

Testing for Linear Autoregressive Dynamics under Heteroskedasticity

Christian M. Hafner, Helmut Herwartz *

November 1998

Abstract

One puzzling behavior of asset returns for various frequencies is the often observed positive autocorrelation at lag 1. To some extent this can be explained by standard asset pricing models when assuming time varying risk premia. However, one often finds better results when directly fitting an autoregressive model, for which there is little economic foundation. One may ask whether the underlying process does in fact contain an autoregressive component. It is therefore of interest to have a statistical test at hand that performs well under the stylized facts of financial returns. In this paper, we investigate empirical properties of competing devices to test for autoregressive dynamics in case of heteroskedastic errors. For the volatility process we assume GARCH, TGARCH and stochastic volatility. The results indicate that standard QML inference for the autoregressive parameter is negatively affected by misspecification of the volatility process. We show that bootstrapped versions of a likelihood ratio and White's t-statistic have better size properties and comparable power properties. Applied to German stock data, the alternative tests in many cases yield very different p-values.

Keywords: *autoregression, heteroskedasticity, bootstrap, GARCH, stochastic volatility*

*SFB 373 and Institut für Statistik und Ökonometrie, Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität zu Berlin, Spandauer Str. 1, D-10178 Berlin, Germany. e-mail hafner@wiwi.hu-berlin.de and helmut@wiwi.hu-berlin.de. Financial support by the Deutsche Forschungsgemeinschaft is gratefully acknowledged. Both authors would like to thank Michael Neumann for helpful remarks and suggestions.

1 Introduction

Standard models of finance theory assume that financial returns have two components, a risk premium plus noise. If the risk premium is assumed to be time-varying, as e.g. in Engle, Lilien and Robins (1987), then returns will in general exhibit some autocorrelation. An example is the capital asset pricing model (CAPM) with time-varying covariances as investigated by Bollerslev, Engle and Wooldridge (1988) and Hafner and Herwartz (1998). A competing device is to directly model the autocorrelation structure by using autoregressive models. In general, there is no economic interpretation of such models because the conditional mean may become negative and thus fails the interpretation of a risk premium. As empirical practice has shown, however, the fit of autoregressive models very often outperforms the one of risk premium models. This may be due to the strong autocorrelation reported for stock returns on various frequencies, see e.g. Chapter 2 of Campbell, Lo and MacKinlay (1997).

As there may be doubts about the prevalence of autoregressive components in the mean, there is no doubt about the observation that volatility is time-varying for financial assets. The predominant models in this domain are the GARCH (Engle, 1982 and Bollerslev, 1986) and stochastic volatility (Taylor, 1986) models, which will serve as the volatility generating processes throughout our paper.

The R^2 of autoregressions for return data is typically small such that the inclusion or exclusion of autoregressive dynamics has only a minor effect on parameter estimates of the volatility model such as GARCH. However, the volatility process itself may be more affected by the choice of linear dynamics because relative to noisy returns an autoregressive specification of the return series yields a different estimate of the history of the volatility process. Thus, whether or not to have an autoregressive component is particularly relevant for applications where the estimated volatility is used, for example option pricing.

In this paper, we examine the empirical properties of competing devices testing for significant autoregressive dynamics of order one in conditionally heteroskedastic time series models. The quasi maximum likelihood (QML) approach as introduced in the GARCH-framework by Bollerslev and Wooldridge (1992) is compared with inference techniques based on least squares (LS) procedures which cope with conditional heteroskedasticity. The correction of standard t -ratios given by White (1980) and a pseudo likelihood ratio (LR) statistic are investigated. Additionally, we use the so-called wild bootstrap introduced by Wu (1986) to obtain bootstrap replications of the White and likelihood ratio statistic. Whereas for QML inference a parametric description of the volatility process is necessary, the latter devices concentrate only on the specification of linear dynamics. We show that the bootstrap tests show in many cases superior size properties and that QML inference is negatively affected by volatility misspecification. In particular, assuming the popular GARCH(1,1) model when the true process is stochastic volatility yields severe size distortions and, hence,

invalid QML inference. If the true process is of the SV type, our bootstrap tests are still valid but lose power especially for the case of very high persistence.

The test statistics were applied to German stock returns, giving in many cases different decisions about acceptance or rejection of the null hypothesis of no autocorrelation.

The remainder of the paper is organized as follows: In Section 2 the time series framework of our analysis is provided and competing devices for inference against an uncorrelated series are encountered. A simulation study which is given in Section 3 sheds light on the empirical properties of the different inference procedures. Section 4 provides an application to German stock data. In Section 5 we give a brief summary of our results and conclude.

2 Testing for AR(1) dynamics in case of heteroskedastic error terms

An autoregressive time series model of order one, AR(1), with conditionally heteroskedastic error is conveniently specified as

$$y_t = c + \phi y_{t-1} + e_t, \quad e_t | \mathcal{F}_{t-1} \sim N(0, h_t), \quad t = 1, 2, \dots, T, \quad (1)$$

where \mathcal{F}_{t-1} denotes the information set generated by the process y_t up to and including time $t - 1$. Complementary to autoregressive dynamics c captures deterministic influences governing the evolution of y_t . Within the so-called autoregressive conditionally heteroskedastic model of order q , ARCH(q), introduced by Engle (1982) h_t is assumed to be completely determined by a time invariant component ω and innovations obtained up to time $t - 1$, i.e.

$$h_t = \omega + \alpha_1 e_{t-1}^2 + \dots + \alpha_q e_{t-q}^2. \quad (2)$$

Bollerslev (1986) generalized (2) introducing autoregressive dynamics of h_t defining the GARCH(p, q) model:

$$h_t = \omega + \alpha_1 e_{t-1}^2 + \dots + \alpha_q e_{t-q}^2 + \beta_1 h_{t-1} + \dots + \beta_p h_{t-p}. \quad (3)$$

In applied work the GARCH(1,1) model turned out to be particularly useful to describe a wide variety of financial market data (see e.g. Bollerslev, Engle and Nelson, 1994). Therefore we concentrate in the following on this parsimonious specification of conditional heteroskedasticity. As given above the GARCH(1,1) model is characterized by a symmetric response of current volatility to positive and negative lagged innovations e_{t-1} . From the empirical literature on the analysis of stock market returns the so-called leverage effect might be seen as a stylized fact of return volatility (see e.g. Black, 1976). To allow for different impacts of lagged positive and negative innovations one may favour a so-called threshold GARCH (TGARCH) model which was proposed Glosten,

Jagannathan and Runkle (1993) for the variance and by Zakoian (1994) for the standard deviation. The TGARCH(1,1) model for h_t takes the following form:

$$h_t = \omega + \alpha_1 e_{t-1}^2 + \alpha_1^- e_{t-1}^2 I_{e_{t-1} < 0} + \beta_1 h_{t-1}. \quad (4)$$

In (4), $I_{(\cdot)}$ denotes an indicator function. The leverage effect describes that current volatility is more affected by negative lagged innovations relative to positive ones. This characteristic is obtained for $\alpha_1^- > 0$. Many extensions of the standard GARCH models have been proposed to capture the leverage effect, the most popular being the exponential GARCH (EGARCH) model of Nelson (1991). We refrain from using the EGARCH model for two reasons: first, there are only few results on the asymptotics of QML estimates, and second, when applied to stock market data many studies showed that EGARCH tends to overestimate the impact of outliers on volatility, see e.g. Engle and Ng (1993).

Error sequences following a GARCH volatility process complicate the specification of the log-likelihood function of a time series model like (1). Whereas the QML estimation of the parameters of the variance model (4) always requires nonlinear optimization routines the parameters governing linear dynamics may still be estimated using standard LS procedures. However, standard inference e.g. against significance of estimated autoregressive parameters in (1) is no longer valid.

Heteroskedasticity consistent inference by means of t -ratios requires the computation of second and first order derivatives of the log-likelihood function with respect to the autoregressive parameter of interest (see e.g. Bollerslev and Wooldridge, 1992). A simultaneous analysis of linear and higher order dynamics might be cumbersome in applied work for at least three reasons: First, the performance of the nonlinear optimization may be negatively affected by large dimensions of the parameter space. In addition, increasing the number of parameters may reduce the power of inference techniques. Second, inference on linear dynamics within such a framework always depends on the specification of the variance process. Since there is a variety of possible variance specifications there is also a variety e.g. of t -ratios that may be used to indicate significance of model parameters governing linear dynamics. A third disadvantage of QML inference which is related to the argument given before is its lack of robustness. Inference on autocorrelation should be valid for alternative types of heteroskedasticity especially if the assumed volatility model, GARCH say, amounts to a misspecification of the volatility process.

Due to the difficulties listed above standard LS techniques are often used to infer on linear dynamics of financial time series. In a second step of the analysis the interest concentrates on the specification of the variance process of residuals obtained from first step residuals. LS estimation bears the advantage of having a unique solution which is straightforward to compute. However, inference along standard lines ignoring the potential of heteroskedasticity involves invalid empirical levels of tests derived under i.i.d. assumptions. White (1980) introduced a correction of standard t -ratios in case of heteroskedastic error terms

which is easily implemented. In addition, bootstrap procedures designed for heteroskedastic innovations may be seen as a reasonable framework to retain the convenience of LS procedures.

In the following, we outline alternative inference procedures against significance of the autoregressive parameter in heteroskedastic time series models such as (1).

2.1 QML inference

Iterative QML estimation of ϕ requires a parametric description of the volatility process. Let θ denote the vector of parameters describing a time series model like (1). In case of a GARCH(1,1) model one has $\theta = (c, \phi, \omega, \alpha_1, \beta_1)'$. The log-likelihood function conditional on a starting value y_0 is obtained as:

$$l(\theta) = \sum_{t=1}^T l_t(\theta) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(h_t) - \frac{1}{2} \sum_{t=1}^T \frac{(y_t - c - \phi y_{t-1})^2}{h_t}.$$

The parameter vector θ is conveniently estimated by nonlinear optimization routines, e.g. using the BHHH-algorithm (see e.g. Judge et al., 1988). Under some regularity conditions the estimator converges at rate \sqrt{T} and is asymptotically normally distributed:

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, D^{-1}SD^{-1}). \quad (5)$$

In (5) S is the expectation of the cross-product of first order derivatives of $l_t(\theta)$ with respect to θ and D is the negative expectation of the matrix of second order derivatives. In case of normally distributed errors one has $D = S$ and the asymptotically valid covariance matrix is

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, S^{-1}). \quad (6)$$

For the symmetric GARCH(1,1) model both matrices are known to be block-diagonal, such that inference on significance of $\hat{\phi}$ involves only the upper left elements of D and S in the case of nonnormal errors and of S alone in case of normally distributed errors. For the autoregressive model of order one the relevant elements of D and S can be estimated as follows using the QML estimate $\hat{\theta}$ (see e.g. Bollerslev, 1986):

$$\hat{S}_\phi = \frac{1}{T} \sum_{t=1}^T \left\{ \frac{e_t y_{t-1}}{h_t} + \frac{1}{2h_t} \frac{\partial h_t}{\partial \phi} \left(\frac{e_t^2}{h_t} - 1 \right) \right\}^2, \quad (7)$$

$$\begin{aligned} \hat{D}_\phi &= \frac{1}{T} \sum_{t=1}^T \frac{y_{t-1}^2}{h_t} + \frac{1}{2h_t^2} \left(\frac{\partial h_t}{\partial \phi} \right)^2 \\ &+ \frac{2e_t y_{t-1}}{h_t^2} \frac{\partial h_t}{\partial \phi} - \left(\frac{e_t^2}{h_t} - 1 \right) \frac{\partial}{\partial \phi} \left\{ \frac{1}{2h_t} \frac{\partial h_t}{\partial \phi} \right\}. \end{aligned} \quad (8)$$

Note that for the GARCH(1,1) model one has:

$$\frac{\partial h_t}{\partial \phi} = -2\alpha_1 y_{t-2} e_{t-1} + \beta_1 \frac{\partial h_{t-1}}{\partial \phi}.$$

2.2 Inference based on least squares

Ignoring the heteroskedasticity of model errors in (1) one may compute along standard lines the t -ratio of the LS estimate of ϕ ,

$$t = (\hat{\sigma}^2 \sum_{t=2}^T \bar{y}_{t-1}^2)^{-1/2} \left(\sum_{t=2}^T \bar{y}_{t-1} \bar{y}_t \right),$$

where $\bar{y}_t = y_t - T^{-1} \sum_{t=1}^T y_t$ and $\hat{\sigma}^2$ is obtained from LS residuals \hat{e}_t as $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{e}_t^2$. In presence of (conditionally) heteroskedastic error terms the standard t -ratio loses its asymptotic $N(0,1)$ -distribution and fails to be pivotal.

To infer against linear dependence in higher order autoregressive models it may be more convenient to use a likelihood ratio type statistic (see e.g. Lütkepohl, 1991). Essentially this statistic compares the log-ratio (times T) of sums of squared residuals of the model under the null hypothesis (RSS_0) and under the alternative (RSS_1). For the AR(1)-model it can be shown that assuming independent and identically distributed error terms the likelihood ratio statistic is almost equal to the squared t -statistic given in (2.2). In presence of heteroskedasticity, however, the true likelihood ratio cannot be computed from sums of unweighted squared LS residuals. Thus we refer to this statistic in the following as the pseudo likelihood ratio statistic (PLR):

$$Q_{PLR} = T \log \left(\frac{RSS_0}{RSS_1} \right). \quad (9)$$

Note that the ratio RSS_0/RSS_1 is easily augmented to obtain a (pseudo) F-statistic.

To derive the asymptotic variance of the LS estimator in the presence of heteroskedasticity knowledge of the underlying true variances σ_t^2 is necessary. From standard theory on generalized linear models asymptotic normality can be derived for the modified t -statistic:

$$\tilde{t} = \left(\sum_{t=2}^T \sigma_t^2 \bar{y}_{t-1}^2 \right)^{-1/2} \left(\sum_{t=2}^T \bar{y}_{t-1} \bar{y}_t \right).$$

In practice, however, σ_t^2 is unknown. White (1980) proposed to replace this unknown quantity by squared residuals obtained from consistent estimation of the linear model in (1):

$$t_{WH} = \left(\sum_{t=2}^T \hat{e}_t^2 \bar{y}_{t-1}^2 \right)^{-1/2} \left(\sum_{t=2}^T \bar{y}_{t-1} \bar{y}_t \right).$$

Sufficient conditions for t_{WH} to follow a standard normal distribution asymptotically are provided by White (1980) and in Chapter 6 of White (1984). Essentially these conditions concern the existence of fourth moments of y_t which for our model (1) means that $E[e_t^4]$ is finite and $|\phi| < 1$. For GARCH-type processes the existence of unconditional fourth order moments depends on the particular parameterization of the process. Since the probability limit of \hat{e}_t^2 is e_t^2 and by definition $E[e_t^2 | \mathcal{F}_{t-1}] = \sigma_t^2$ the construction of t_{WH} still appears to be reasonable irrespective of the existence of fourth order moments.

2.3 Inference using bootstrap methods

Except for White's corrected t -ratio the asymptotic distribution of all LS based test statistics given above is at least cumbersome to derive analytically in the presence of heteroskedastic error terms. In this case bootstrap methods become a convenient means to estimate the distribution e.g. of the pseudo LR statistic if the resampling scheme allows for heteroskedasticity. Wu (1986) introduced the wild bootstrap coping with heteroskedastic error distributions. This procedure is advocated by Mammen (1993) to estimate the distribution of F -type statistics in parametric regression models with random explanatory variables under heteroskedasticity. Adopting a nonparametric framework, Neumann and Kreiss (1998) show that the validity of regression type bootstrap procedures is often maintained for autoregressive models if $y_t - E[y_t | \mathcal{F}_{t-1}]$ follows a martingale difference. For the application we have in mind here it can be shown that bootstrap inference in parametric autoregressive models is analogous to the regression case as discussed in Mammen (1993). In Herwartz (1998) the wild bootstrap procedure is used to mimic the stochastic behaviour of the pseudo LR statistic in the framework of so-called periodic autoregressive time series models.

Since a resampling scheme is easily implemented under the null hypothesis of uncorrelated observations y_t , bootstrap inference is regarded as a promising alternative to nonlinear estimation and inference techniques designed for heteroskedastic error distributions. For the present investigation we use the wild bootstrap to estimate the distribution of the pseudo LR statistic. It is known from asymptotic theory on bootstrap procedures that with respect to interval estimation the bootstrap procedure is especially fruitful if the simulated quantity is (asymptotically) pivotal (see e.g. Hall, 1992). Thus, we also evaluate the distribution of White's corrected t -ratio by means of bootstrap techniques.

An appropriate resampling procedure is given as follows:

1. LS estimation of (1) provides test statistics Q_{PLR} and the absolute value of t_{WH} , i.e. $|t_{WH}|$. Under the null hypothesis $H_0 : \phi = 0$ the model reads compactly as:

$$y = c_0 + e_0. \tag{10}$$

In (10), the T -dimensional vector y collects the available observations on

y_t , c_0 is a constant vector and e_0 contains the error terms implied by the null hypothesis. LS provides estimates of c_0 and e_0 denoted as \hat{c}_0 and \hat{e}_0 , respectively.

2. The residuals obtained under the null hypothesis are used to generate a wild bootstrap sample as follows:

$$y^* = \hat{c}_0 + e^*.$$

The elements e_t^* of e^* are obtained from residuals \hat{e}_{0t} by mimicing their low order moments. For the present analysis two procedures generating e_t^* are considered.

First, imagine for each $t = 1, \dots, T$ a random variable Z and a distribution F_t satisfying

$$\begin{aligned} E(Z|F_t) &= 0, \\ E(Z^2|F_t) &= \hat{e}_{0t}^2, \\ E(Z^3|F_t) &= \hat{e}_{0t}^3. \end{aligned}$$

A convenient way to construct F_t is to use a two point distribution such that

$$\begin{aligned} \text{prob} \left(Z = (1 - \sqrt{5}) \frac{\hat{e}_{0t}}{2} \right) &= \frac{\sqrt{5} + 1}{2\sqrt{5}}, \\ \text{prob} \left(Z = (1 + \sqrt{5}) \frac{\hat{e}_{0t}}{2} \right) &= 1 - \frac{\sqrt{5} + 1}{2\sqrt{5}}. \end{aligned}$$

The distribution F_t may serve to sample e_t^* , the low order moments of which reproduce those of the estimated error sequence. Wild bootstrap sampling as outlined above bears the advantage of mimicing the moments of \hat{e}_{0t} up to order 3. To obtain e_t^* , however, estimated first step residuals are scaled down by a factor $(1 - \sqrt{5})/2$ with a probability of about 0.72. This suggests that the estimated conditional bootstrap variance may in these cases underestimate the true conditional variance.

A symmetric version of the wild bootstrap is easily implemented by sampling e_t^* directly from a normal distribution with mean zero and time dependent variance, i.e.

$$e_t^* \sim N(0, \hat{e}_{0t}^2). \tag{11}$$

In the following we refer to the sampling scheme in (11) as the symmetric wild bootstrap. Another possible scheme using the normal distribution is to generate standard normal and independent random variables U_t and V_t and construct $Z_t = U_t/\sqrt{2} + (V_t^2 - 1)/2$, thus having $E[Z_t] = 0$, $E[Z_t^2] = 1$ and $E[Z_t^3] = 1$. Multiplying with \hat{e}_{0t} , one obtains the accordance of the

first three moments, as above for the two-point distribution. We tried both bootstrap procedures based on normal distributions, but no major difference was detected for any of the investigated processes. Therefore we only document the results for the symmetric bootstrap given in (11) and for the two-point distribution.

3. From the generated series y_t^* it is easy to obtain bootstrap estimates of the pseudo LR statistic and the corrected t -ratio which are denoted as Q_{PLR}^* and $|t_{WH}^*|$ respectively.
4. Steps (2) and (3) are performed R times with R sufficiently large. For each bootstrap sample the statistics of interest are recorded. For the experiments discussed in the next section $R = 1000$ was used.
5. The hypothesis $H_0 : \phi = 0$ is rejected with significance level α if Q_{PLR} ($|t_{WH}|$) exceeds the $(1 - \alpha)$ quantile of Q_{PLR}^* ($|t_{WH}^*|$).

Note that in principle one may also adopt QML based bootstrapping schemes. In this case one would first specify and estimate a volatility process, a GARCH(1,1) say, and obtain standardized residuals $\hat{\xi}_t = \hat{e}_{0t} / \sqrt{\hat{h}_t}$. Drawing innovations with replacement from $\hat{\xi}_t$ one may easily build up a bootstrap sample of y_t using the specification of the dynamic model under the respective null hypothesis. Whereas the wild bootstrap samples by construction show the same clustering of volatility as the estimated error sequence a model based bootstrap bears the advantage that different patterns of volatility clustering are generated for a given sample. Since bootstrap QML inference again requires the estimation of a volatility model we did not follow these lines to discuss robust inference procedures.

3 A Simulation Study

To provide insight into the empirical performance of the alternative inference procedures discussed above we performed a Monte Carlo investigation.

3.1 The Employed Time Series Models

To characterize empirical size properties we generated white noise sequences

$$y_t = \sqrt{h_t} \xi_t, \quad \xi_t \sim \text{i.i.d. } N(0, 1)$$

and to estimate the empirical power of competing procedures we used

$$y_t = 0.1y_{t-1} + \sqrt{h_t} \xi_t.$$

The parameter $\phi = 0.1$ was selected for two reasons. First, one may expect only weak autoregressive dependence in real return data, and second, stronger dependence may be trivial to detect by statistical inference.

For the volatility process we concentrate on two basic specifications: on the one hand we have the GARCH model class, on the other we have stochastic volatility (SV). In particular, we used the following specifications for the GARCH model class, letting $e_t = \sqrt{h_t}\xi_t$:

I. Symmetric GARCH(1,1) models

$$\text{GARCH1: } h_t = 0.1 + 0.05e_{t-1}^2 + 0.9h_{t-1},$$

$$\text{GARCH2: } h_t = 0.1 + 0.09e_{t-1}^2 + 0.9h_{t-1},$$

$$\text{GARCH3: } h_t = 0.1 + 0.1e_{t-1}^2 + 0.9h_{t-1},$$

II. Threshold GARCH(1,1) models

$$\text{GARCH4: } h_t = 0.1 + 0.05e_{t-1}^2 + 0.08e_{t-1}^2 I_{e_{t-1} < 0} + 0.9h_{t-1},$$

$$\text{GARCH5: } h_t = 0.1 + 0.18e_{t-1}^2 I_{e_{t-1} < 0} + 0.9h_{t-1},$$

$$\text{GARCH6: } h_t = 0.1 + 0.05e_{t-1}^2 + 0.1e_{t-1}^2 I_{e_{t-1} < 0} + 0.9h_{t-1},$$

$$\text{GARCH7: } h_t = 0.1 + 0.2e_{t-1}^2 I_{e_{t-1} < 0} + 0.9h_{t-1}.$$

The deterministic components of the volatility processes ($\omega = 0.1$) were used only to generate the volatility path. After the generation step each of the obtained sequences $e_t = \sqrt{h_t}\xi_t$ was standardized to have a sample variance of 1. The employed volatility models cover a wide range of scenarios. All volatility models are characterized by a high degree of persistence, measured in the symmetric GARCH case as $\alpha + \beta$ and in the TGARCH case, assuming a symmetric error distribution, as $\alpha_1 + \alpha_1^- / 2 + \beta$. If these degrees are smaller than one, the unconditional variance is finite. If they are equal to one, we have the important case of a model integrated in variance. In this case the unconditional variance of e_t does not exist. We deal with three processes which are integrated in variance: GARCH3, which is the integrated GARCH (IGARCH) model of Engle and Bollerslev (1986), GARCH6 and GARCH7, which may be regarded as integrated TGARCH models. For the TGARCH models, the asymmetric impact of lagged innovations on current volatility was chosen relatively small which is in accordance with many results in the empirical literature. From Bollerslev (1986) it is known that under normality the unconditional fourth moment of a GARCH sequence exists if and only if

$$\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 < 1$$

holds. For the TGARCH models with $\alpha_1^- \neq 0$, the condition extends to

$$\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 + \frac{3}{2}(\alpha_1^-)^2 + 3\alpha_1\alpha_1^- + \alpha_1^- \beta < 1$$

which is a consequence of Theorem 1 of He and Teräsvirta (1998). For the processes considered here, it can be verified that the unconditional fourth moment of e_t exists only in the cases GARCH1 and GARCH2. It may be interesting to investigate how the non-existence of fourth or even second moments, respectively, affects the performance of the test statistics applied to the integrated and threshold GARCH models.

For SV models, we consider AR(1) specifications for log volatility, as it is standard in the literature since Taylor (1986), i.e.

$$\log h_t = \omega + \alpha \log h_{t-1} + \vartheta \eta_t. \quad (12)$$

In (12), ω , α , and θ are fixed parameters, and η_t is independent standard normally distributed and independent of ξ_t . Since for SV models we have two random sources, one needs to determine the noise to noise ratio. The volatility h_t is lognormal with mean $\mu = E[h_t] = \exp(\omega/(1-\alpha) + 0.5\theta^2/(1-\alpha^2))$. To generate sequences of errors with stochastic volatility, μ was chosen alternatively $\mu = 0.5, 1$ and 2 , thus giving three different levels of the volatility process while leaving the unit variance of ξ_t unchanged. For the persistence parameter α values close to one are frequently reported for financial time series, so we choose 0.95 and 0.99 . Since we have two remaining parameters, ω and ϑ , we fix one, i.e. $\omega = -0.1$, and solve for ϑ . This gives us the following processes SV1 to SV6.

I. $\alpha = 0.95$

$$\text{SV1: } \log h_t = -0.1 + 0.95 \log h_{t-1} + 0.624\eta_t,$$

$$\text{SV2: } \log h_t = -0.1 + 0.95 \log h_{t-1} + 0.725\eta_t,$$

$$\text{SV3: } \log h_t = -0.1 + 0.95 \log h_{t-1} + 0.505\eta_t,$$

II. $\alpha = 0.99$

$$\text{SV4: } \log h_t = -0.1 + 0.99 \log h_{t-1} + 0.631\eta_t,$$

$$\text{SV5: } \log h_t = -0.1 + 0.99 \log h_{t-1} + 0.652\eta_t,$$

$$\text{SV6: } \log h_t = -0.1 + 0.99 \log h_{t-1} + 0.609\eta_t.$$

As above $\omega = -0.1$ was only used to generate the volatility path. Each obtained error sequence e_t was standardized to have a sample variance of unity.

All processes under study were generated using initial values of $e_{-50} = y_{-50} = 0$ and $\sigma_{-50}^2 = 1$. Observations running from $t = -50$ to $t = 0$ were deleted from each sample to have the empirical model under study less affected by the initial conditions.

3.2 Competing Test Procedures

To test the hypothesis $H_0 : \phi = 0$ for the simulated processes we used alternatively the standard t -ratio obtained from LS estimation and the heteroskedasticity consistent t -ratio proposed by White (1980). The asymmetric and symmetric version of wild bootstrap inference was employed to estimate the distribution of the pseudo LR statistic and White's t -ratio. In order to get insights into the effects of volatility misspecification, we performed QML inference in all cases under the assumption that the e_t sequence follows a GARCH(1,1) model. Note that this specification does not match the true volatility process for the threshold GARCH models and the SV models under study. Each of the employed time series models was generated 2000 times. To provide empirical small and large sample properties of the test procedures under study we used alternatively $T = 50, 100, 250, 1000$ as sample sizes.

3.3 Empirical Results

Standard t -ratios obtained from LS inference have empirical significance levels that differ significantly from their nominal counterparts in all cases under study. The strongest violations of the nominal level are obtained for the SV-models. In large samples ($T = 1000$) the empirical rejection frequencies of $H_0 : \phi = 0$ for this model class vary between 37% and 57% with respect to a nominal level of 5%. Size distortions for this procedure increase with the sample size and with the assumed persistence ($\alpha_1 + \beta_1$ or $\alpha_1 + 0.5\alpha_1^- + \beta_1$) of the volatility process. Integrated GARCH-type processes yield empirical rejection frequencies at the 5% nominal level of about 20%. To economize on space we do not provide further results concerning empirical size and power of standard LS inference.

Selected results characterizing empirical size and power properties of the remaining test procedures are given in Table 1 and Table 2 for the GARCH and SV models respectively. We only report the empirical significance levels and all power estimates with respect to the nominal level $\alpha = 0.05$. The results for $\alpha = 0.01$ and $\alpha = 0.1$ were very similar. Violations of the nominal level which are significant with significance level 0.05 are indicated by an asterisk.

In a first step we discuss empirical results obtained for GARCH-type data generating processes:

- Using heteroskedasticity consistent t -ratios the empirical significance levels come close to their nominal counterparts especially in larger samples ($T = 250, 1000$). However, numerous significant violations of the nominal level are encountered in small samples ($T = 50, 100$) for which the empirical levels vary between 6.7% and 8.1%. For the (nearly) integrated processes GARCH2, GARCH3 and GARCH7 the empirical levels are significantly different from 5% using White's t -test even for samples of size $T = 250$.

- QML inference under the assumption that the underlying volatility process is GARCH(1,1) yields empirical rejection frequencies which are equal to their nominal counterparts at the 5% level for almost all GARCH-type processes and sample sizes. The GARCH5 and GARCH7 processes are characterized by the strongest threshold effects. Misspecification of the volatility process, however, only weakly affects the empirical size of QML inference. For $T = 50$ empirical rejection frequencies of about 6.8% are obtained for these processes which differ significantly from the nominal significance level.
- Wild bootstrap inference shows superior size properties relative to all other test procedures. For all GARCH-type processes the obtained empirical significance levels are not significantly different from the nominal level. Both versions of the wild bootstrap procedure used to simulate White's t -statistic behave similarly. Results for the symmetric bootstrap applied to the pseudo LR statistic are thus omitted. It turns out that the symmetric version of the wild bootstrap yields empirical significance levels which are slightly larger compared to those obtained from its asymmetric counterpart. The latter result holds for almost all processes (even the SV models) and sample sizes. In addition, asymmetric versions of the wild bootstrap obtain similar results in simulating the pseudo LR statistic compared to White's t -ratio.
- QML inference assuming a GARCH(1,1) error sequence has superior power properties relative to the remaining inference procedures. In large samples ($T = 1000$) the hypothesis $H_0 : \phi = 0$ is rejected in about 86% of the performed experiments irrespective of the underlying true volatility process.
- Relative to QML inference the remaining procedures are slightly inferior with respect to the GARCH1 model. For the remaining models, which are either asymmetric, integrated or nearly integrated, power estimates are considerably less than those obtained from QML. In the worst case, i.e. for the GARCH7 process which is characterized by strong asymmetry and non-existing unconditional variance, power estimates of the bootstrap procedures are about 35% less than QML estimates. Since the power estimates for bootstrap procedures are increasing with the sample sizes considered one may still conjecture that bootstrap inference is still consistent.
- Power estimates of wild bootstrap inference are slightly less compared to those obtained for White's t -test. Since the empirical size of the latter is larger compared to bootstrap inference one may not expect both procedures to have different power properties conditional on a common empirical size.

A number of results discussing empirical size properties given above does not carry over to inference against linear dependence if error terms are generated by stochastic volatility. First, QML inference shows empirical significance levels which are up to nine times larger compared to the nominal level of 5%. Empirical rejection frequencies of QML inference increase and are mostly severe for the highly persistent processes SV4 to SV6.

Second, White's t -test provides empirical levels up to 12% which are significantly different from 5% for all but one experiment. Compared to QML based inference, however, the empirical size of White's t -test is decreasing with the sample size such that one may conjecture this test to be valid for very large sample sizes in the case of error terms exhibiting stochastic volatility.

Third, for the processes SV1 to SV3 the empirical size of bootstrap procedures is significantly less than the nominal level in a few experiments. The symmetric version of the wild bootstrap used to estimate the distribution of White's t -ratio yields empirical sizes that do not differ from the nominal level. Using the asymmetric wild bootstrap technique both test statistics under study show empirical sizes of 3.9% to 4.1% in a number of experiments amounting to a conservative testing device. Significant violations of the nominal level are, however, more often observed for bootstrap inference via White's t -test.

The estimated power of the competing procedures mirrors to some extent the size properties given above. QML inference shows highest power estimates which is not surprising in the light of the observed size distortions. White's t -test is also characterized by higher power estimates relative to bootstrap inference but regarding the empirical size differences of both procedures, differences in empirical power may be neglected. Bootstrap procedures have almost no power at all for samples of size $T = 50$ for which power and size estimates are close to each other. Power estimates increase with the sample size for all investigated processes. With respect to the highly persistent specifications SV4 to SV6, however, even for $T = 1000$ the obtained power estimates are only about 11%. For the remaining SV-models bootstrap power estimates are at most 30.1% for $T = 1000$.

4 Empirical Application to German Stock Data

The alternative test procedures discussed in the previous sections were applied to the daily returns of 25 major German stocks at the Frankfurt stock exchange from January 2, 1990 to December 30, 1996, which amounts to 1753 observations for each stock. All prices are adjusted for dividends.

As discussed in the introduction, an often observed feature of daily stock returns is a positive autocorrelation at lag 1. In this section, we want to test whether this feature is due to an autoregressive process of order 1. Table 4 and 5 present estimation results for the TGARCH(1,1) model with AR(1) component. Only for Siemens and Thyssen the estimate of the autoregressive parameter is

negative, for all others positive. The same result was obtained for a TGARCH-M model. Note that in the latter case the parameter λ was significant according to QML p-values less frequently than the autoregressive parameter ϕ . Although positive for most stocks as predicted by finance theory, the significance of λ is even less clear than the significance of ϕ . The asymmetry parameter α_1^- for the estimated AR(1)-TGARCH(1,1) models is positive for most cases, supporting the frequently reported leverage effect. However, 5% significance according to QML p-values is obtained for only nine stocks.

The results concerning p-values for ϕ using different test procedures are presented in Table 3. As the bootstrapped versions of White's t -ratio and the pseudo LR statistic yield relatively similar results, we will mainly compare the QML statistics with the bootstrap statistics. Taking a nominal level of 5%, the QML statistic rejects H_0 in twelve cases, whereas the symmetrically bootstrapped White statistic only rejects in five cases. For a nominal level of 5% there are seven stocks (Allianz, Bayerische Vereinsbank, Dresdner Bank, Hoechst, MAN, RWE and Schering) for which QML suggests to reject the null hypothesis and the bootstrap tests do not reject. For a level of 10% there are five stocks (Daimler, MAN, RWE, Schering and Veba) where this occurs. The reverse case of a rejection by the bootstrap tests and no rejection by QML does not occur, except for Bayer when setting a level larger than 15%.

For all estimated models the conditions for existing unconditional variances are satisfied. The stocks with highest persistence in volatility as measured by $\alpha_1 + \alpha_1^-/2 + \beta_1$ in the TGARCH(1,1) model are Allianz, Bayerische Vereinsbank and Commerzbank with a value of 0.99, whereas the smallest degree of persistence is found for Linde with a value of 0.68. The mean degree of persistence is 0.94. Recall from Section 3 that the bootstrap tests lose power for the case of an integrated GARCH process or an SV process with $\alpha = 0.99$. If we assume that the stock return volatility is well described by TGARCH or SV-type models, we have to conclude that only in a few cases the bootstrap tests will not have enough power to reject the null hypothesis.

On the other hand, recall that QML with the assumption of GARCH(1,1) is not a valid inference procedure if the true volatility process is stochastic. As there are many claims in the literature that SV models describe real data better than GARCH models (see e.g. Shephard, 1996, for an overview) and only lack the simplicity of estimation, we conclude that for our data the QML statistics may reject the null hypothesis too often. We clearly prefer the bootstrapped versions of LS based statistics, since for all scenarios investigated we have valid test statistics that lack power in cases that do not seem to occur frequently for our data.

When comparing the three estimated models, we can say that for many stocks the inclusion of an AR(1) component substantially increased the value of the likelihood function. This was true somewhat less for the inclusion of the risk premium parameter λ . This tends to give stronger support for autoregressive models than for risk premium models, as discussed in the introduction. Nev-

ertheless, we showed that the statistical significance of these puzzling effects are much less evident when taking into account possible misspecifications of the volatility process.

5 Conclusions

In this paper, we investigate the inference on linear AR(1) dynamics under heteroskedasticity. We compare several test procedures including standard QML inference and an LS based wild bootstrap test via Monte Carlo simulations. For the volatility process we allow for both threshold GARCH and stochastic volatility dynamics. It is found that among all tests the wild bootstrap inference delivers empirical levels that are closest to the nominal levels. The power of the bootstrap tests is low in the situation of an SV process with very high persistence. However, standard QML inference assuming GARCH(1,1) completely fails for all scenarios of underlying SV processes, where the null hypothesis of no AR dynamics is rejected too often. QML inference is less affected by misspecification with respect to the volatility asymmetry. Applying the alternative tests to German stock returns, we show that in many cases different decisions would be taken with respect to acceptance or rejection of the null hypothesis. Our conclusion is the following: In practice one does not know whether the true volatility process is better described by GARCH or SV, so for estimation convenience one very often prefers GARCH. Keeping in mind possible misspecifications of the volatility process, one should be careful interpreting standard QML inference results.

We observed that in general the volatility parameters remained relatively stable irrespective of the inclusion or exclusion of an autoregressive component. However, the estimated volatility processes differed in many cases substantially. As volatility estimates and predictions may be used in many fields of empirical finance, this observation appears to be of broad relevance. For example, in a related work we investigate the impact of these effects on option pricing (Hafner and Herwartz, 1998).

References

- Black, F. (1976). Studies in Stock Price Volatility Changes, *Proceedings of the 1976 Meeting of the Business and Economics Statistics Section, American Statistical Association*, 177–181.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31, 307–327.
- Bollerslev, T., R.F. Engle, D.B. Nelson (1994). GARCH Models, in: Engle, R.F., D.L. McFadden (eds.) *Handbook of Econometrics*, Vol. 4, Elsevier,

Amsterdam, 2961–3038.

- Bollerslev, T., R.F. Engle, J.M. Wooldridge (1988). A capital asset pricing model with time-varying covariances, *Journal of Political Economy* 96, 116–131.
- Bollerslev, T., J.M. Wooldridge (1992). Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances *Econometric Reviews* 11, 143-172.
- Campbell, J., A. Lo, A. MacKinlay (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.
- Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. inflation, *Econometrica* 50, 987–1008.
- Engle, R.F., Bollerslev, T. (1986). Modelling the Persistence of Conditional Variances, *Econometric Reviews* 5, 81–87.
- Engle, R.F., D.M. Lilien, R.P. Robins (1987). Estimating the Varying Risk Premia in the Term Structure: the GARCH-M Model, *Econometrica* 55, 391–407.
- Engle, R.F., Ng, V.K. (1993). Measuring and Testing the Impact of News on Volatility, *Journal of Finance* 48, 1749–1778.
- Glosten, L., Jagannathan, R., Runkle, D. (1993). Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance* 48, 1779–1801.
- Hafner, C.M., H. Herwartz (1998). Structural Analysis of Portfolio Risk Using Beta Impulse Response Functions, *Statistica Neerlandica*, in press.
- Hafner, C.M., H. Herwartz (1998). Option Pricing under Linear Autoregressive Dynamics and Heteroskedasticity, *mimeo*, Humboldt-Universität zu Berlin.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer Verlag.
- He, C. and Teräsvirta, T. (1998). Properties of Moments of a Family of GARCH Processes, *Journal of Econometrics*, forthcoming.
- Herwartz, H. (1998). Testing Periodicity in Time Series Models – A Recommendation of Bootstrap Methods, *Computational Statistics* 13, 283–300.
- Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lütkepohl, T.C. Lee (1988). *Introduction to the Theory and Practice of Econometrics*. New York: Wiley.

- Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Berlin: Springer Verlag.
- Mammen, E. (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models, *The Annals of Statistics* 21, 255–285.
- Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.
- Neumann, M.H., Kreiss, J.-P. (1998). Regression-type Inference in Nonparametric Autoregression, *Annals of Statistics*, in press.
- Shephard, N. (1996). Statistical Aspects of ARCH and Stochastic Volatility, in: *Time Series Models in Econometrics, Finance and Other Fields*, ed. by D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, Monographs on Statistics and Applied Probability 65, pp. 1–67. Chapman & Hall.
- Taylor, S.J. (1986). *Modelling Financial Time Series*, Chichester: John Wiley.
- White, H. (1980). A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica* 48, 817–838.
- White, H., (1984). *Asymptotic Theory for Econometricians*, San Diego: Academic Press.
- Wu, C.F.J. (1986). Jackknife, Bootstrap, and Other Resampling Methods in Regression Analysis (with discussion), *Annals of Statistics* 14, 1261–1295.
- Zakoian, J.M. (1994). Threshold Heteroskedastic Functions, *Journal of Economic Dynamics and Control* 18, 931–955.

T	Size 5%					Power 5%				
	QML	WH	LRB	WHB	WHS	QML	WH	LRB	WHB	WHS
	GARCH1: $h_t = 0.1 + 0.05e_{t-1}^2 + 0.90h_{t-1}$									
50	5.70	7.00*	5.10	4.65	4.65	10.15	11.25	7.90	7.30	8.10
100	4.75	6.70*	5.10	5.00	5.20	16.10	16.25	14.70	13.20	14.10
250	4.75	5.90	4.95	4.50	4.70	32.80	31.95	30.10	28.90	28.85
1000	5.10	4.90	4.60	4.50	4.70	87.60	84.35	84.25	83.95	83.90
	GARCH2: $h_t = 0.1 + 0.09e_{t-1}^2 + 0.90h_{t-1}$									
50	6.00	7.55*	4.95	4.85	4.90	10.55	11.70	8.10	7.30	8.10
100	5.40	7.10*	5.35	4.70	5.10	17.05	16.50	13.85	12.15	13.50
250	4.75	6.30*	5.20	4.80	4.65	32.20	28.85	27.50	25.95	25.70
1000	4.30	5.20	5.00	4.90	4.95	88.00	69.95	69.50	68.85	68.45
	GARCH3: $h_t = 0.1 + 0.10e_{t-1}^2 + 0.90h_{t-1}$									
50	5.95	7.75*	5.15	5.00	5.00	10.65	11.70	8.00	7.10	8.10
100	5.15	6.85*	5.40	4.85	5.30	16.60	16.15	13.85	12.75	13.50
250	4.70	6.40*	5.05	5.05	4.65	31.90	26.85	25.15	23.80	23.60
1000	4.35	5.45	4.90	4.90	4.70	87.55	61.15	61.00	59.35	59.80
	GARCH4: $h_t = 0.1 + 0.05e_{t-1}^2 + 0.08e_{t-1}^2 I(e_{t-1} < 0) + 0.90h_{t-1}$									
50	5.90	7.90*	5.25	4.70	5.15	11.05	11.75	7.90	7.40	7.75
100	5.45	7.20*	5.45	4.80	5.35	16.70	15.80	13.20	12.25	13.05
250	4.65	5.80	5.10	4.45	4.45	32.10	28.00	25.85	24.25	23.70
1000	4.80	5.35	4.75	4.80	4.95	87.25	66.70	66.10	65.65	64.90
	GARCH5: $h_t = 0.1 + 0.18e_{t-1}^2 I(e_{t-1} < 0) + 0.90h_{t-1}$									
50	6.85*	7.95*	5.25	5.30	5.40	11.80	11.50	7.45	6.70	7.45
100	5.65	7.55*	5.35	5.15	5.80	17.55	15.10	11.95	10.75	11.95
250	5.20	6.00	4.85	4.70	4.80	32.05	24.80	23.20	21.40	22.10
1000	4.95	5.05	4.70	4.55	4.70	85.80	57.75	57.20	56.00	56.35
	GARCH6: $h_t = 0.1 + 0.05e_{t-1}^2 + 0.10e_{t-1}^2 I(e_{t-1} < 0) + 0.90h_{t-1}$									
50	6.00	8.10*	5.50	5.05	5.40	10.70	11.70	7.70	7.15	7.75
100	5.30	7.25*	5.20	4.65	5.35	17.10	15.80	13.20	11.90	12.90
250	4.50	5.65	4.75	4.45	4.75	31.55	25.75	24.45	22.45	22.55
1000	4.55	4.95	4.45	4.35	4.50	86.80	57.50	56.85	55.25	55.20
	GARCH7: $h_t = 0.1 + 0.20e_{t-1}^2 I(e_{t-1} < 0) + 0.90h_{t-1}$									
50	6.75*	7.95*	5.20	5.35	5.65	11.75	11.30	7.60	6.90	7.40
100	6.00	7.75*	5.35	5.25	5.75	17.30	14.75	11.35	10.25	11.80
250	5.40	6.35*	4.80	4.75	4.90	32.20	23.05	21.70	19.85	20.35
1000	5.05	5.20	4.80	4.30	4.50	85.30	48.50	47.50	45.40	46.55

Table 1: Empirical rejection percentiles of $H_0 : \phi = 0$ for competing tests against significant autoregressive dynamics in GARCH-type series of length T at a nominal level of 5%. QML is quasi maximum likelihood using GARCH(1,1), WH White's heteroskedasticity correction, LRB is a bootstrapped pseudo LR statistic and WHB and WHS are two bootstrapped versions of WH.

T	Size 5%					Power 5%				
	QML	WH	LRB	WHB	WHS	QML	WH	LRB	WHB	WHS
	SV1: $\log h_t = -0.1 + 0.95 \log h_{t-1} + 0.624\eta_t$									
50	18.75*	10.10*	4.20	4.40	4.60	22.55	12.75	5.25	5.20	5.45
100	22.15*	8.55*	4.15	4.25	4.25	30.65	13.30	7.55	6.80	7.50
250	26.65*	8.25*	4.70	4.10*	4.70	44.25	16.20	11.75	10.35	11.35
1000	33.15*	6.70*	4.15	4.00*	4.40	79.30	25.65	22.95	21.10	21.60
	SV2: $\log h_t = -0.1 + 0.95 \log h_{t-1} + 0.725\eta_t$									
50	23.75*	10.55*	4.05*	4.15	4.40	26.75	12.95	5.05	4.85	5.25
100	29.70*	9.00*	4.00*	4.10*	4.20	36.70	13.85	7.05	5.85	6.60
250	36.80*	9.00*	4.70	4.15	4.45	50.25	15.40	10.25	9.40	9.60
1000	46.90*	7.50*	4.20	3.90*	4.60	79.60	22.35	18.30	16.60	17.30
	SV3: $\log h_t = -0.1 + 0.95 \log h_{t-1} + 0.505\eta_t$									
50	12.90*	9.75*	4.35	4.45	4.75	17.25	11.60	5.40	5.30	5.90
100	14.85*	7.75*	4.20	3.95*	4.40	24.25	12.85	8.60	7.70	7.90
250	18.15*	7.30*	4.95	4.10*	4.55	38.40	16.70	12.75	11.90	12.75
1000	19.75*	6.00	4.15	4.35	4.45	79.30	33.55	31.35	29.65	30.10
	SV4: $\log h_t = -0.1 + 0.99 \log h_{t-1} + 0.631\eta_t$									
50	23.25*	11.95*	4.40	4.45	4.75	27.60	13.55	5.95	5.35	5.20
100	31.15*	11.70*	4.60	4.40	5.15	37.70	15.25	7.60	7.20	7.25
250	39.05*	11.05*	5.00	4.80	5.35	53.75	15.85	10.30	8.85	9.65
1000	46.70*	9.20*	5.40	4.75	5.60	81.30	16.65	12.15	10.95	11.35
	SV5: $\log h_t = -0.1 + 0.99 \log h_{t-1} + 0.652\eta_t$									
50	24.50*	12.10*	4.30	4.40	4.65	28.60	13.65	6.00	5.15	5.50
100	32.80*	11.90*	4.75	4.40	5.10	39.80	15.35	7.60	7.10	7.30
250	41.20*	11.20*	4.85	4.85	5.10	54.95	16.15	10.10	8.90	9.50
1000	48.95*	9.60*	5.35	5.05	5.60	81.35	16.80	11.80	10.50	11.05
	SV6: $\log h_t = -0.1 + 0.99 \log h_{t-1} + 0.609\eta_t$									
50	21.80*	11.75*	4.45	4.55	4.70	26.80	13.65	6.00	5.25	5.25
100	29.25*	11.50*	4.55	4.45	5.10	36.40	15.00	7.65	7.20	7.25
250	36.95*	10.95*	5.15	5.20	5.45	52.90	15.90	10.45	9.05	9.95
1000	43.65*	9.20*	5.55	4.85	5.55	80.50	16.70	12.65	11.25	11.90

Table 2: Empirical rejection percentiles of $H_0 : \phi = 0$ for competing tests against significant autoregressive dynamics in SV-type series of length T at a nominal level of 5%. QML is quasi maximum likelihood using GARCH(1,1), WH White's heteroskedasticity correction, LRB is a bootstrapped pseudo LR statistic and WHB and WHS are two bootstrapped versions of WH.

Stock	TGARCH(1,1)-AR(1)				TG-M	TG-M, AR(1)	
	$\hat{\alpha}$	$\hat{\alpha}^-$	$\hat{\beta}$	$\hat{\phi}$	$\hat{\lambda}$	$\hat{\phi}$	$\hat{\lambda}$
ALLI	0.03 (0.04)	0.08 (0.00)	0.92 (0.00)	0.09 (0.00)	-0.02 (0.80)	0.09 (0.00)	0.00 (0.99)
BASF	0.10 (0.00)	0.05 (0.20)	0.79 (0.00)	0.03 (0.35)	0.15 (0.19)	0.03 (0.36)	0.16 (0.16)
BAYE	0.07 (0.04)	0.08 (0.06)	0.81 (0.00)	0.01 (0.68)	0.11 (0.34)	0.01 (0.80)	0.12 (0.34)
BAYH	0.06 (0.01)	-0.03 (0.30)	0.92 (0.00)	0.07 (0.01)	0.05 (0.66)	0.08 (0.01)	0.09 (0.45)
BAYV	0.09 (0.02)	0.01 (0.92)	0.89 (0.00)	0.07 (0.02)	0.02 (0.81)	0.07 (0.02)	0.01 (0.88)
BMW	0.09 (0.01)	0.01 (0.80)	0.85 (0.00)	0.12 (0.00)	0.10 (0.32)	0.12 (0.00)	0.11 (0.31)
COMM	0.07 (0.01)	-0.03 (0.53)	0.93 (0.00)	0.01 (0.63)	0.06 (0.48)	0.01 (0.65)	0.11 (0.28)
DAIM	0.03 (0.15)	0.03 (0.07)	0.93 (0.00)	0.04 (0.09)	0.02 (0.87)	0.04 (0.08)	0.04 (0.75)
DEUT	0.01 (0.27)	0.04 (0.12)	0.94 (0.00)	0.03 (0.29)	0.64 (0.06)	0.03 (0.13)	0.50 (0.09)
DEGU	0.06 (0.01)	0.00 (0.89)	0.87 (0.00)	0.04 (0.11)	0.04 (0.85)	0.04 (0.15)	-0.02 (0.88)
DRES	0.03 (0.12)	0.02 (0.35)	0.94 (0.00)	0.08 (0.00)	0.08 (0.48)	0.08 (0.01)	0.08 (0.46)
HENK	0.00 (0.94)	0.01 (0.48)	0.95 (0.00)	0.02 (0.51)	0.10 (0.82)	0.03 (0.32)	0.46 (0.43)
HOEC	0.07 (0.01)	0.10 (0.04)	0.71 (0.00)	0.08 (0.00)	0.26 (0.14)	0.09 (0.00)	0.45 (0.03)

Table 3: Parameter estimates and QML p-values (in parentheses) of the TGARCH(1,1) model for the German stocks (2nd to 5th column). Column 6 reports the estimates of λ for the TGARCH(1,1)-M model, and the last two columns the estimates of λ and ϕ for the AR(1)-TGARCH(1,1)-M model.

Stock	TGARCH(1,1)-AR(1)				TG-M	TG-M, AR(1)	
	$\hat{\alpha}$	$\hat{\alpha}^-$	$\hat{\beta}$	$\hat{\phi}$	$\hat{\lambda}$	$\hat{\phi}$	$\hat{\lambda}$
LIND	0.04 (0.22)	0.10 (0.06)	0.59 (0.00)	0.10 (0.00)	0.23 (0.37)	0.11 (0.00)	0.81 (0.12)
MAN	0.00 (0.89)	0.08 (0.03)	0.81 (0.00)	0.09 (0.00)	0.34 (0.46)	0.10 (0.00)	0.28 (0.48)
MANN	0.03 (0.46)	0.12 (0.04)	0.81 (0.00)	0.05 (0.11)	-0.04 (0.78)	0.05 (0.11)	-0.01 (0.97)
MUEN	0.04 (0.02)	0.07 (0.04)	0.83 (0.00)	0.18 (0.00)	0.30 (0.03)	0.18 (0.00)	0.34 (0.03)
PREU	0.01 (0.34)	0.04 (0.02)	0.95 (0.00)	0.10 (0.00)	0.25 (0.04)	0.10 (0.00)	0.21 (0.07)
RWE	0.02 (0.25)	0.09 (0.01)	0.91 (0.00)	0.08 (0.01)	0.02 (0.85)	0.07 (0.02)	0.04 (0.69)
SCHE	0.17 (0.00)	-0.01 (0.79)	0.71 (0.00)	0.06 (0.02)	0.25 (0.05)	0.06 (0.03)	0.23 (0.07)
SIEM	-0.00 (0.89)	0.05 (0.00)	0.95 (0.00)	-0.04 (0.10)	0.06 (0.71)	-0.03 (0.16)	0.05 (0.77)
THYS	0.03 (0.15)	0.05 (0.07)	0.91 (0.00)	-0.01 (0.59)	0.07 (0.68)	-0.01 (0.61)	0.09 (0.61)
VEBA	0.07 (0.10)	0.02 (0.81)	0.88 (0.00)	0.04 (0.08)	-0.04 (0.66)	0.02 (0.37)	-0.03 (0.68)
VIAG	0.07 (0.13)	-0.02 (0.75)	0.91 (0.00)	0.03 (0.18)	0.08 (0.48)	0.04 (0.21)	0.08 (0.48)
VOLK	0.01 (0.75)	0.06 (0.02)	0.92 (0.00)	0.02 (0.36)	-0.03 (0.85)	0.02 (0.39)	-0.02 (0.90)

Table 4: Parameter estimates and QML p-values (in parentheses) of the TGARCH(1,1) model for the German stocks (2nd to 5th column). Column 6 reports the estimates of λ for the T GARCH(1,1)-M model, and the last two columns the estimates of λ and ϕ for the AR(1)-TGARCH(1,1)-M model.

Stock	QML	PLR	WH	LRB	LRS	WHB	WHS
ALLI	0.00	0.03	0.10	0.07	0.10	0.11	0.13
BASF	0.35	0.57	0.66	0.67	0.66	0.68	0.71
BAYE	0.68	0.04	0.14	0.14	0.15	0.16	0.18
BAYH	0.01	0.01	0.05	0.04	0.05	0.04	0.05
BAYV	0.02	0.05	0.19	0.19	0.17	0.21	0.23
BMW	0.00	0.00	0.01	0.01	0.01	0.01	0.02
COMM	0.63	0.56	0.66	0.66	0.62	0.65	0.71
DAIM	0.09	0.08	0.14	0.12	0.14	0.16	0.16
DEUT	0.29	0.43	0.53	0.52	0.51	0.54	0.60
DEGU	0.11	0.58	0.67	0.69	0.67	0.68	0.68
DRES	0.00	0.02	0.08	0.06	0.06	0.09	0.10
HENK	0.51	0.63	0.70	0.69	0.67	0.75	0.74
HOEC	0.00	0.02	0.09	0.07	0.08	0.10	0.11
LIND	0.00	0.00	0.01	0.00	0.00	0.01	0.01
MAN	0.00	0.03	0.23	0.21	0.20	0.31	0.27
MANN	0.11	0.06	0.19	0.18	0.19	0.22	0.23
MUEN	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PREU	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RWE	0.01	0.43	0.65	0.66	0.66	0.70	0.72
SCHE	0.02	0.20	0.39	0.42	0.39	0.41	0.44
SIEM	0.10	0.35	0.42	0.42	0.42	0.45	0.45
THYS	0.59	0.45	0.59	0.62	0.58	0.60	0.64
VEBA	0.08	0.71	0.79	0.76	0.79	0.81	0.81
VIAG	0.18	0.22	0.37	0.36	0.36	0.36	0.42
VOLK	0.36	0.42	0.50	0.50	0.48	0.52	0.55

Table 5: P-values of the AR(1) coefficient for German stocks. QML is quasi maximum likelihood, PLR pseudo likelihood ratio and WH White's heteroskedasticity correction. The last four columns report bootstrapped versions of the LR and WH p-values, where B uses the two point distribution and S a normal distribution.