Central European Journal of Economic Modelling and Econometrics

Modelling and Forecasting WIG20 Daily Returns

Cristina Amado[*], Annastiina Silvennoinen[†] and Timo Teräsvirta[‡]

Submitted: 30.03.2017, Accepted: 15.05.2017

Abstract

The purpose of this paper is to model daily returns of the WIG20 index. The idea is to consider a model that explicitly takes changes in the amplitude of the clusters of volatility into account. This variation is modelled by a positive-valued deterministic component. A novelty in specification of the model is that the deterministic component is specified before estimating the multiplicative conditional variance component. The resulting model is subjected to misspecification tests and its forecasting performance is compared with that of commonly applied models of conditional heteroskedasticity.

Keywords: autoregressive conditional heteroskedasticity, forecasting volatility, modelling volatility, multiplicative time-varying GARCH, smooth transition

JEL Classification: C32, C52, C58

[*]University of Minho, Braga; CREATES, Aarhus University

[†]School of Economics and Finance, Queensland University of Technology, Brisbane

[‡]CREATES, Aarhus University; C.A.S.E., Humboldt-Universität zu Berlin; e-mail: tterasvirta@econ.au.dk

1 Introduction

The WIG20 index of the Warsaw Stock Exchange has been published since 16 April 1994. It is based on the value of a portfolio with shares in 20 major and most liquid companies in the main stock market. A detailed description and history of the index can be found in Brdyś, Borowa, Idźkowiak and Brdyś (2009). During its first year it did not yet comprise 20 companies and was very volatile. The index has now been published for more than 22 years, so its daily values form a rather long financial time series. There are not many published studies (in English) that analyse the (logarithmic) returns of the WIG20 index. The set of papers in which GARCH(1,1) models are fitted to the daily returns of WIG20 contains two in which the object of interest was the performance of GARCH in estimating the Value at Risk (Makiel, 2012, or Malecka, 2013). Joint volatility of WIG20 and a large number of foreign stock indices using Copula-GARCH was the concern of Czapkiewicz and Basiura (2014). These papers did not report any GARCH parameter estimates.

In these papers modelling returns of WIG20 using GARCH an implicit assumption has been that the return process is weakly stationary. In this work we question this assumption using a rather long series of WIG20 returns and test weak stationarity against the alternative that the variance of the process is time-varying. Early proponents of this view were Diebold (1986) and Lamoureux and Lastrapes (1990) who argued that high persistence in return series as viewed through GARCH may be due to shifts in the unconditional variance of the process. There is a growing literature based on a multiplicative decomposition of the return variance into a conditional variance component and a deterministic component describing changes in the unconditional variance. Examples include Feng (2004), van Bellegem and von Sachs (2004), Engle and Rangel (2008), Brownlees and Gallo (2010) and Mazur and Pipień (2012). In this work we follow the line of research started by Amado and Teräsvirta (2008), see also Amado and Teräsvirta (2013, 2014, 2017). The deterministic component is a linear combination of logistic or generalised logistic functions in which the transition variable is (rescaled) time.

One of our aims is to complete the previous literature by providing a comprehensive analysis of daily log returns of the WIG20 index using the best practices. We follow the modelling procedure outlined in Amado and Teräsvirta (2017) with modifications suggested in Silvennoinen and Teräsvirta (2016). This will be interesting *per se*, but we also consider the performance of these multiplicative time-varying GARCH models in forecasting and compare it to that of the standard GARCH(1,1) model. For this purpose we save a part of the log return series for out-of-sample forecasting.

The paper is also intended to serve as an example of what a careful analysis of a return series in the GARCH framework may require. It is structured as follows. The model is introduced in Section 2. Specification issues are discussed in Section 3, parameter estimation in Section 4 and model evaluation in Section 5. The application of the modelling strategy to the WIG20 series is described in Section 6. There is also a brief description of the data. Results from fitting two variants of the Spline-GARCH model of Engle and Rangel (2008) to the WIG20 series are reported in Section 7. In Section 8 the early observations until 1 April 2004 are discarded and our model as well as the Spline-GARCH one are fitted to the remaining subseries. Section 9 is devoted to out-of-sample forecasting and forecast comparisons. Section 10 concludes.

2 The model

The model under consideration is the time-varying GJR–GARCH model of Amado and Teräsvirta (2008, 2013, 2017). It contains a deterministic component that changes smoothly over time. To define the model, let

$$y_t = \mathsf{E}(y_t|\mathcal{F}_{t-1}) + \varepsilon_t \tag{1}$$

where $\mathcal{F}_{t-1}$ contains the historical information available at time $t-1$. For simplicity, it is assumed that $\mathsf{E}(y_t|\mathcal{F}_{t-1}) = 0$. The innovation sequence $\{\varepsilon_t\}$ has a conditional mean $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, and variance σ_t^2. The innovations are assumed to have the standard decomposition

$$\varepsilon_t = \zeta_t \sigma_t \tag{2}$$

where $\{\zeta_t\} \sim \text{iid}(0,1)$, $\mathsf{E}\zeta_t^3 = 0$, and $\mathsf{E}|\zeta_t^2|^{2+\phi} < \infty$, $\phi > 0$. The time-varying variance σ_t^2 is further decomposed multiplicatively such that

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

where h_t describes the short-run dynamics of the variance of the returns, whereas g_t is a positive-valued deterministic component. The conditional variance component h_t is modelled as the GJR–GARCH$(1,1)$ process of Glosten, Jagannathan and Runkle (1993):

$$h_t = \alpha_0 + \alpha_1 \phi_{t-1}^2 + \kappa_1 \phi_{t-1}^2 I\left(\phi_{t-1} < 0\right) + \beta_1 h_{t-1} \tag{4}$$

where $\phi_t = \varepsilon_t / g_t^{1/2}$ and $I(A)$ is the indicator variable, defined as $I(A) = 1$ when A is true, and zero otherwise. Equation (4) is assumed to satisfy the set of conditions for positivity and stationarity of the conditional variance of ϕ_t. This implies $\alpha_0 > 0$, $\alpha_1 > 0$, $\alpha_1 + \kappa_1 > 0$, $\beta_1 \geq 0$, and $\alpha_1 + \kappa_1/2 + \beta_1 < 1$. Higher-order representations of (4) are possible, but in applications of the GJR–GARCH model found in the literature the order invariably equals one.
The GJR–GARCH$(1,1)$ model is nested in (3) when $g_t \equiv 1$. The unconditional variance component g_t is smooth and time-varying, introducing nonstationarity into σ_t^2. It is defined as follows:

$$g_t\left(\frac{t}{T};\boldsymbol{\theta}_1\right) = g_t = \delta_0 + \sum_{l=1}^{r} \delta_l G_l\left(\frac{t}{T};\gamma_l,\mathbf{c}_l\right) \tag{5}$$

where $\boldsymbol{\theta}_1 = (\boldsymbol{\delta}', \boldsymbol{\gamma}', \mathbf{c}_1', ..., \mathbf{c}_r')' \in \Theta_1 = (\Delta \times \Gamma \times C)$, with $\boldsymbol{\delta} = (\delta_0, \delta_1, \ldots, \delta_r)'$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)'$, $\mathbf{c}_l' = (c_{l1}, \ldots, c_{lK_l})'$, $l = 1, \ldots, r$, is an element of the parameter space of g_t. The transition function in (5) is the general logistic transition function:

$$G_l\left(\frac{t}{T}; \gamma_l, \mathbf{c}_l\right) = \left(1 + \exp\left\{-\gamma_l \prod_{k=1}^{K_l}\left(\frac{t}{T} - c_{lk}\right)\right\}\right)^{-1} \tag{6}$$

It is a continuous and non-negative function bounded between zero and one. We make the following assumptions about (5) and (6); see Amado and Teräsvirta (2017):

AG1. The elements of $\boldsymbol{\delta} \in \Delta$ are restricted such that $\delta_0 > 0$ is fixed, $\max_{j=1,\ldots,q} |\delta_j| \leq M_\delta < \infty$ and $\inf_{\boldsymbol{\theta}_1 \in \Theta_1} g_t(\boldsymbol{\theta}_1, t/T) \geq g_{\min} > 0$.

AG2. The slope parameter $\gamma_l > 0$, $l = 1, \ldots, r$, and the location parameters $c_{1k} < c_{2k} < \cdots < c_{rk}$.

AG2 and δ_0 fixed in AG1 are identifying restrictions. The latter is needed because h_t contains a free intercept and the decomposition (3) is multiplicative. The restriction $\delta_0 = 1$ is notationally convenient. Any positive constant will do, although from the computational point of view some choices are better than some others. This constant has to be fixed to achieve identification.

The transition function allows the unconditional variance to change smoothly between regimes as a function of the transition variable $\frac{t}{T}$. The parameters $\mathbf{c}_l$ and γ_l determine the location and the speed of the transition between different regimes. When $r = K_1 = 1$, the function g_t increases monotonically over time from 1 to $1+\delta_1$ when $\delta_1 > 0$ or decreases from 1 to $1+\delta_1$ when $-1 < \delta_1 < 0$, with the location centred at $t/T = c_1$. The slope parameter γ_l in (6) controls the degree of smoothness of the transition: the larger γ_l, the faster the transition is between the extreme regimes. For example, when $r = K_1 = 1$ and $\gamma_1 \to \infty$, g_t collapses into a step function. When γ_l is large, it is numerically convenient to use a transform $\gamma_l = \exp\{\eta_l\}$ and estimate η_l; see Goodwin, Holt and Prestemon (2011) or Silvennoinen and Teräsvirta (2016). For other transformations that alleviate potential numerical problems when γ_l is large, see Chan and Theoharakis (2011) and Ekner and Nejstgaard (2013).
Equations (1)–(6) define the time-varying GJR–GARCH (TVGJR–GARCH) model. The unconditional variance in this model is time-varying and equals $\mathsf{E}\varepsilon_t^2 = \mathsf{E}\zeta_t^2 h_t g_t = g_t \mathsf{E} h_t$. This means that when $\delta_1 = \ldots = \delta_r = 0$, the unconditional variance $\mathsf{E}\varepsilon_t^2 = \delta_0 \mathsf{E} h_t$ (constant). When $\delta_l \neq 0$ for $r > 1$ and $K_l \geq 1$, equations (5) and (6) form a very flexible parameterisation capable of describing nonmonotonic deterministic changes in the unconditional variance. How to find out that g_t is a positive constant or, more generally, how to select r, will be discussed in the next section.

3 Specification of the model

Specification of g_t is a data-driven process. As already indicated, the number of transitions r is not known and has to be determined ($r = 0$ is also possible). In each transition function, K_l has to be decided. The common alternatives are $K_l = 1$ and $K_l = 2$. Since a model with $r + s$ transitions, $s > 0$, is not identified when the true number of transitions equals r, this number has to be found by proceeding from specific to general. Amado and Teräsvirta (2017) first fit a GARCH or GJR–GARCH to the series and then determine r by a sequence of specification tests, adding one transition to the model at a time. The order of the logistic function can be determined by a sequence of tests as in smooth transition autoregressive models; see Teräsvirta (1994).

Silvennoinen and Teräsvirta (2016) observed that power of the test Amado and Teräsvirta (2017) suggested may suffer from the fact that under the alternative the estimates of the sum $\alpha_1 + \kappa_1/2 + \beta_1$ tend to one. This is a natural outcome as $\alpha_1 + \kappa_1/2 + \beta_1 < 1$ is a necessary and sufficient condition for weak stationarity in first-order GJR–GARCH models. The model thus tries to accommodate as much nonstationarity generated by the deterministic component as possible. To avoid this, their solution was to specify r first, without estimating the GARCH component.

Due to leaving the conditional variance unspecified for the purpose of focusing on the specification of the deterministic, unconditional component has the implication that the test statistic no longer has its standard asymptotic distribution. Ignoring this will lead to a size distortion of the test. This problem is overcome by computing the p-values for the tests via simulation, where an artificial GARCH process is used to generate an imitation of the actual data set. This works well in simulations in which the GARCH model is known, and it turns out that size-adjusted power vastly exceeds that of the misspecification test applied in Amado and Teräsvirta (2017). The situation becomes slightly more complicated in applications where the form of neglected heteroskedasticity is unknown.

Investigations in Hall, Silvennoinen and Teräsvirta (2017) have led to the conclusion that special attention is to be placed on matching the persistence of the GARCH process present in the data. On the contrary, other features, such as implied kurtosis or relative balance of weights of the GARCH parameters only have a negligible effect on the performance of the test. As this measure of persistence is quite obviously difficult to estimate in the presence of the time-varying variance component, we proceed to estimate the standard GARCH(1,1) model over a rolling window of length 1000 observations. For each, we compute the implied persistence as well as the measure of kurtosis. If the unconditional variance indeed varies over time, most such windows may be expected to contain a source for nonstationarity, and hence the persistence estimate appears to be higher than it actually is. We therefore choose the 10th and the 25th percentiles of the resulting persistence distribution. The corresponding persistence measures are 0.95 and 0.97. Excess kurtosis turns out to stay fairly constant, just below one, over the windows with the aforementioned persistence levels, thus being a close match with the kurtosis implied by the estimated GARCH models. Using the results in He and Teräsvirta (1999) we then backtrack the corresponding GARCH parameters to be used in simulating the distribution of test statistic and calculating the p-values. Due to the higher level of persistence, using the latter measure (0.97) is expected to result in more conservative conclusions than the ones from using the former (0.95).

To summarise, the approach used here proceeds with sequential testing for an additional transition in g_t by increasing r in equation (5) while keeping $K_l = 1$ in equation (6). After the final shape of the deterministic component is specified, the resulting sequence of single transitions may be simplified and merged into fewer transitions but with $K_l = 2$. We note that the sequence of tests proposed in Teräsvirta (1994) could in principle be used for specifying the order K_l of the additive transitions. However, as we have to rely on simulations to obtain p-values for the test statistic, it turns out to be impractical to carry out the test sequence as it was originally proposed. This has mostly to do with controlling for convergence and acceptance of particular simulation rounds. This has been left for future research, as one can devise alternative methods for overcoming such issues. The performance of each of them must, however, be analysed before making any recommendations.

Instead, our approach here is to assess the strength of rejection of the null hypothesis in the test in Silvennoinen and Teräsvirta (2016) where the linear approximation for the transition under test is of linear, quadratic, or cubic form. These tests are labelled here as LM_1, LM_2, and LM_3, respectively, the last one being the test originally proposed in Silvennoinen and Teräsvirta (2016). Lack of power in each is most likely to be due to either under- or over-fitting the transition that is being tested. Hence, comparison of p-values and the test statistic values (recall that in absence of GARCH, the distributions of the three statistics are χ^2_1, χ^2_2, and χ^2_3, respectively), together with visual inspection of the data series may be used to guide the choice K_l.

4 Estimation of the model

Estimation of the parameters of the TVGJR–GARCH model is carried out by maximum likelihood. In previous work it has turned out that straightforward maximisation runs into convergence problems. A better way of maximise the log-likelihood is to do it by dividing each iteration into two parts. This has been discussed by Song, Fan and Kalbfleisch (2005). In addition to a numerical superiority this approach has a strong theoretical advantage. Using the results of Song *et al.* (2005), Amado and Teräsvirta (2013) were able to show that, under regularity conditions, maximum likelihood estimators of the parameters of the TVGJR–GARCH model are consistent and asymptotically normal. This makes it possible to consider misspecification tests for the model, see Amado and Teräsvirta (2017). This result also applies to time-varying variance (TVV) models where $h_t \equiv 1$. This fact justifies the sequential testing approach to determining the number of transitions.

5 Evaluation of the model

The estimated TVGJR–GARCH model can be evaluated both formally using misspecification tests and informally by looking at the estimates of h_t. They can be expected to satisfy $\widehat{\alpha}_1 + \widehat{\kappa}_1/2 + \widehat{\beta}_1 < 1$ by some margin. If this is not the case, there may be unmodelled nonstationary left in the process, so the model would not be satisfactory. Formal misspecification tests are generalisations of tests in Lundbergh and Teräsvirta (2002) as discussed in Amado and Teräsvirta (2017). In this work we apply the test called 'ARCH nested in GARCH'. Combining (2) and (3) gives $\varepsilon_t = \zeta_t(h_t g_t)^{1/2}$, where $\zeta_t \sim \text{iid}(0,1)$. This is the situation under the null hypothesis. Under the alternative, $\varepsilon_t = z_t(h_t g_t f_t)^{1/2}$, where now $z_t \sim \text{iid}(0,1)$, and $f_t = 1 + \sum_{j=1}^{r} \psi_j \zeta_{t-j}^2$. This means that under the alternative there is unmodelled dependence in ζ_t, characterised by an ARCH(r) process. Another alternative, not considered here, is that $f_t = 1 + \sum_{j=1}^{r} \psi_j x_{t-j}^2$, where x_t^2 is an observable positive-valued stationary random variable.
There is another, more straightforward, way of looking for unmodelled structure: testing for higher-order GARCH. Bollerslev (1986) already derived the relevant test statistics. Another test applied in this work is the test of TVGJR–GARCH against smooth transition TVGJR–GARCH, proposed by Hagerud (1997). It is a test of linearity of h_t. Under the alternative,

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1\phi_{t-1}^2 + \kappa_1\phi_{t-1}^2 I(\phi_{t-1} < 0) + \beta_1 h_{t-1} \\ &+ \{\alpha_{01} + \alpha_{11}\phi_{t-1}^2 + \kappa_{11}\phi_{t-1}^2 I(\phi_{t-1} < 0)\} G(\phi_{t-1}; \gamma, c) \end{aligned} \tag{7}$$

where

$$G(\phi_{t-1}; \gamma, c) = \left(1 + \exp\left\{-\gamma \prod_{l=1}^{L} (\phi_{t-1} - c)\right\}\right)^{-1}, \ \gamma > 0. \tag{8}$$

Under H_0: $\gamma = 0$, the transition function (8) is constant and the model thus a TVGJR–GARCH model. For details of the test, see Hagerud (1997) or Lundbergh and Teräsvirta (2002). It should be mentioned, however, that the smooth transition GJR–GARCH is a generalisation of the standard GJR–GARCH model, where (in the first-order case) $I(\phi_{t-1} < 0)$ is replaced by (8):

$$h_t = \alpha_0 + \alpha_1\phi_{t-1}^2 + \kappa_1\phi_{t-1}^2 G(\phi_{t-1}; \gamma, c) + \beta_1 h_{t-1}.$$

The null hypothesis is $\gamma = 0$ in (7). The test of this hypothesis can be viewed as a test of the hypothesis that asymmetry in the response of the conditional variance to ϕ_{t-1}, the lagged rescaled return, is adequately described by the component $\kappa_1\phi_{t-1}^2 I(\phi_{t-1} < 0)$ in (4). Test results appear in Section 6.4.

6 Fitting the model to WIG20 returns (full sample)

6.1 The data

In this section we consider modelling the WIG20 daily percentage logarithmic returns from 3 January 1996 until 31 March 2015. The series that has 4777 observations appears in Figure 1. We exclude the early observations of the index, established in April 1994, from the analysis because early on the index comprised only a small number (less than 20) stocks and was very volatile. The most recent observations from 1 April 2015 up until 30 April 2016 in our time series are saved for forecasting.

Figure 1: Daily logarithmic returns of WIG20, from 3 January 1996 to 31 March 2015

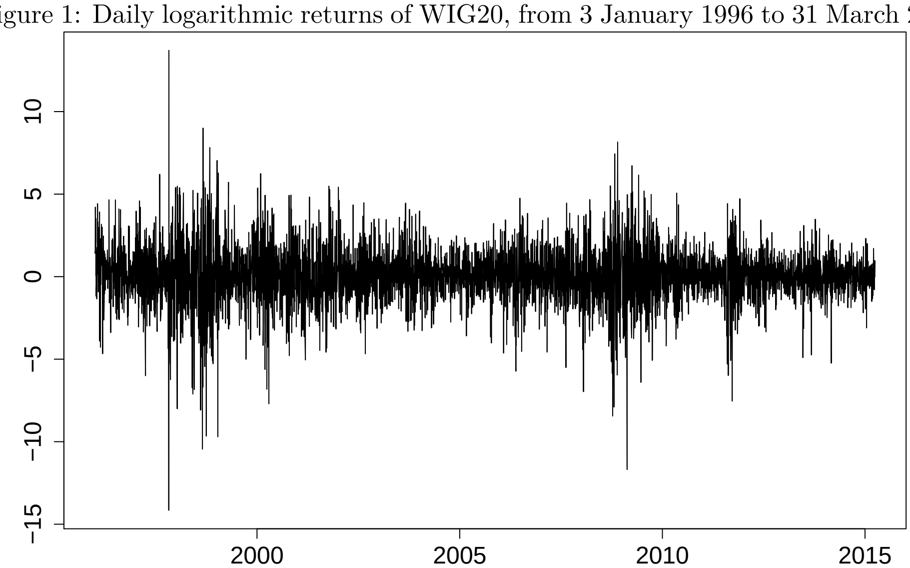


Figure 1 shows that the amplitude of the clusters is fairly high between 1997 and 2003 with a couple of very high (in absolute values) returns and decreases thereafter. There is a new increase that culminates around 2009 and a short but distinct spurt in 2011. Table 1 contains some statistics of the returns. It is seen that while the standard skewness measure indicates negative skewness, this is completely non-existent in the robust skewness measure based on quantiles. The observed skewness is due to a very small number of negative returns that do not have a counterweight on the positive side. The idea of skewing the whole error distribution is not supported by the robust skewness measure, and is not considered here.

Table 1: Statistics for the WIG20 log return series and the rescaled series $\varepsilon_t/\hat{g}_t^{1/2}$, 3 January 1996 to 31 March 2015

	WIG20	$\varepsilon_t/\hat{g}_t^{1/2}$
Minimum	−14.16	−6.295
Maximum	13.71	6.093
Mean	0.023	0.017
Std. Dev.	1.739	0.965
Skewness	−0.241	−0.226
Rob. SK	−0.006	0.008
Ex. Kurt.	4.581	2.387
Rob. KR	0.180	0.123

6.2 Specification of the model

As already mentioned, specification begins by assuming that the conditional heteroskedasticity is constant, $h_t \equiv 1$ in (3). Constancy of the deterministic component is tested using the Lagrange multiplier test for LM_3 described in Hall *et al.* (2017), as well as using the LM_2 and LM_1 tests, as explained in Section 3. The nonrobust and robust versions are used, and the p-values are found by simulation as described in Section 3. The results can be found in Table 2.

The null hypothesis is rejected, but attempts to estimate the alternative fail, which is due to the fact that the third-order Taylor approximation to the alternative is not sufficiently adequate. Consequently, the tests used to determine K_1 in (6) yield inconclusive results. The solution, resembling the one in Amado and Teräsvirta (2014) who were modelling an approximately 23000 observations long daily return series, consists of splitting the time series into two and specifying g_t separately for these two subseries. Constancy is rejected for both subseries. The shape of the transitions is determined as described in Section 3. Based on the p-values, a quadratic shape is preferred for both subseries (this being most clear in the robust test results), and hence the conclusion is that both transitions have $K_1 = 2$ in (6).

However, to avoid compromising the fit of the model for the sake of saving a few parameters at this stage, a single transition with $K_1 = 2$ is estimated as two transitions with $K = 1$ in each instead. There is no evidence to suggest that the first subseries would require additional transitions. Testing for an additional transition in the second subseries results in adding another one, and this new transition is deemed to be a second-order one. Hence, two first-order transitions are added to represent it, and a model with $r = 4$ is estimated. At this point, the nonrobust LM test does not provide evidence of yet another transition, but the robust test is pointing in the opposite direction.

The subseries are then joined and the TVV model with the six first-order 'subtransitions' estimated. The estimated model is then tested against a TVV model

Table 2: Nonrobust and robust specification tests for time-varying deterministic component in the model for the WIG20 log return series. For each transition r, $K = 1$. The simulated p-values are reported for both tests, and the subscripts refer to the measure of persistence used in the simulation.

	$r = 0$			$r = 2$			$r = 4$			$r = 6$		
	LM_1	LM_2	LM_3	LM_1	LM_2	LM_3	LM_1	LM_2	LM_3	LM_1	LM_2	LM_3
3 January 1996 – 31 March 2015												
LM test	92.2	84.9	84.1				118	113	84.5	46.8	36.3	35.3
p_{95}-value	0.000	0.000	0.000				0.006	0.009	0.027	0.848	0.874	0.866
p_{97}-value	0.000	0.000	0.000				0.008	0.010	0.029	0.844	0.869	0.852
LM Rob test	162	126	64.3				33.9	27.5	4.15	31.6	20.8	19.9
p_{95}-value	0.000	0.000	0.000				0.001	0.000	0.146	0.101	0.130	0.058
p_{97}-value	0.000	0.000	0.000				0.002	0.001	0.195	0.121	0.146	0.068
3 January 1996 – 31 March 2004												
LM test	63.3	39.8	20.6	17.7	13.7	12.1						
p_{95}-value	0.000	0.000	0.001	0.062	0.072	0.039						
p_{97}-value	0.000	0.000	0.006	0.119	0.131	0.074						
LM Rob test	62.1	60.2	28.9	6.71	1.95	0.620						
p_{95}-value	0.000	0.000	0.000	0.331	0.597	0.562						
p_{97}-value	0.000	0.000	0.001	0.443	0.669	0.624						
1 April 2004 – 31 March 2015												
LM test	116	99.4	8.58	68.3	53.3	48.9	31.6	22.6	22.6			
p_{95}-value	0.000	0.000	0.026	0.000	0.000	0.000	0.254	0.285	0.274			
p_{97}-value	0.000	0.000	0.065	0.000	0.000	0.000	0.272	0.307	0.291			
LM Rob test	103	103	28.5	51.1	50.8	48.3	29.1	22.9	22.7			
p_{95}-value	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.019	0.002			
p_{97}-value	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.023	0.006			

with an additional transition. The p-values exceed 5%, and the model is thus deemed adequate. As a robustness check against overfitting, the full data set is used to estimate a model with four first-order 'subtransitions'. The null hypothesis of four transitions is rejected for the LM_1 and LM_2 tests, however, and the shape of this transition found to be similar to the preceding ones, $K = 2$ in (6). This again points towards a model with six first-order transitions.

The final step consists of assessing the parameter estimates from the TVV model with six transitions. It turns out that the speed and location parameter estimates coincide such that the six transitions can be paired to form three second-order transitions. These transitions form the final specification of the deterministic component of the model.

6.3 Parameter estimation

The parameter estimates of the TVV model serve as starting-values for estimating the TVGJR–GARCH model, which is carried out by estimation by parts; see Song *et al.* (2005) and Amado and Teräsvirta (2013). The estimates can be found in Table 3. The estimates of the GJR–GARCH model are included in the same table for comparison. It is seen that the persistence, as measured by $\widehat{\alpha}_1 + \widehat{\kappa}_1/2 + \widehat{\beta}_1$ decreases considerably from 0.990 to 0.968 when g_t is included in the model. This has implications for forecasting. The decrease is mainly ascribed to β_1, the coefficient of the lagged conditional variance. This is in line with previous studies; see for example Amado and Teräsvirta (2014, 2017).

Table 3: Estimated GARCH components of the GJR–GARCH and TVGJR–GARCH model for the WIG20 log return series, 2 January 1996 to 31 March 2015. Standard deviation estimates in parentheses

GJR–GARCH	α_0	α_1	κ_1	β_1	$\alpha_1 + \kappa_1/2 + \beta_1$
	0.023 (0.008)	0.046 (0.009)	0.043 (0.012)	0.927 (0.013)	0.994
TVGJR–GARCH	α_0	α_1	κ_1	β_1	$\alpha_1 + \kappa_1/2 + \beta_1$
	0.030 (0.009)	0.039 (0.009)	0.055 (0.016)	0.901 (0.018)	0.968

The estimated deterministic component equals (standard deviation estimates in parentheses)

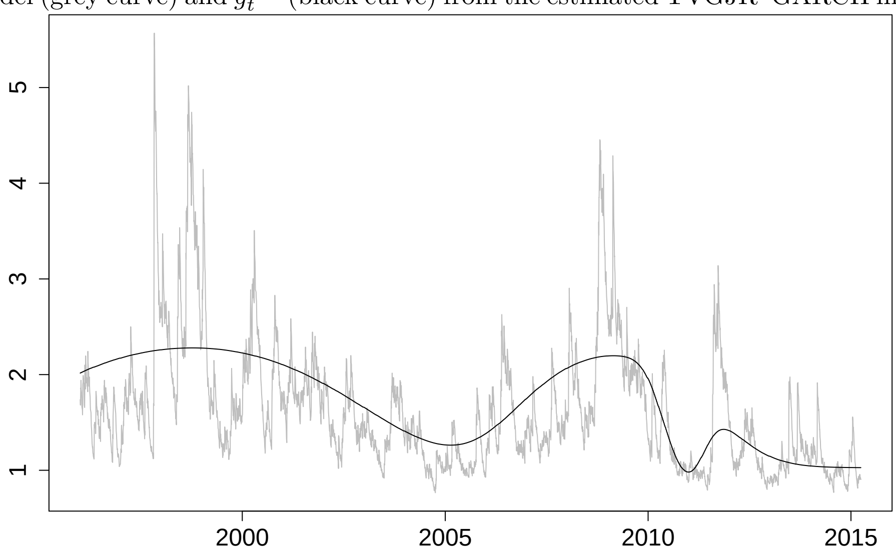
$$\widehat{g}_t = \underset{(-)}{11.643} - \underset{(0.328)}{8.269} G_1\left(\frac{t}{T}; \widehat{\gamma}_1, \widehat{\mathbf{c}}_1\right) - \underset{(0.602)}{7.543} G_2\left(\frac{t}{T}; \widehat{\gamma}_2, \widehat{\mathbf{c}}_2\right) + \underset{(0.584)}{5.232} G_3\left(\frac{t}{T}; \widehat{\gamma}_3, \widehat{\mathbf{c}}_3\right), \quad (9)$$

where the transitions are

$$
\begin{aligned}
G_1\left(\frac{t}{T};\widehat{\gamma}_1,\widehat{\mathbf{c}}_1\right) &= \left(1+\exp\left\{-\underset{(0.155)}{2.104}\left(\frac{t}{T}-\underset{(0.016)}{0.141}\right)^2\right\}\right)^{-1} \\
G_2\left(\frac{t}{T};\widehat{\gamma}_2,\widehat{\mathbf{c}}_2\right) &= \left(1+\exp\left\{-\underset{(0.103)}{3.176}\left(\frac{t}{T}-\underset{(0.006)}{0.683}\right)^2\right\}\right)^{-1} \\
G_3\left(\frac{t}{T};\widehat{\gamma}_3,\widehat{\mathbf{c}}_3\right) &= \left(1+\exp\left\{-\underset{(0.243)}{5.677}(\frac{t}{T}-\underset{(0.003)}{0.770})^2\right\}\right)^{-1}.
\end{aligned}
$$

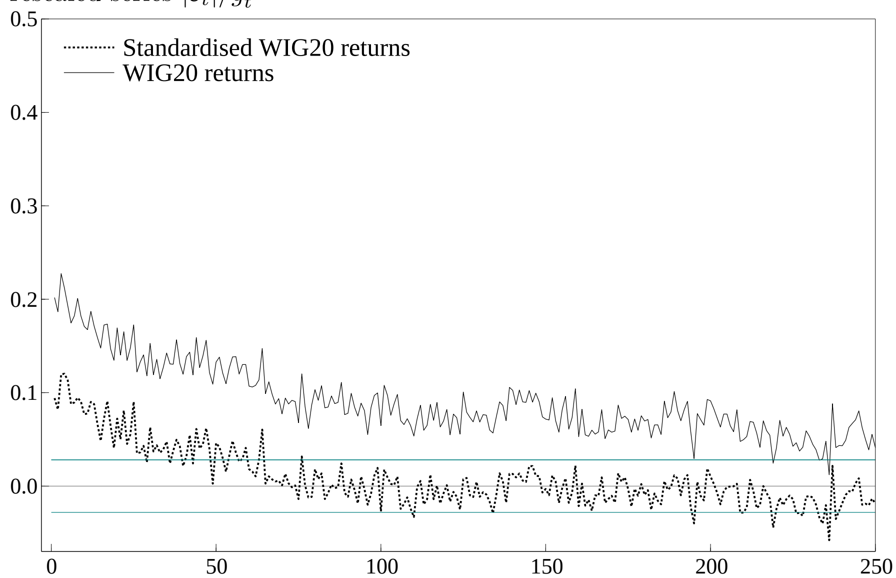
As already discussed, the intercept in (9) is fixed and so does not have a standard deviation. Interestingly, in all transitions the location parameters c_1 and c_2 are estimated to be equal. This means that the transitions are not very 'broad-shouldered' but instead rather smooth. This can be seen from Figure 2. The apparent asymmetry of the second and the third transition is due to the fact that they overlap.

Figure 2: Conditional standard deviations of WIG20 returns from the GJR–GARCH model (grey curve) and $\widehat{g}_t^{1/2}$ (black curve) from the estimated TVGJR–GARCH model



The effect of the estimated deterministic component (9) on the dependence structure of the absolute returns is visible in Figure 3. The original autocorrelations decay very slowly as a function of the lag. This phenomenon, present in many sufficiently long daily return series, has prompted researchers to model these series as a long memory process using Fractionally Integrated GARCH; see for example Baillie, Bollerslev and Mikkelsen (1996) or Davidson (2004). The autocorrelations of rescaled or standardised absolute returns decay appreciably faster than the original ones. They are from the first lag clearly lower than the original ones that are positive up until lag 250. There is still a bump between lags 30 and 60 in the former suggesting that the deterministic component may not have removed all long run dependence. What the deterministic component does to the conditional variance can be seen from Figure 4. The figure shows that rescaling removes trendlike movements in conditional standard deviations between the years 1998 and 2003, and 2007 and 2011. Furthermore, the spikes in the graph of conditional standard deviations from the GJR–GARCH model are of different magnitude, whereas the ones from the TVGJR–GARCH model are approximately of the same size. These standard deviations are determined up to a constant, that is, they are relative, as opposed to absolute, entities. This is because changes in the fixed intercept δ_0 affect the level of the two curves in Figure 4.

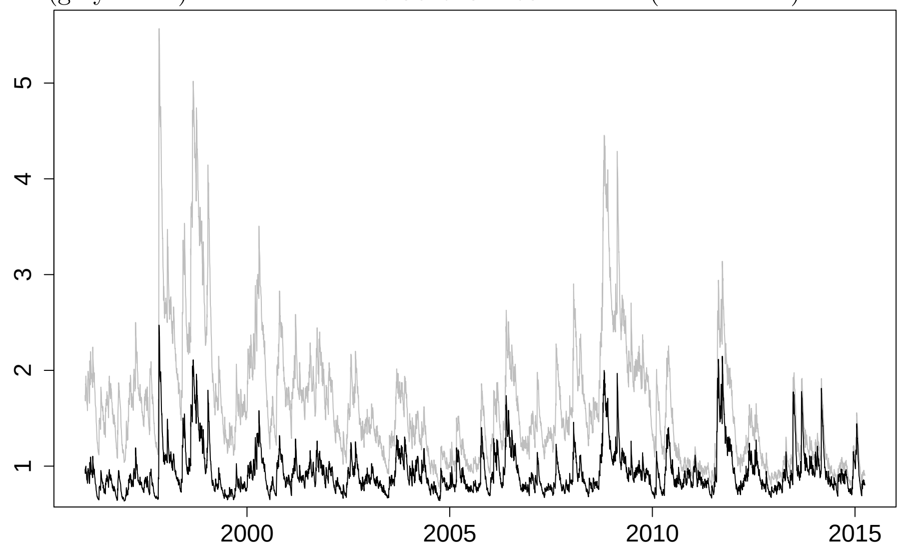
Figure 3: First 250 autocorrelations of absolute values $|\varepsilon_t|$ of the WIG20 index and the rescaled series $|\varepsilon_t|/\widehat{g}_t^{1/2}$



6.4 Evaluation

The estimated model is evaluated using misspecification tests discussed in Section 5. The results appear in Table 4. The test no ARCH in GARCH is an extension of a corresponding test in Lundbergh and Teräsvirta (2002), whereas the tests of higher-order GARCH are the ones by Bollerslev (1986) and modified for testing the TVGJR–GARCH. The robust tests (LM Rob) are the previous tests robustified as

Figure 4: Conditional standard deviations of WIG20 retuns from the GJR–GARCH model (grey curve) and from the TVGJR–GARCH model (black curve)



suggested by Wooldridge (1991), see Lundbergh and Teräsvirta (2002). Since it is well known that the nonrobust tests are positively size-distorted even in large samples, use of robust tests is encouraged. The results show that if we trust the robust versions the GJR–GARCH model passes all tests (it did fail the test of g_t being constant). The TVGJR–GARCH model fails one of the tests of higher-order GARCH. This may be surprising at first because the GJR–GARCH model passes the same test. Sometimes, however, it becomes possible to 'see' a defect in an estimated model only after bigger problems have been taken care of.

7 Spline-GARCH results

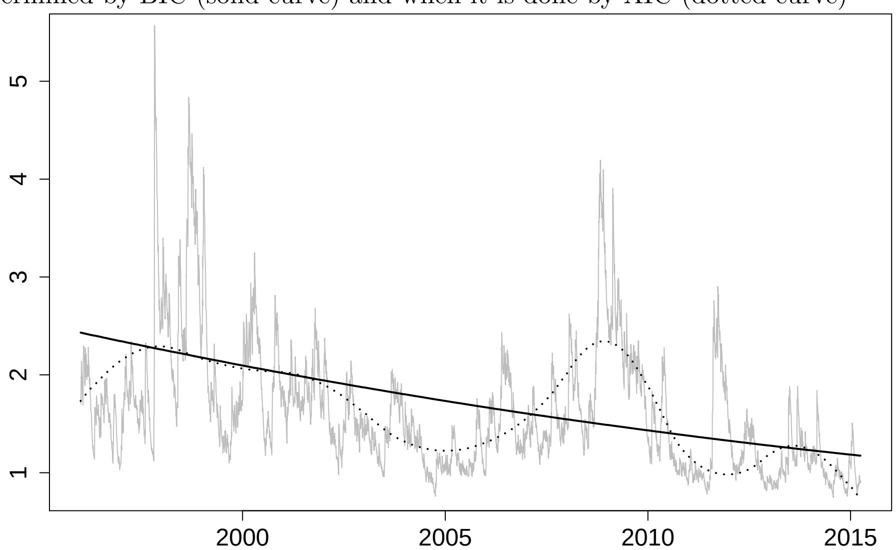
For comparison, we also present results of fitting the Spline-GARCH model to the series. Engle and Rangel (2008) used BIC of Rissanen (1978) and Schwarz (1978) to determine the number of (equidistant) knots in quadratic spline. The result can be seen in Figure 5. If, instead, the number of knots is selected by AIC of Akaike (1974), the corresponding curve in this figure follows the series more closely and bears some resemblance to Figure 2. Both have three local maxima, whereas the curve selected by BIC is very close to a straight line. This suggests that one might want to compute the deterministic component using different numbers of knots beginning from a small number quite like in sequential testing. Where to stop would be an interesting research question.

The GARCH equations of Spline-GARCH can be found in Appendix A. As may be

Table 4: Nonrobust and robust misspecification tests for the estimated GJR–GARCH (upper panel) and TVGJR–GARCH (lower panel) model for the WIG20 log return series

	No ARCH-in-GARCH			GARCH(1,1) vs. GARCH(1,2)	GARCH(1,1) vs. GARCH(2,1)	No ST–GARCH
	$r = 1$	$r = 5$	$r = 10$			
GJR						
LM test	8.326	18.34	20.61	2.841	0.003	9.357
p-value	0.004	0.003	0.024	0.092	0.953	0.009
LM Rob test	1.588	5.135	7.287	2.352	0.002	1.152
p-value	0.208	0.400	0.698	0.125	0.968	0.562
TV–GJR						
LM test	5.279	11.22	13.09	9.393	3.244	7.396
p-value	0.022	0.047	0.219	0.002	0.072	0.025
LM Rob test	1.030	7.867	10.37	7.360	0.616	0.897
p-value	0.310	0.160	0.409	0.007	0.433	0.639

Figure 5: Conditional standard deviations of WIG20 returns from the GJR–GARCH model (grey curve), exponential quadratic spline when the number of knots is determined by BIC (solid curve) and when it is done by AIC (dotted curve)



expected, the GARCH equation of the AIC-based model has lower persistence than the BIC-based one. In fact, the persistence of the latter equals 0.986 and is thus still quite close to one. For the former, the corresponding figure equals 0.967, which is practically the same as the one in Table 3 for the TVGJR–GARCH model.

In general, our test results indicate that the deterministic component cannot be neglected when modelling the WIG20 daily returns. The effect of excluding or including this component on forecasting is studied in the next section.

8 Fitting the model to WIG20 log returns (partial sample)

In this section we discuss modelling a more recent part of the return series from 1 April 2004 to 31 March 2015. This is done for two main reasons. First, it is interesting to see how much $\widehat{g}_t$ changes, if it does, compared to the corresponding part of this component in (9). Second, it may not be necessary to use all observations when constructing a TV–GARCH model for forecasting. There is evidence of this in Amado and Teräsvirta (2014) who modelled daily returns of the Dow-Jones index with almost 23000 observations using the TVGJR–GARCH model. It turned out, perhaps not surprisingly, that a model based on a much shorter subseries generated more accurate forecasts than the model estimated from the original series. Whether or not something similar occurs here will be investigated.
Sequential testing for the number of transitions from the previous Section suggests two shifts, each with $K = 2$. The estimated GARCH equations in Table 5 have not changed much compared to the ones in Table 3. The change in persistence when one moves from GJR–GARCH to TVGJR–GARCH is of the same magnitude as before. However, in the TVGJR–GARCH model the evidence of asymmetric response of the conditional variance to shocks is now quite pronounced as $\widehat{\kappa}_1$ has increased whereas $\widehat{\alpha}_1$ has shrunk and is no longer significant. The deterministic component has the following form:

$$\widehat{g}_t = \underset{(-)}{20.689} - \underset{(0.549)}{4.405} G_1\left(\frac{t}{T};\widehat{\gamma}_1,\widehat{\mathbf{c}}_1\right) - \underset{(0.537)}{14.596} G_2\left(\frac{t}{T};\widehat{\gamma}_2,\widehat{\mathbf{c}}_2\right)$$

where

$$\begin{aligned} G_1\left(\frac{t}{T};\widehat{\gamma}_1,\widehat{\mathbf{c}}_1\right) &= \left(1+\exp\left\{-\underset{(0.207)}{2.356}\left(\frac{t}{T}-\underset{(0.017)}{0.339}\right)^2\right\}\right)^{-1} \\ G_2\left(\frac{t}{T};\widehat{\gamma}_2,\widehat{\mathbf{c}}_2\right) &= \left(1+\exp\left\{-\underset{(0.188)}{5.097}\left(\frac{t}{T}-\underset{(0.004)}{0.469}\right)^2\right\}\right)^{-1}. \end{aligned}$$

There are now two transitions and the shape of $\widehat{g}_t^{1/2}$ is depicted in Figure 6. By comparing this figure with Figure 2 it is seen that the increase in $\widehat{g}_t^{1/2}$ around 2009 is more pronounced in the former than the latter, but otherwise the two curves look fairly similar. The small hump around 2012 in Figure 2 has, however, vanished in Figure 6. As the counterpart of Figure 4, Figure 8 shows the same situation: the apparent nonstationarity around 2007–2010 in conditional heteroskedasticity is

Figure 6: Conditional standard deviations from the GJR–GARCH(1,1) model (grey curve) and estimated $\widehat{g}_t^{1/2}$ (black curve) from the TVGJR–GARCH(1,1) model for the WIG20 daily returns, both from 1 April 2004 to 31 March 2015

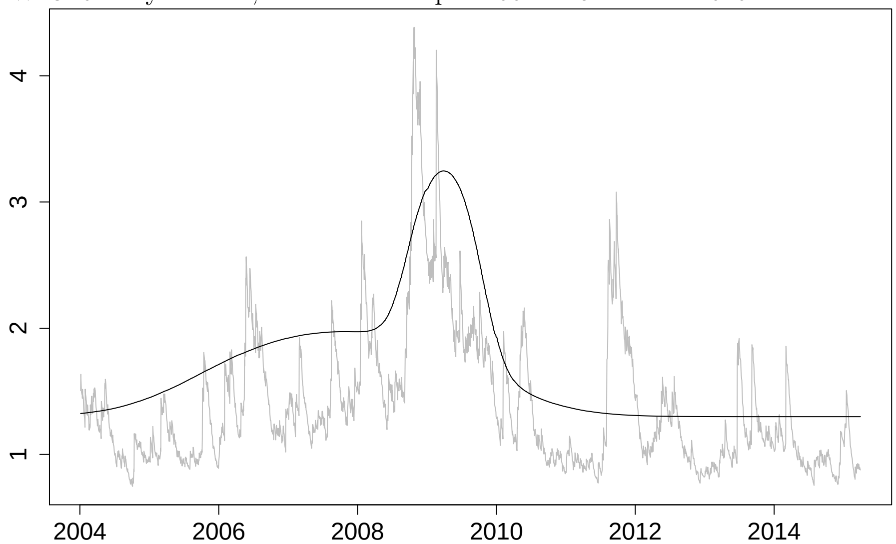


Figure 7: First 250 autocorrelations of absolute values $|\varepsilon_t|$ of the WIG20 index and the rescaled series $|\varepsilon_t|/\widehat{g}_t^{1/2}$, subsample

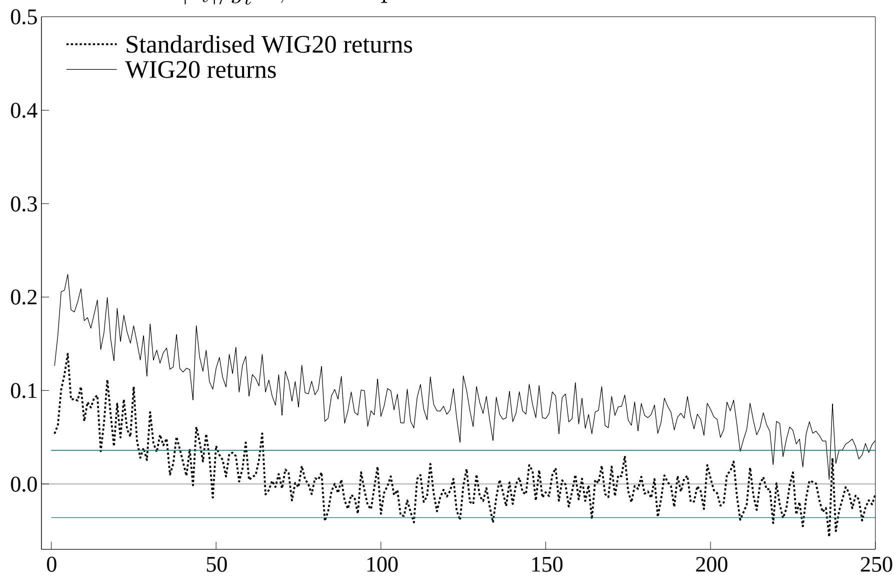
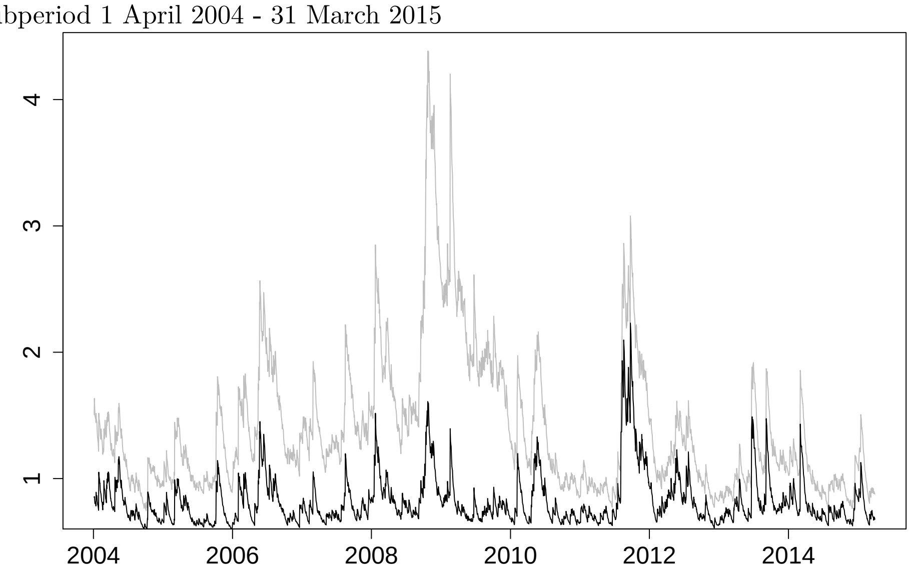


Figure 8: Conditional standard deviations of WIG20 retuns from the GJR–GARCH model (grey curve) and from the TVGJR–GARCH model (black curve), for the subperiod 1 April 2004 - 31 March 2015



evened out after rescaling.

The autocorrelations of the absolute returns and those of $\varepsilon_t/\widehat{g}_t^{1/2}$ appear in Figure 7. The former autocorrelations are positive for all 250 lags. The plateau between lags 30 and 60 in rescaled absolute returns is still visible but is much weaker than in Figure 3. The difference between the two curves is as substantial as before. To save space, results of the misspecification tests for the models are not reported here. They are rather similar to the ones in Table 4.

For comparison, we fitted two Spline-GARCH models to the subseries. Even here we selected the knots using both AIC and BIC. The results can be found in Figure 9. The spline obtained by BIC is still very parsimonious, no knots, but it now bends more than the previous one. The main reason for this is that the period containing the first hump visible in Figure 5 is not included in the shorter series. As before, the spline generated by AIC follows the conditional standard deviation from GARCH quite closely. Which one of the two choices is more appropriate when the Spline-GARCH model is put into practical use will be discussed in the next section.

Figure 9: Conditional standard deviations of WIG20 returns from the GJR–GARCH model (grey curve), exponential quadratic spline when the number of knots is determined by BIC (solid curve) and when it is done by AIC (dotted curve), for the period 1 April 2004 - 31 March 2015

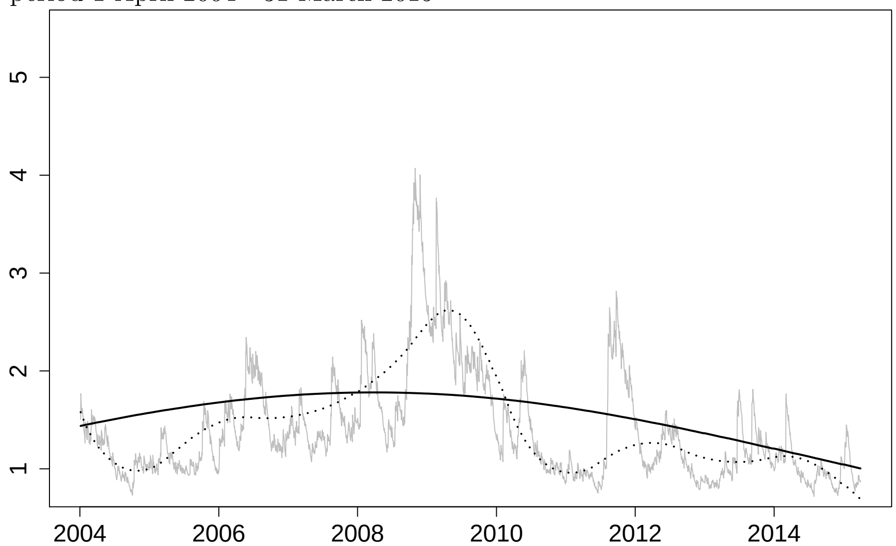


Table 5: Estimated GARCH components of the GJR–GARCH and TVGJR–GARCH model for the WIG20 log return series, 1 April 2004 to 31 March 2015. Standard deviation estimates in parentheses

	α_0	α_1	κ_1	β_1	$\alpha_1 + \kappa_1/2 + \beta_1$
GJR–GARCH	0.020 (0.007)	0.034 (0.009)	0.050 (0.016)	0.931 (0.011)	0.990
	α_0	α_1	κ_1	β_1	$\alpha_1 + \kappa_1/2 + \beta_1$
TVGJR–GARCH	0.029 (0.009)	0.005 (0.012)	0.088 (0.020)	0.908 (0.026)	0.957

9 Forecasting

9.1 Full sample

In this section we consider forecasting with the TVGJR–GARCH model and compare the forecasts with corresponding outcomes from GJR–GARCH and Spline-GARCH models. The forecasting period comprises the returns from 1 April 2015 up until 30 April 2016. In place of the unobserved volatility we use the squared daily return which, as discussed in Patton (2011), is an unbiased volatility proxy. The measures

of forecasting accuracy are thus formed with respect to this proxy. We also want to find out whether the accuracy of the forecasts depends on the estimation period. This question becomes interesting when the deterministic component of the model is not constant. For a standard GARCH or GJR–GARCH model using a subperiod instead of the whole series to estimate the parameters does not make much sense when the model is correctly specified.
For the TVGJR–GARCH model the forecasts are constructed by assuming that the value of the deterministic component does not change during the forecasting period from what it is at the end of the estimation period. The same rule is applied to Spline-GARCH, which means that the spline is not extrapolated into the forecasting period. The forecasting horizon varies from one to 120 days. In Table 6 we report the Root Mean Square Forecast Error (RMSFE) which for our volatility proxy is a robust loss function, see Patton (2011). In the same table, however, we also include results based on the Mean Absolute Forecast Error (MAFE) and the Median Squared Forecast Error (MedSFE).

Table 6: Root mean squared, mean absolute and median squared forecast errors for forecasts from various models, estimation period 2 January 1996 - 31 March 2015

	GJR–GARCH			TVGJR–GARCH		
Horizon	MedSFE	MAFE	RMSFE	MedSFE	MAFE	RMSFE
$h = 1$	1.607	1.681	2.825	0.811	1.437	2.656
$h = 5$	1.667	1.713	2.833	0.779	1.417	2.653
$h = 10$	2.021	1.787	2.872	0.840	1.433	2.679
$h = 20$	2.415	1.909	2.928	0.810	1.422	2.716
$h = 60$	3.319	2.167	3.250	0.744	1.474	2.977
$h = 90$	3.987	2.219	3.298	0.670	1.458	3.103
$h = 120$	5.006	2.278	2.570	0.650	1.353	2.261
	Spline-GARCH (AIC)			Spline-GARCH (BIC)		
Horizon	MedSFE	MAFE	RMSFE	MedSFE	MAFE	RMSFE
$h = 1$	2.549	2.706	5.215	0.629	1.133	2.039
$h = 5$	2.289	2.647	5.224	0.593	1.134	2.038
$h = 10$	2.250	2.661	5.282	0.653	1.154	2.054
$h = 20$	1.664	2.608	5.405	0.713	1.182	2.088
$h = 60$	1.127	2.764	5.998	0.700	1.228	2.267
$h = 90$	0.974	2.747	6.264	0.691	1.197	2.331
$h = 120$	0.925	2.538	4.665	0.696	1.129	1.684

The results show that Spline-GARCH (BIC) yields the most accurate forecasts by all criteria of comparison, whereas Spline-GARCH (AIC) generates the least accurate ones. The difference is due to the fact that the end-point of the spline in the latter is 'too low' when compared to the former, see Figure 5. This has a dramatic effect on the accuracy of the forecasts. Note that the final level of the spline cannot be compared to the level of the deterministic component of the TVGJR–GARCH model in Figure 2. As already discussed, the level in that model is a relative concept. What

matters is $\widehat{\alpha}_0\delta_0$ as the change in δ_0 in (5) affects the estimate of α_0 in (4). Table 7 contains the values of the out-of-sample F-test (OOS-F) for the TVGJR- and Spline-GARCH models. The benchmark is the GARCH or GJR–GARCH model. All values indicate significance at the level 0.05. The minus sign shows that the roles of the null and the alternative have changed: Spline-GARCH (AIC) is the null model and produces significantly more inaccurate forecasts than GARCH. These figures agree with the ones in Table 6.
Another way of sorting out inferior models is to construct model confidence sets (MCS), see Hansen, Lunde and Nason (2011). The results in Table 8 show that when the mean squared error is used for comparing the models, only Spline-GARCH (AIC) falls outside the confidence set when the forecasting horizon is sufficiently short. When the selection is based on MAFE, see Table 9, the distinctions are sharper, and at short horizons Spline-GARCH (BIC) is the sole member of MCS up until $h = 20$. When $h \geq 60$, the TVGJR–GARCH model also belongs to MCS.

Table 7: Values of the out-of-sample F-test for the models. Benchmark: GARCH or GJR–GARCH

	Forecasting horizon						
Model	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 60$	$h = 90$	$h = 120$
TVGJR–GARCH	35.47	37.16	38.68	40.38	40.29	23.41	52.69
Spline-GARCH (AIC)	−190.0	−187.1	−183.8	−176.4	−149.8	−131.1	−127.1
Spline-GARCH (BIC)	246.1	244.6	240.8	242.4	210.6	172.6	225.7

Table 8: The model confidence set when models are compared using the mean squared forecast error

	Horizon					
Model	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 60$	$h = 90$
GARCH	0.641*	0.500*	0.556*	0.310*	0.374*	0.881*
GJR–GARCH	0.541*	0.518*	0.458*	0.327*	0.240*	0.741*
TVGJR–GARCH	1.000*	1.000*	1.000*	1.000*	0.374*	0.780*
Spline-GARCH (BIC)	0.641*	0.518*	0.556*	0.327*	0.374*	0.879*
Spline-GARCH (AIC)	0.012	0.088	0.316*	0.446*	1.000*	1.000*

9.2 Subsample from April 2004

As already mentioned, in the light of results in Amado and Teräsvirta (2014) studying the effect of the estimation period and thus that of the deterministic component on forecasts should be quite interesting. To this end we forecast with models estimated in Section 8. Results can be found in Table 10. It can be seen that the TVGJR–GARCH model generates the most accurate forecasts. They are more accurate than

Table 9: The model confidence set when models are compared using the mean absolute forecast error

	Horizon					
Model	$h=1$	$h=5$	$h=10$	$h=20$	$h=60$	$h=90$
GARCH	0.000	0.000	0.000	0.000	0.000	0.000
GJR–GARCH	0.000	0.000	0.000	0.000	0.000	0.000
TVGJR–GARCH	0.000	0.000	0.000	0.010	1.000*	0.408*
Spline-GARCH (BIC)	1.000*	1.000*	1.000*	1.000*	0.038	1.000*
Spline-GARCH (AIC)	0.000	0.000	0.000	0.000	0.000	0.000

the corresponding ones from the full-sample model and the most accurate of all models. One can conclude that the estimation period matters. In this case, the starting 'level' for forecasting, $\widehat{\alpha}_0\delta_0$, is more favourable than in the model based on the full sample. For Spline-GARCH (BIC) the situation is the opposite, but the forecasts from this model are still far more accurate than those from Spline-GARCH (AIC). The former Spline-GARCH model is now roughly at par with GJR–GARCH. This is also seen from Table 11. The Spline-GARCH model is superior to GJR–GARCH for $h \leq 20$ but loses its edge at longer horizons. This is also obvious from results of the OOS-F test in Table 11. Forecasts from Spline-GARCH (AIC) continue to be inferior to the others and even less accurate than the ones from GJR–GARCH.

Table 10: Root mean squared, mean absolute and median squared forecast errors for forecasts from various models, estimation period 1 April 2004 - 31 March 2015

	GJR–GARCH			TVGJR–GARCH		
Horizon	MedSFE	MAFE	RMSFE	MedSFE	MAFE	RMSFE
$h=1$	1.387	1.635	2.811	0.337	0.908	1.655
$h=5$	1.423	1.654	2.817	0.350	0.898	1.655
$h=10$	1.682	1.703	2.850	0.357	0.905	1.673
$h=20$	1.824	1.776	2.890	0.361	0.900	1.694
$h=60$	2.124	1.901	3.152	0.373	0.939	1.849
$h=90$	2.313	1.870	3.205	0.333	0.936	1.930
$h=120$	2.550	1.821	2.338	0.336	0.871	1.401
	Spline-GARCH (AIC)			Spline-GARCH (BIC)		
Horizon	MedSFE	MAFE	RMSFE	MedSFE	MAFE	RMSFE
$h=1$	2.920	2.981	5.905	0.974	1.517	2.788
$h=5$	2.489	2.930	5.934	0.964	1.513	2.792
$h=10$	2.405	2.938	6.010	1.071	1.534	2.818
$h=20$	1.793	2.895	6.167	1.046	1.551	2.871
$h=60$	1.120	3.113	6.840	0.793	1.583	3.143
$h=90$	0.983	3.114	7.144	0.674	1.511	3.240
$h=120$	0.983	2.871	5.340	0.587	1.405	2.365

Table 11: Values of the out-of-sample F-test for the models. Benchmark: GARCH or GJR–GARCH

	Forecasting horizon						
Model	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 60$	$h = 90$	$h = 120$
TVGJR–GARCH	503.0	497.1	493.2	470.0	400.7	319.6	325.7
Spline-GARCH (AIC)	−208.1	−205.4	−201.8	−194.8	−165.5	−144.0	−145.4
Spline-GARCH (BIC)	4.535	4.521	4.683	4.763	0.808	−4.971	−3.669

9.3 Comparing full sample and subsample forecasts

It may be asked after seeing these forecasts is whether longer return series lead to more accurate models and volatility forecasts than shorter series. A comparison of forecasts from models based on these samples shows that there is no clear-cut answer to this question. Tables 6 and 10 show that accuracy of forecasts from the TVGJR–GARCH model increases when the model is built only on the time series starting in 2004, whereas the situation is the opposite for the Spline-GARCH (BIC) model. When the model is a GJR–GARCH one, there is hardly any difference in RMSFE between forecasts from the two variants of the model. Obviously, the parameter estimates do not change much when one moves from the subsample to the complete one, although their precision should improve.

The accuracy of forecasts from TV–GARCH and Spline-GARCH models is very dependent on the last value of the deterministic component because this value forms the starting-point for forecasting. This is why there are differences in RMSFE between the variants of the same model. This also explains why in one case estimating the model from the subsample leads to more precise forecasts than using the whole time series, whereas in another case the situation is the opposite. It is reassuring, however, that both the subsample and full sample forecasts are more accurate than the ones from the GJR–GARCH model. This suggests that using models with a multiplicative deterministic component for forecasting is a good idea, although it may not always be possible to tell in advance which multiplicative model and which observation period one should use. In the present case it seems that the Spline-GARCH (AIC) gives a deterministic component that is quite flexible. Nevertheless, and perhaps because of this property, the final value of the deterministic component becomes too low and the forecasts thereby less competitive when compared to other approaches.

10 Conclusions

In this paper we model daily logarithmic returns of the WIG20 index acknowledging the fact that the series may be nonstationarity in the sense that the amplitude of volatility clusters is not constant over time. Modelling is carried out in a systematic fashion, which is emphasised in the paper. The form of the model is specified first, the parameters of the fully specified model estimated thereafter and, finally, the estimated model is subjected to misspecification tests. This is done both using the whole sample from the beginning of 1995 and a subsample in which the observation period starts 1 April 2004.

Forecasting performance of the TVGJR–GARCH model is compared with that of two variants of the Spline-GARCH model. It turns out that the most accurate forecasts of volatility are obtained using the TVGJR–GARCH model fitted to the subperiod. The conclusion is that the length of the observation period matters, and that, measured by the root mean squared error, models built on the longest series do not automatically provide the best forecasts. This accords with findings reported in Amado and Teräsvirta (2014).

A general conclusion is that when the return series to be modelled are sufficiently long, the deterministic component in the variance cannot be ignored. It may, at least in theory, be replaced or completed by a stochastic component, although finding economic variables that would explain variation in daily return series does not seem to be easy. Time used in this work is a proxy for the factors and phenomena that are moving daily equity prices but may be remarkably difficult to quantify.

Acknowledgements

This research has been supported by Center for Research in Econometric Analysis of Time Series (CREATES), funded by the Danish National Research Foundation, Grant No. DNFR 78. The first author also acknowledges funding from COMPETE (Ref. No. POCI-01-0145-FEDER-006683), with the FCT/MEC's (Fundação para a Ciência e a Tecnologia, I.P.) financial support through national funding and by the ERDF through the Operational Programme on "Competitiveness and Internationalization" – COMPETE 2020 under the PT2020 Partnership Agreement. Part of the work for this paper was carried out when the first and the third author were visiting School of Economics and Finance of Queensland University of Technology, Brisbane, whose kind hospitality is gratefully acknowledged. We wish to thank Dobromil Serwa for providing the data and José Rangel for the Spline-GARCH code. Material from this paper has been presented at the Second International Workshop on Financial Markets and Nonlinear Dynamics, Paris, and the Workshop on Forecasting Economic Time Series: In Honour of Antoni Espasa, Carlos III de Madrid, both June 2015, the Second International Workshop in Financial Econometrics, Salvador, October 2015, Advances in Empirical and Theoretical Econometrics, November 2015, 9th International Conference on Computational and Financial Econometrics, University of London, December 2015, the 24th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, University of Alabama, Tuscaloosa, AL, and Oxmetrics Users' Meeting, George Washington University, Washington, DC, both March 2016, and in seminars at City University of Hong Kong, Hong Kong University, Monash

University, University of Helsinki and University of Tasmania, Hobart. Comments from participants in these occasions are gratefully acknowledged. We would also like to thank an anonymous referee for helpful remarks. Any errors and shortcomings in this work are the authors' responsibility.

References

[1] Akaike H. (1974), A new look at statistical model identification, *IEEE Transactions on Automatic Control* AC–19, 716–723.

[2] Amado C. and Teräsvirta T. (2008), Modelling conditional and unconditional heteroskedasticity with smoothly time-varying structure, *SSE/EFI Working Paper Series in Economics and Finance* 691, Stockholm School of Economics.

[3] Amado C. and Teräsvirta T. (2013), Modelling volatility by variance decomposition, *Journal of Econometrics* 175, 153–165.

[4] Amado C. and Teräsvirta T. (2014), Modelling changes in the unconditional variance of long stock return series, *Journal of Empirical Finance* 25, 15–35.

[5] Amado C. and Teräsvirta T. (2017), Specification and testing of multiplicative time-varying GARCH models with applications, *Econometric Reviews* 36, 421–446.

[6] Baillie R. T., Bollerslev T. and Mikkelsen H. O. (1996), Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 74, 3–30.

[7] Bollerslev T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307–327.

[8] Brdyś M. A., Borowa A., Idźkowiak P. and Brdyś M. T.: 2009, Adaptive prediction of stock exchange indices by state space wavelet networks, *International Journal of Applied Mathematics and Computer Science* 19, 337–348.

[9] Brownlees C. T. and Gallo G. M. (2010), Comparison of volatility measures: A risk management perspective, *Journal of Financial Econometrics* 8, 29–56.

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

[11] Czapkiewicz A. and Basiura B. (2014), The position of the WIG index in comparison with selected market indices in boom and bust periods, *Statistics in Transition* 15, 427–436.

[12] Davidson J. (2004), Moment and memory properties of linear conditional heteroscedasticity models, and a new model, *Journal of Business and Economic Statistics* 22, 16–29.

[13] Diebold F. X. (1986), Modeling the persistence of conditional variances: A comment, *Econometric Reviews* 5, 51–56.

[14] Ekner L. E. and Nejstgaard E. (2013), Parameter identification in the logistic STAR model, *Discussion Paper 13–07*, Department of Economics, University of Copenhagen.

[15] Engle R. F. and Rangel J. G. (2008), The spline-GARCH model for low-frequency volatility and its global macroeconomic causes, *Review of Financial Studies* 21, 1187–1222.

[16] Feng Y. (2004), Simultaneously modeling conditional heteroskedasticity and scale change, *Econometric Theory* 20, 563–596.

[17] Glosten L. W., Jagannathan R. and Runkle D. E. (1993), On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.

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

[19] Hagerud G. E. (1997), A new non-linear GARCH model, EFI Economic Research Institute, Stockholm.

[20] Hall A. D., Silvennoinen A. and Teräsvirta T. (2017), Building multiplicative time-varying smooth transition conditional correlation GARCH models, *work in progress*, School of Economics and Finance, Queensland University of Technology.

[21] Hansen P. R., Lunde A. and Nason J. M. (2011), The model confidence set, *Econometrica* 79, 453–497.

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

[23] Lamoureux C. G. and Lastrapes W. G. (1990), Persistence in variance, structural change and the GARCH model, *Journal of Business and Economic Statistics* 8, 225–234.

[24] Lundbergh S. and Teräsvirta T. (2002), Evaluating GARCH models, *Journal of Econometrics* 110, 417–435.

[25] Makiel K. (2012), ARIMA–GARCH models in estimating market risk using value at risk for WIG20 index, *Financial Internet Quarterly "e-Finance"* 8, 25–33.

[26] Malecka M. (2013), GARCH process application in risk valuation for WIG20 index, *Acta Universitatis Lodziensis Folia Oeconomica* 285, 209–220.

[27] Mazur B. and Pipień M. (2012), On the empirical importance of periodicity in the volatility of financial returns - time varying GARCH as a second order APC(2) process, *Central European Journal of Economic Modelling and Econometrics* 4, 95–116.

[28] Patton A. J. (2011), Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics* 160, 246–256.

[29] Rissanen J. (1978), Modeling by shortest data description, *Automatica* 14, 465–471.

[30] Schwarz G. (1978), Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.

[31] Silvennoinen A. and Teräsvirta T. (2016), Testing constancy of unconditional variance in volatility models by misspecification and specification tests, *Studies in Nonlinear Dynamics and Econometrics* 20, 347–364.

[32] Song P. X., Fan Y. and Kalbfleisch J. D. (2005), Maximization by parts in likelihood inference, *Journal of the American Statistical Association* 100, 1145–1158.

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

[34] van Bellegem S. and von Sachs R. (2004), Forecasting economic time series with unconditional time-varying variance, *International Journal of Forecasting* 20, 611–627.

[35] Wooldridge J. M. (1991), On the application of robust, regression-based diagnostics to models of conditional mean and conditional variances, *Journal of Econometrics* 47, 5–46.

A Estimated GARCH equations in the Spline-GARCH model

The estimated GARCH equations of the Spline-GARCH model when the estimation period consists of the whole sample from 2 January 1996 to 31 March 2015 ($T = 4777$) are as follows. The equation of the AIC-based model equals

$$\widehat{h}_t = \underset{(-)}{0.033} + \underset{(0.005)}{0.073}\phi_{t-1}^2 + \underset{(0.008)}{0.894}h_{t-1}$$

so the persistence, that is, $\widehat{\alpha}_1 + \widehat{\beta}_1 = 0.967$. When the number of equidistant knots is selected using BIC as in Engle and Rangel (2008), the equation is

$$\widehat{h}_t = \underset{(-)}{0.014} + \underset{(0.004)}{0.072}\phi_{t-1}^2 + \underset{(0.005)}{0.913}h_{t-1}$$

where $\widehat{\alpha}_1 + \widehat{\beta}_1 = 0.986$.

For the subperiod from 2 January 2004 to 31 March 2015 ($T = 2808$) the equation for the AIC-based model is

$$\widehat{h}_t = \underset{(-)}{0.034} + \underset{(0.008)}{0.057}\phi_{t-1}^2 + \underset{(0.013)}{0.909}h_{t-1} \tag{10}$$

where $\widehat{\alpha}_1 + \widehat{\beta}_1 = 0.966$. For the BIC-selected splines,

$$\widehat{h}_t = \underset{(-)}{0.010} + \underset{(0.007)}{0.064}\phi_{t-1}^2 + \underset{(0.008)}{0.926}h_{t-1} \tag{11}$$

where $\widehat{\alpha}_1 + \widehat{\beta}_1 = 0.990$. Even here, the persistence is clearly lower in (10) than in (11).