25**	**24.00**	**65.80**	**46.95**	**91.45**	**78.10**
		$r\geq 2$	2.00	3.35	2.45	2.50	1.85	3.65
	γ_1=10	$r=0$	46.50	58.10	9.45	21.90	0.20	1.20
		$r=1$	**51.35**	**37.35**	**88.35**	**75.75**	**97.45**	**96.65**
		$r\geq 2$	2.15	4.55	2.20	2.35	2.35	2.15
α_{01}=0.30	γ_1=5	$r=0$	17.90	33.55	0.25	2.85	0.00	0.00
		$r=1$	**78.35**	**60.90**	**97.15**	**93.75**	**97.35**	**96.75**
		$r\geq 2$	3.75	5.55	2.60	3.40	2.65	3.25
	γ_1=10	$r=0$	5.75	11.15	0.00	0.00	0.00	0.00
		$r=1$	**90.45**	**82.00**	**98.05**	**96.70**	**97.15**	**96.25**
		$r\geq 2$	3.80	6.85	1.95	3.30	2.85	3.75
DGP (iii): Change only in the ARCH component								
α_{11}=0.05	γ_1=5	$r=0$	80.25	84.55	55.80	69.65	22.65	41.30
		$r=1$	**18.70**	**13.40**	**42.70**	**28.95**	**75.15**	**56.45**
		$r\geq 2$	1.05	2.05	1.50	1.40	2.20	2.25
	γ_1=10	$r=0$	68.10	76.35	29.05	47.00	3.65	10.80
		$r=1$	**30.20**	**21.60**	**68.85**	**51.00**	**93.55**	**86.95**
		$r\geq 2$	1.70	2.05	2.10	2.00	2.80	2.25
α_{11}=0.09	γ_1=5	$r=0$	50.30	62.05	7.75	22.20	0.00	1.10
		$r=1$	**46.95**	**34.40**	**89.20**	**74.95**	**96.35**	**95.80**
		$r\geq 2$	2.75	3.55	3.05	2.85	3.65	3.10
	γ_1=10	$r=0$	27.80	38.65	0.95	3.05	0.00	0.00
		$r=1$	**68.95**	**56.90**	**95.70**	**94.70**	**96.85**	**96.80**
		$r\geq 2$	3.35	4.45	3.35	2.25	3.15	3.20
DGP (v): Change only in the GARCH component								
β_{11}=0.05	γ_1=5	$r=0$	68.40	76.50	30.60	50.80	5.00	14.90
		$r=1$	**30.00**	**21.10**	**67.40**	**47.05**	**93.30**	**83.05**
		$r\geq 2$	1.60	2.40	2.00	2.15	1.70	2.05
	γ_1=10	$r=0$	50.15	62.30	8.35	21.20	0.10	0.70
		$r=1$	**47.90**	**34.90**	**89.55**	**76.50**	**97.35**	**97.25**
		$r\geq 2$	1.95	2.80	2.30	2.30	2.55	2.05
β_{11}=0.09	γ_1=5	$r=0$	18.75	30.10	0.00	1.80	0.00	0.00
		$r=1$	**77.75**	**64.90**	**95.95**	**94.85**	**96.35**	**96.65**
		$r\geq 2$	3.50	5.00	4.05	3.35	3.65	3.35
	γ_1=10	$r=0$	9.80	11.25	0.00	0.00	0.00	0.00
		$r=1$	**85.90**	**81.70**	**96.60**	**96.20**	**97.05**	**96.00**
		$r\geq 2$	4.30	7.05	3.40	3.80	1.95	4.00

Notes: Selection frequencies in percentage of the standard LM parameter constancy test based on 2000 replications. 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. Model selection frequencies based on the additive sequential procedure

α_{01}	α_{11}	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$_3$
DGP (iv): Change in the intercept and ARCH component									
0.10	0.05	5	$r=0$	24.55	41.40	0.75	3.80	0.00	0.00
			$r=1$	**72.75**	**55.05**	**96.70**	**93.85**	**97.65**	**97.50**
			$r\geq 2$	2.70	3.55	2.55	2.35	2.35	2.50
		10	$r=0$	7.80	17.05	0.00	0.05	0.00	0.00
			$r=1$	**88.80**	**79.45**	**97.80**	**97.65**	**97.35**	**97.40**
			$r\geq 2$	3.40	3.50	2.20	2.30	2.65	2.60
0.10	0.09	5	$r=0$	11.80	25.50	0.00	0.80	0.00	0.00
			$r=1$	**84.35**	**69.45**	**96.90**	**95.85**	**96.35**	**96.20**
			$r\geq 2$	3.85	5.05	3.10	3.35	3.65	3.80
		10	$r=0$	3.25	7.45	0.00	0.00	0.00	0.00
			$r=1$	**91.95**	**87.25**	**96.90**	**96.90**	**96.40**	**97.10**
			$r\geq 2$	4.80	5.30	3.10	3.10	3.60	2.90
0.30	0.05	5	$r=0$	2.30	8.80	0.00	0.00	0.00	0.00
			$r=1$	**93.20**	**86.35**	**97.35**	**97.40**	**97.05**	**96.90**
			$r\geq 2$	4.50	4.85	2.65	2.60	2.95	3.10
		10	$r=0$	0.95	0.90	0.00	0.00	0.00	0.00
			$r=1$	**94.95**	**93.65**	**96.95**	**97.15**	**97.15**	**97.20**
			$r\geq 2$	4.10	5.45	3.05	0.50	2.85	2.80
0.30	0.09	5	$r=0$	1.65	5.80	0.00	0.00	0.00	0.00
			$r=1$	**92.40**	**86.65**	**96.60**	**96.25**	**96.60**	**96.60**
			$r\geq 2$	5.95	7.55	3.40	3.75	3.40	3.40
		10	$r=0$	0.55	0.45	0.00	0.00	0.00	0.00
			$r=1$	**92.40**	**91.30**	**96.05**	**95.50**	**95.80**	**95.70**
			$r\geq 2$	7.05	8.25	3.95	4.50	4.20	4.30

Notes: Selection frequencies in percentage of the standard LM parameter constancy test based on 2000 replications. 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 4. Model selection frequencies based on the additive sequential procedure

			Number of	$T = 1000$		$T = 2500$		$T = 5000$	
α_{01}	β_{11}	γ_1	transitions	LM_1	LM_3	LM_1	LM_3	LM_1	LM_3
DGP (vi): Change in the intercept and GARCH component									
0.10	0.05	5	$r = 0$	12.45	27.20	0.00	0.70	0.00	0.00
			$r = 1$	**84.85**	**69.55**	**97.30**	**96.60**	**97.60**	**97.25**
			$r \geq 2$	2.70	3.25	2.70	2.70	2.40	2.75
		10	$r = 0$	4.50	8.45	0.00	0.00	0.00	0.00
			$r = 1$	**92.50**	**87.85**	**97.75**	**97.50**	**97.05**	**97.30**
			$r \geq 2$	3.00	3.70	2.25	2.50	2.95	2.70
0.10	0.09	5	$r = 0$	2.95	7.45	0.00	0.00	0.00	0.00
			$r = 1$	**91.70**	**86.35**	**95.65**	**95.55**	**96.50**	**96.50**
			$r \geq 2$	5.35	6.20	4.35	4.45	3.50	3.50
		10	$r = 0$	3.20	1.55	0.00	0.00	0.00	0.00
			$r = 1$	**90.90**	**89.85**	**95.40**	**93.90**	**96.15**	**95.15**
			$r \geq 2$	5.90	8.60	4.60	6.10	3.85	4.85
0.30	0.05	5	$r = 0$	1.25	4.35	0.00	0.00	0.00	0.00
			$r = 1$	**94.65**	**90.00**	**96.70**	**96.50**	**97.25**	**97.15**
			$r \geq 2$	4.10	5.65	3.30	3.50	2.75	2.85
		10	$r = 0$	1.15	0.55	0.00	0.00	0.00	0.00
			$r = 1$	**95.05**	**95.00**	**96.80**	**96.20**	**97.20**	**96.50**
			$r \geq 2$	3.80	4.45	3.20	3.80	2.80	3.50
0.30	0.09	5	$r = 0$	0.60	1.70	0.00	0.00	0.00	0.00
			$r = 1$	**91.70**	**89.10**	**95.35**	**94.15**	**96.10**	**95.95**
			$r \geq 2$	7.70	9.20	4.65	5.85	3.90	4.05
		10	$r = 0$	1.25	0.20	0.00	0.00	0.00	0.00
			$r = 1$	**91.60**	**88.00**	**94.60**	**91.35**	**94.30**	**92.70**
			$r \geq 2$	7.15	11.80	5.40	8.65	5.70	7.30

Notes: Selection frequencies in percentage of the standard LM parameter constancy test based on 2000 replications. 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 5. 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	−4.3309	−6.3139	−4.1444	−1.9042	−6.0724
Maximum	4.9887	3.0374	4.0498	2.7618	1.4231	4.0671
Skewness	−0.3678	−0.3361	−0.3898	−0.9045	−0.2839	−0.2424
Robust SK	0.0325	0.0318	0.0229	−0.0045	−0.0165	−0.0217
Ex.kurtosis	5.2867	2.7996	2.1736	14.593	3.2055	2.1941
Robust KR	0.2541	0.1737	0.1503	0.1662	0.1120	0.1030
Std. dev.	0.8912	0.6120	0.9980	0.4150	0.2887	0.9971
Mean	0.0538	0.0407	0.0621	0.0077	0.0035	0.0142
LJB	3004.53 (0.0000)	874.21 (0.0000)	562.31 (0.0000)	18558.57 (0.0000)	909.62 (0.0000)	433.38 (0.0000)
ARCH(4)	154.19 (3×10^{-32})	55.340 (3×10^{-11})	4.056 (0.3478)	339.69 (3×10^{-72})	108.07 (2×10^{-22})	5.111 (0.2761)
T	2531	2531	2531	2060	2060	2060

Notes: LJB denotes the Lomnicki-Jarque-Bera 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 6. Estimation results for the GJR-GARCH and GARCH models in the two applications

GJR-GARCH model

	S&P 500		$\varepsilon_t/\hat{g}_{t_{S\&P500}}^{1/2}$ First iteration		$\varepsilon_t/\hat{g}_{t_{S\&P500}}^{1/2}$ Last iteration	
	Estimate	(Std. error)	Estimate	(Std. error)	Estimate	(Std. error)
$\hat{\alpha}_0$	0.009	(0.003)	0.024	(0.007)	0.033	(0.008)
$\hat{\alpha}_1$	0.014	(0.007)	–	(–)	–	(–)
$\hat{\beta}_1$	0.939	(0.012)	0.901	(0.019)	0.855	(0.031)
$\hat{\gamma}_1$	0.079	(0.017)	0.123	(0.024)	0.125	(0.023)
$\hat{\alpha}_1 + \hat{\beta}_1 + \frac{\hat{\gamma}_1}{2}$	0.993		0.963		0.918	
Log-Lik	−3034.98		−2760.91		−2280.98	

GARCH model

	SPD/USD		$\varepsilon_t/\hat{g}_{t_{SPD/USD}}^{1/2}$ First iteration		$\varepsilon_t/\hat{g}_{t_{SPD/USD}}^{1/2}$ Last iteration	
	Estimate	(Std. error)	Estimate	(Std. error)	Estimate	(Std. error)
$\hat{\alpha}_0$	9×10^{-4}	(3×10^{-4})	0.002	(0.001)	0.003	(0.001)
$\hat{\alpha}_1$	0.057	(0.017)	0.058	(0.019)	0.065	(0.021)
$\hat{\beta}_1$	0.937	(0.020)	0.929	(0.027)	0.901	(0.031)
$\hat{\alpha}_1 + \hat{\beta}_1$	0.994		0.987		0.966	
Log-Lik	−635.66		−455.38		−283.45	

Table 7. Test of parameter constancy of the GARCH(1,1) model against a time-varying GARCH model with additive structure for several combinations of parameters

S&P 500 returns

	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
α_0	16.695	8×10^{-4}	0.087	0.7680	15.713	7×10^{-5}	0.901	0.3425
α_1	9.477	0.0236	3.364	0.0667	4.652	0.031	1.473	0.2249
β_1	8.779	0.0324	1.886	0.1697	6.083	0.014	0.817	0.3660
α_0 and α_1	22.829	7×10^{-4}	4.293	0.1169	17.071	2×10^{-4}	1.507	0.4708
α_0 and β_1	19.974	0.0028	2.181	0.3360	16.900	2×10^{-4}	0.914	0.6331
α_0, α_1 and β_1	26.415	0.0017	6.694	0.0823	16.518	9×10^{-4}	3.277	0.3509

SPD/USD returns

	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
α_0	5.169	0.1598	1.879	0.1705	3.011	0.0827	0.283	0.5946
α_1	15.291	0.0016	3.841	0.0500	7.598	0.0058	3.888	0.0486
β_1	10.967	0.0119	1.944	0.1633	6.721	0.0095	2.318	0.1279
α_0 and α_1	17.435	0.0078	3.750	0.1533	4.901	0.0863	8.830	0.0121
α_0 and β_1	13.523	0.0354	1.679	0.4319	4.209	0.1219	7.660	0.0217
α_0, α_1 and β_1	20.816	0.0135	6.886	0.0756	4.795	0.1874	9.203	0.0267

Table 8. 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_2	p-value	LM_2	p-value	LM_1	p-value
Single transition								
S&P 500	107.79	3×10^{-23}	5.96	0.0146	66.99	3×10^{-16}	36.03	2×10^{-9}
SPD/USD	95.74	1×10^{-20}	0.10	0.7528	32.73	1×10^{-8}	63.93	1×10^{-15}
Double transition								
S&P 500	14.10	0.0028	1.31	0.2523	0.06	0.8006	12.74	0.0004
SPD/USD	5.51	0.1380	3.09	0.0788	0.71	0.3993	1.72	0.1903
Triple transition								
S&P 500	7.32	0.0623	1.86	0.1730	0.03	0.8611	5.44	0.0197
SPD/USD	–	–	–	–	–	–	–	–

Table 9. Misspecification tests for the GARCH models in the two applications

	No ARCH-in-GARCH			GARCH(1,1) vs. GARCH(1,2)	GARCH(1,1) vs GARCH(2,1)	No additional transition	No STGARCH with $K = 1$
	$r = 1$	$r = 5$	$r = 10$				
S&P 500							
LM test	0.841	6.167	7.223	0.001	0.303	9.993	20.088
p-value	0.359	0.290	0.704	0.971	0.582	0.019	4×10^{-5}
$\varepsilon_t/\hat{g}_{t_{S\&P500}}^{1/2}$							
LM test	0.155	5.115	6.824	0.558	0.738	0.677	45.387
p-value	0.694	0.402	0.742	0.455	0.390	0.879	1×10^{-10}

	No ARCH-in-GARCH			GARCH(1,1) vs. GARCH(1,2)	GARCH(1,1) vs GARCH(2,1)	No additional transition	No STGARCH with $K = 1$
	$r = 1$	$r = 5$	$r = 10$				
SPD/USD							
LM test	0.343	12.83	15.94	0.666	2.772	14.93	2.741
p-value	0.558	0.025	0.101	0.415	0.096	0.002	0.254
$\varepsilon_t/\hat{g}_{t_{SPD/USD}}^{1/2}$							
LM test	0.014	6.988	10.34	0.064	2.607	2.133	1.167
p-value	0.905	0.222	0.411	0.780	0.106	0.545	0.558

Notes: The tests are those against remaining ARCH in the standardized residuals, GARCH(1,2) and GARCH(2,1) models, additional transition in the function g_t, and STGARCH(1,1) model of order 1.

Table 10. Misspecification tests for the GJR-GARCH model in the stock returns application

	No ARCH-in-GARCH			GARCH(1,1) vs.	GARCH(1,1) vs	No additional	No STGARCH
	$r = 1$	$r = 5$	$r = 10$	GARCH(1,2)	GARCH(2,1)	transition	with $K = 1$
S&P 500							
LM test	0.518	2.094	3.060	0.039	8.317	17.176	13.479
p-value	0.472	0.836	0.980	0.844	0.004	7×10^{-4}	0.001
$\varepsilon_t/\hat{g}_{t_{S\&P500}}^{1/2}$							
LM test	2.450	4.963	7.821	0.060	3.293	0.184	6.704
p-value	0.118	0.420	0.646	0.806	0.070	0.980	0.035

Notes: The tests are those against remaining ARCH in the standardized residuals, GARCH(1,2) and GARCH(2,1) models, additional transition in the function g_t, and STGARCH(1,1) model of order 1.

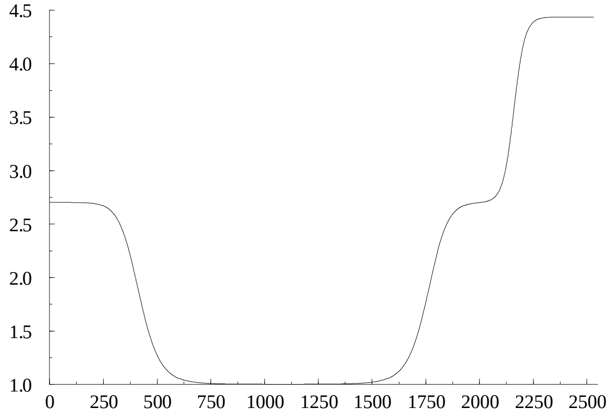


Figure 9. Graph of the final estimated function g_t for the S&P 500 returns model as a smooth function of the rescaled time variable t^* as given in (54)-(56).

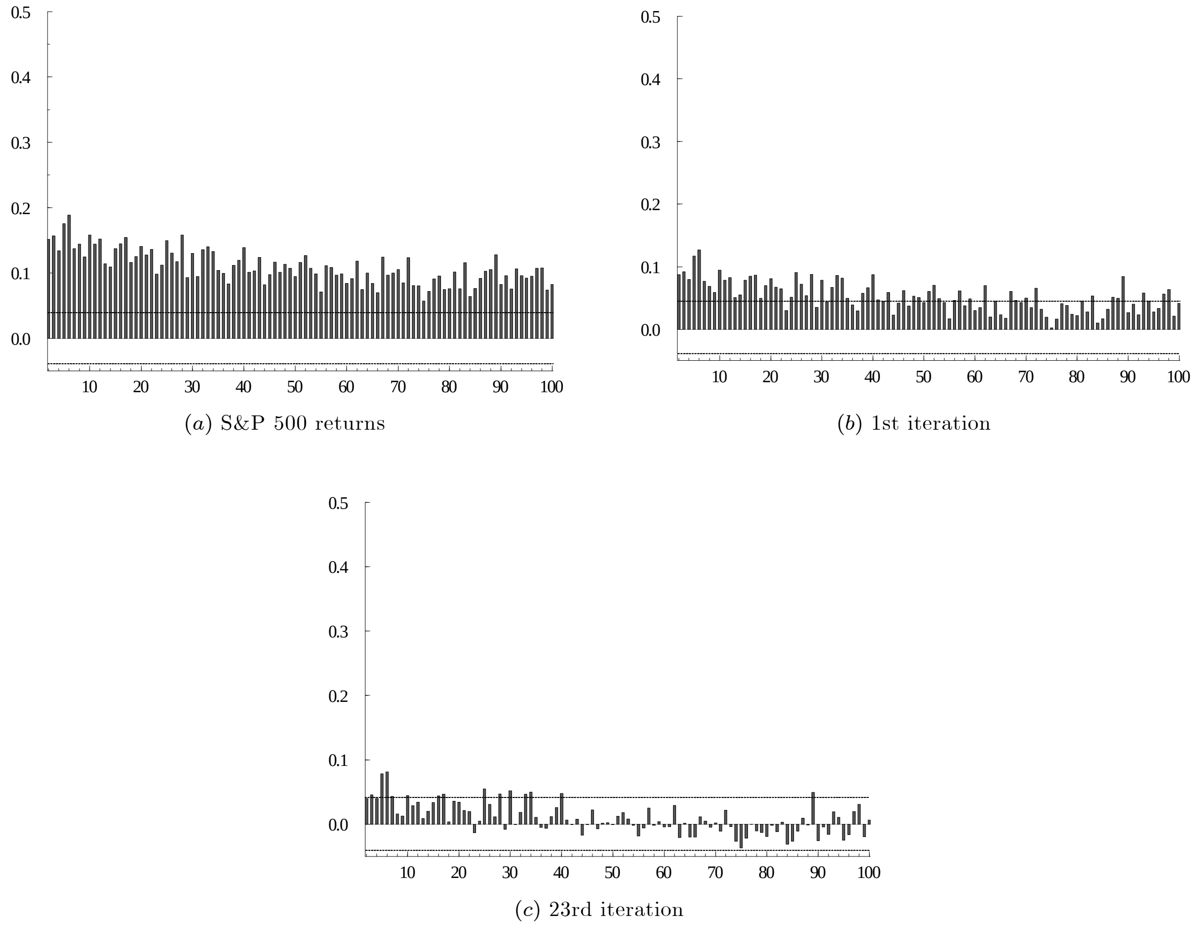
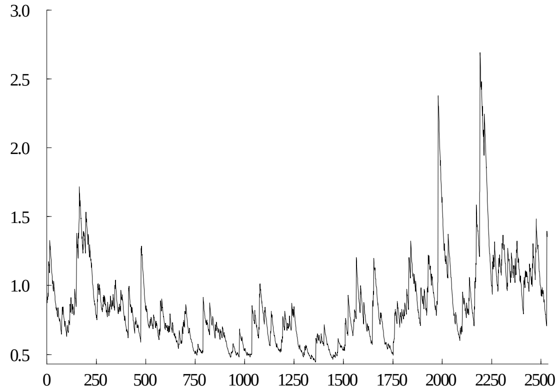
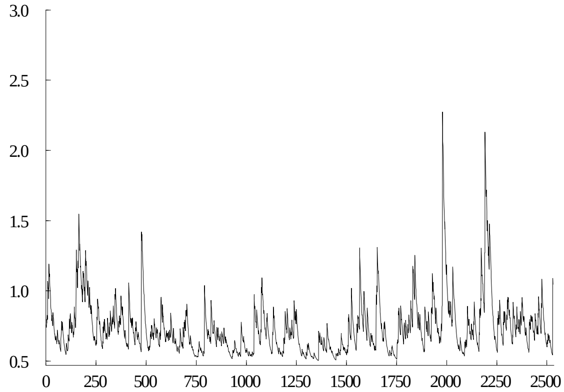


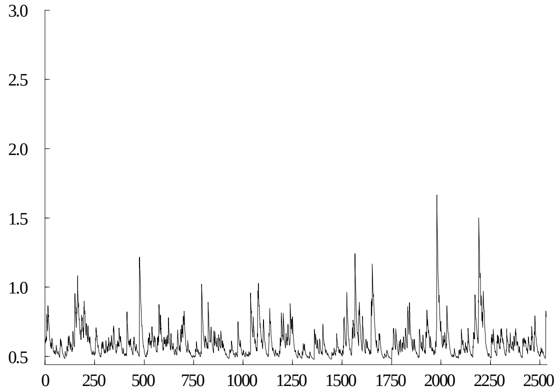
Figure 10. Sample autocorrelations of absolute log returns of the S&P 500 returns and the standardized variable $|\varepsilon_t|/\hat{g}_{t_{S\&P500}}^{1/2}$ for the first and the final iterations with the 95% confidence bounds.



(a) S&P 500 returns



(b) 1st iteration



(c) 23rd iteration

Figure 11. 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.

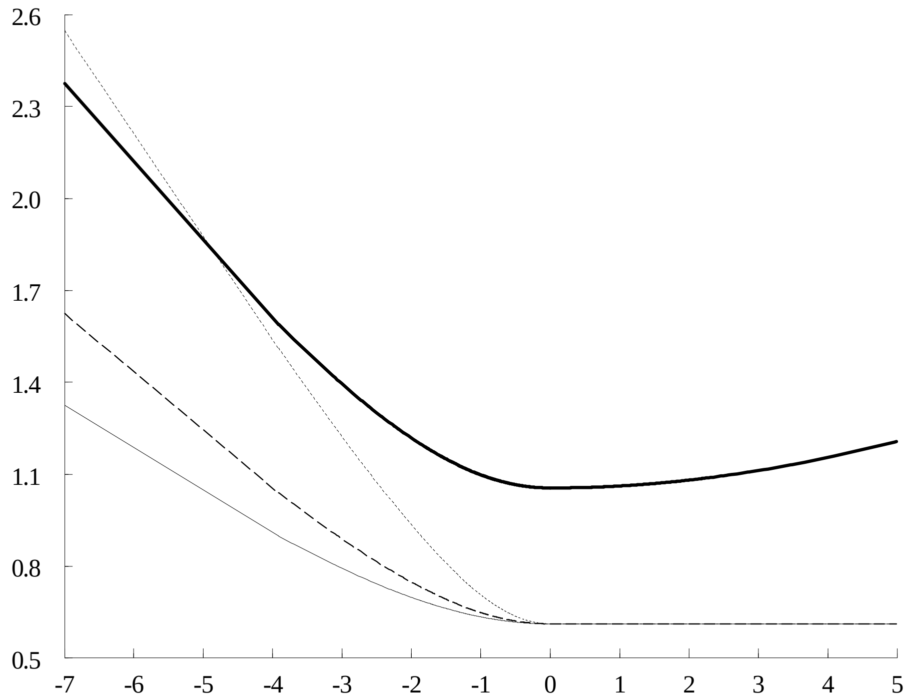


Figure 12. News impact curves of the GJR-GARCH(1,1) (solid line in boldface) and the TV-GJR-GARCH(1,1) models for several regimes. The time-varying news impact curves are plotted for the lower regime, i.e. $G_1(t^*) = G_2(t^*) = 0$ (dotted line), for an intermediate regime, i.e. $G_1(t^*) = 1$ and $G_2(t^*) = 0$ (dashed line) and for the higher regime, i.e. $G_1(t^*) = G_2(t^*) = 1$ (solid line).

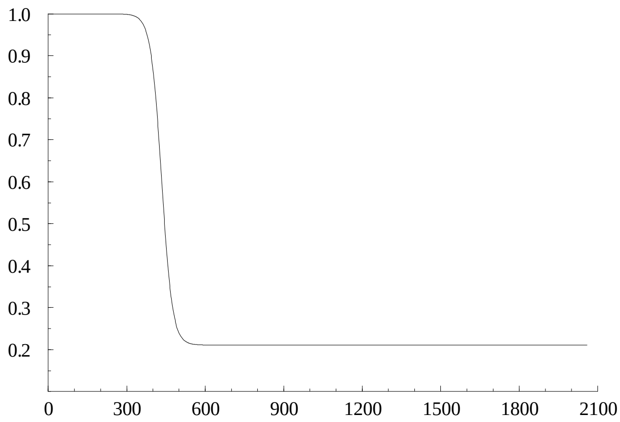
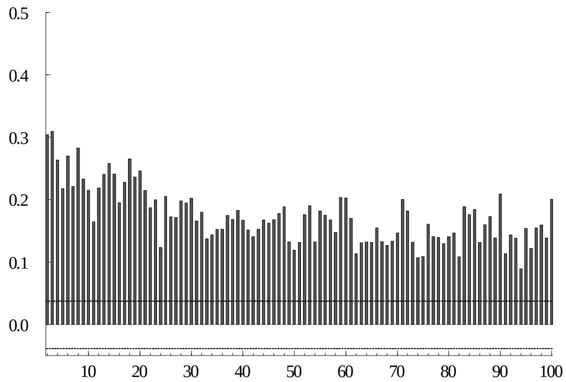
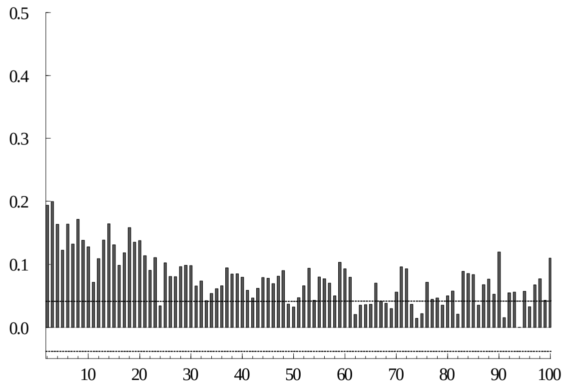


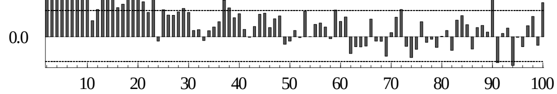
Figure 13. Graph of the final estimated function g_t for the SPD/USD returns model as a smooth function of the rescaled time variable t^* as given in (57)-(58).



(a) SPD/USD returns

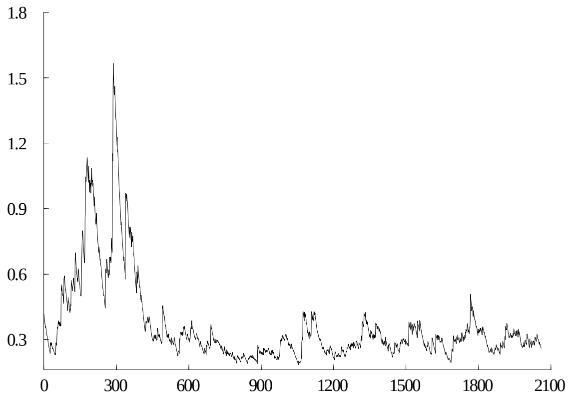


(b) 1st iteration

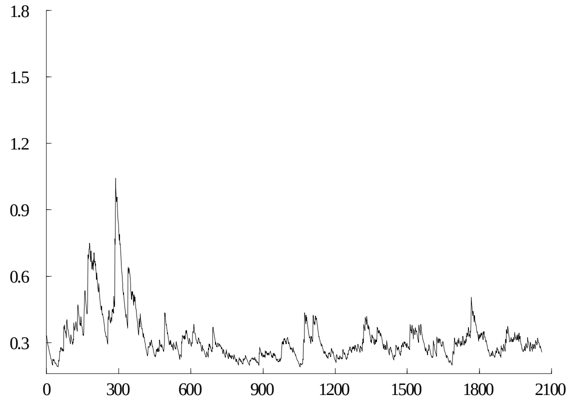


(c) 41st iteration

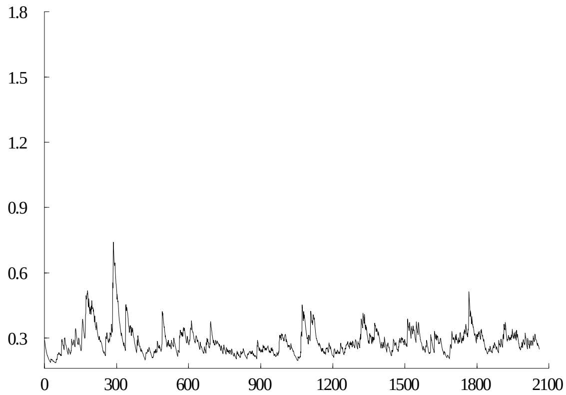
Figure 14. Sample autocorrelations of absolute log returns of the SPD/USD returns and for the standardized variable $|\varepsilon_t|/\hat{g}_{t_{SPD/USD}}^{1/2}$ for the first and the final iterations with the 95% confidence bounds.



(a) SPD/USD returns



(b) 1st iteration



(c) 41st iteration

Figure 15. Conditional standard deviation of the GARCH(1,1) model for the SPD/USD returns and for the standardized variable $\varepsilon_t/\hat{g}_{t_{SPD/USD}}^{1/2}$ for the first and the final iterations.

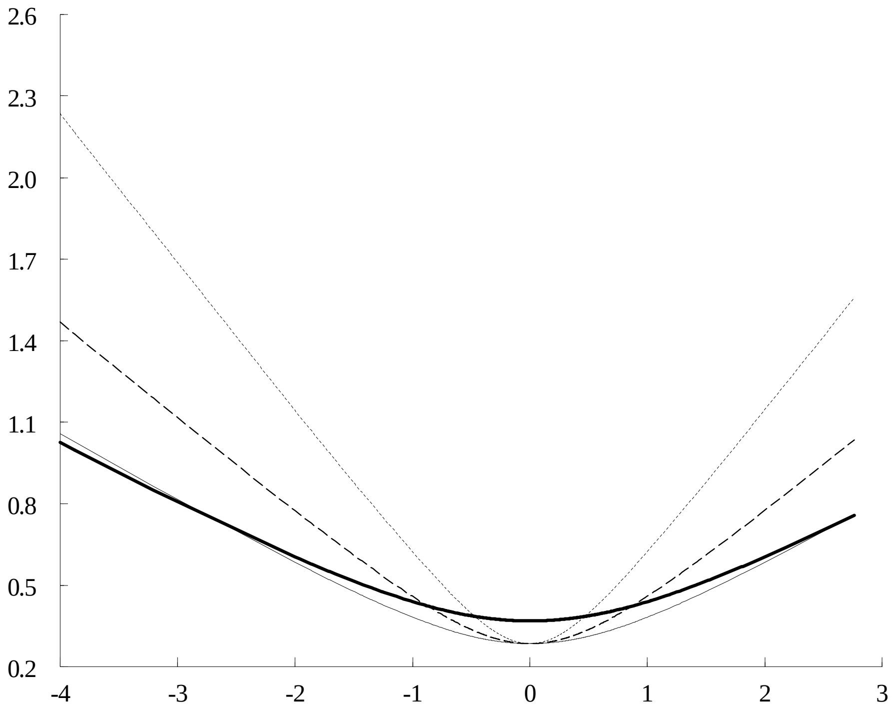


Figure 16. News impact curves of the GARCH(1,1) (solid line in boldface) and the TV-GARCH(1,1) models for several regimes. The time-varying news impact curves are plotted for the lower regime, i.e. $G_1(t^*) = 0$ (dotted line), for an intermediate regime, i.e. $G_1(t^*) = 0.5$ (dashed line) and for the higher regime, i.e. $G_1(t^*) = 1$ (solid line).