Asymmetries in Conditional Mean and Variance: Modelling Stock Returns by asMA-asQGARCH

Kurt Brännäs
Jan G. De Gooijer

ASYMMETRIES IN CONDITIONAL MEAN AND VARIANCE: MODELLING STOCK RETURNS BY asMA-asQGARCH[*]

Kurt Brännäs[1] and Jan G. De Gooijer[2]

[1] Department of Economics, Umeå University
SE-901 87 Umeå, Sweden
Telephone: +46–90–786 6101; Fax: +46–90–772302
e-mail: kurt.brannas@econ.umu.se

[2] Department of Economic Statistics, University of Amsterdam
Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
Telephone: +31–20–525 4244; Fax: +31–20–525 4349
e-mail: jandeg@fee.uva.nl

Abstract

The asymmetric moving average model (asMA) is extended to allow for asymmetric quadratic conditional heteroskedasticity (asQGARCH). The asymmetric parametrization of the conditional variance encompasses the quadratic GARCH model of Sentana (1995). We introduce a framework for testing asymmetries in the conditional mean and the conditional variance, separately or jointly. Some of the new model's moment properties are also derived. Empirical results are given for the daily returns of the composite index of the New York Stock Exchange. There is strong evidence of asymmetry in both the conditional mean and conditional variance functions. In a genuine out-of-sample forecasting experiment the performance of the best fitted asMA-asQGARCH model is compared to pure asMA and no-change forecasts. This is done both in terms of conditional mean forecasting as well as in terms of risk forecasting.

Key Words: Time series, finance, nonlinearity, estimation, testing, forecasting, NYSE.

JEL: C22, C51, C52, C53, G14.

---

1 Introduction

There is an overwhelming amount of empirical evidence that the conditional variance (volatility) of economic and financial time series, say $V(Y_t|Y_{t-1}, Y_{t-2}, \ldots)$, is asymmetric in the sense that large negative shocks are often followed by larger increases in volatility than from equally large positive shocks. Various parametric models for this conditional variance asymmetry have been proposed in the generalized autoregressive heteroskedasticity (GARCH) literature; see, e.g., Hentschel (1995) and Pagan (1996) for overviews. The structure of most existing GARCH models only describe one particular feature of the asymmetric conditional probability density function. Therefore, it seems natural to develop models for financial time series that include other aspects of the conditional distribution function as well. In particular, a different delay in reacting to positive rather than negative innovations (up markets versus down markets) can be represented by an asymmetric parametrization of the conditional mean. The asymmetric moving average (asMA) model proposed by Wecker (1981) is a suitable candidate for this purpose; see, e.g., Brännäs and De Gooijer (1994). The model employs two separate (linear) filters one for positive and one for negative shocks, i.e. the response of the system is asymmetric with respect to the sign of the innovation.

In this article we combine the asMA model for the conditional mean $E(Y_t|Y_{t-1}, Y_{t-2}, \ldots)$ with an analogously defined asymmetric parametrization of the conditional variance. The latter model is an asymmetric extension of the quadratic GARCH model of Sentana (1995). Thus the resulting model allows for both types of asymmetry in stock return data, and is for ease of reference denoted asMA-asQGARCH. In particular, we are concerned with the asymmetric behavior of the NYSE composite daily returns Y_t, defined as $Y_t = 100[\ln(I_t) - \ln(I_{t-1})]$, where I_t is the daily price index. The data cover the period January 1, 1981 – December 31, 1999 (4957 observations) and were obtained from Datastream. Figures 1.a and 1.b show plots of respectively the returns and squared returns. In Figure 1.a there is a sharp decline in the series in October 1987 due to the world-wide crash in stock markets. The squared returns plotted in Figure 1.b display this phenomenon even more significantly. Figures 1.c and 1.d show scatter plots of respectively, Y_t versus $Y_{t-1}^{\pm}$, and Y_t^2 versus $Y_{t-1}^{\pm}$, where $Y_t^{\pm}$ are the positive/negative returns. The straight lines represent linear regression equations. Evidently Figure 1.c shows an asymmetry in the conditional mean whereas the lines in Figure 1.d refer to an asymmetry in the conditional variance.[1]

The idea to add a nonlinear model for the conditional variance to a nonlinear model for the conditional mean has been adopted by various authors; see, e.g., Glosten, Jagannathan and Runkle (1993), Li and Li (1996), Lee and Li (1998), and Lundbergh and Teräsvirta (1998). The asMA-asQGARCH model proposed here is a very flexible model. The asMA part is well-suited to describe asymmetry of the type that frequently is encountered in macro-economic time series. For instance, the model has been successfully applied by Brännäs and De Gooijer (1994) to characterize asymmetries in U.S. business cycle data. Further, since the asQGARCH part nests the QGARCH model of Sentana (1995), it shares the same attractive features of this latter model. Finally, the asMA-asQGARCH model easily allows for testing the hypothesis that only

---

[1]The two-regime self-exciting threshold autoregressive (SETAR) model of Tong (1990) can be regarded as a special case of Wecker's (1981) asMA model; see, e.g., Brännäs and De Gooijer (1994, Appendix) and Section 2 below. Hence, the regression lines in Figures 1.c and 1.d also indicate the presence of regime-switching dynamics in Y_t.
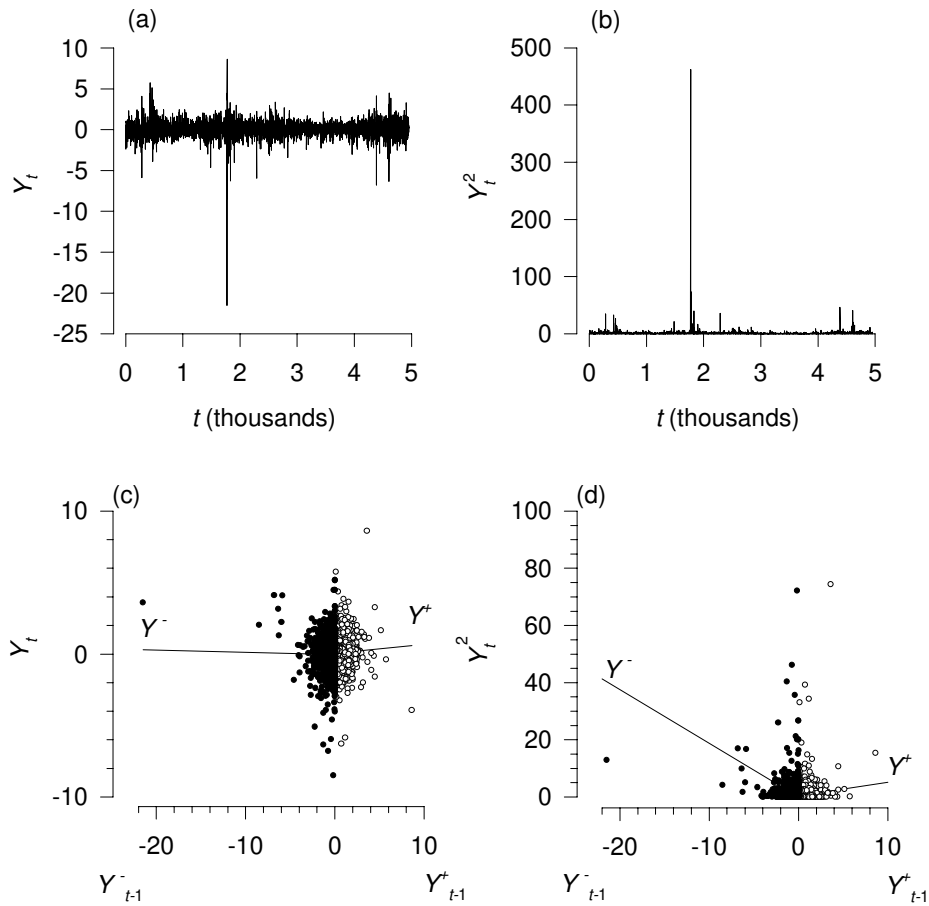
Figure 1: Daily returns of the New York Stock Exchange, 1981–1999 (sample mean 0.043, standard deviation 0.912). Plot (a) graphs the time series Y_t, plot (b) contains the squared series Y_t^2, in (c) Y_t is plotted separately against Y_{t-1}^+ (white marker) and Y_{t-1}^- (black marker) with regression lines for each subgroup, and in (d) Y_t^2 is graphed against Y_{t-1}^+ (white marker) and Y_{t-1}^- (black marker) with regression lines for each subgroup. (Y_{t-1} minus the sample mean is denoted Y_{t-1}^+ if the difference is positive and is not defined otherwise, Y_{t-1} minus the sample mean is denoted Y_{t-1}^- and defined if negative and not defined otherwise).

the conditional mean model captures the asymmetric impact of shocks on volatility or, more elaborate, that both the conditional mean and conditional variance specifications capture this phenomenon jointly. In other words, one can detect asymmetries of various kind relatively easy.

The outline of this paper is as follows. In the next section the asMA-asQGARCH model is introduced. Some basic statistical properties are given in Section 3. Section 4 describes the estimation method and the associated tests of symmetry in the conditional mean and the conditional variance or both. Estimation and testing results for asMA-asQGARCH model fitted to the NYSE composite daily stock returns are presented in Section 5. This section also contains genuine out-of-sample forecasts for prediction horizons of 1-4 days through a rolling forecasting approach. Section 6 provides some concluding remarks.

2   asMA model with asQGARCH errors

Let $\{u_t\}$ be a real-valued discrete time stochastic process generated by

$$u_t = \varepsilon_t h_t \tag{1}$$

where $\{\varepsilon_t\}$ is a sequence of independent, identically distributed (iid) random variables with mean zero and unit variance, and the conditional standard deviation h_t is independent of ε_t as well as non-negative for all t. Further, let

$$u_t^+ = \max(0, u_t) = \varepsilon_t^+ h_t \quad \text{and} \quad u_t^- = \min(u_t, 0) = \varepsilon_t^- h_t$$

where $\varepsilon_t^+ = \max(0, \varepsilon_t)$ and $\varepsilon_t^- = \min(\varepsilon_t, 0)$. Now the general asymmetric moving average process of order q, which we shall abbreviate by asMA(q), is defined by the stochastic sequence $\{Y_t\}$ satisfying

$$
\begin{aligned}
Y_t &= u_t + \sum_{i=1}^{q} \theta_i^+ u_{t-i}^+ + \sum_{i=1}^{q} \theta_i^- u_{t-i}^- \\
&= u_t + \sum_{i=1}^{q} \theta_i^- u_{t-i} + \sum_{i=1}^{q} \omega_i I(u_{t-i} \geqslant 0) u_{t-i}
\end{aligned}
\tag{2}
$$

where $\omega_i = \theta_i^+ - \theta_i^-$ $(i = 1, \ldots, q)$, and where $I(\cdot)$ is the indicator function. Since the values of the θ_i^+ and θ_i^- parameters at the ith lag may be different, the response to equally sized positive and negative shocks may be asymmetric. This effect will prevail in the conditional mean of Y_t given past observations. If there is a threshold level r instead of 0 in the u_t^+ and u_t^- functions it can be accounted for by including a constant term, say θ_0, in (1) and retaining an $r = 0$ threshold in u_t^+ and u_t^-.[2]

To illustrate the asymmetry inherent in (2), consider the case $q = 1$. Let L denote the lag operator such that $L^j Y_t = Y_{t-j}$. Rewrite the asMA(1) model to obtain $Y_1 = u_1$ at $t = 1$, $(1 - \theta^+ L) Y_2 = u_2$ for $u_1 > 0$, and $(1 - \theta^- L) Y_2 = u_2$ for $u_1 < 0$ at time $t = 2$. Repeating this procedure the asMA(1) model gives rise to a "dual" nonlinear AR representation with s^{t-1}

---

[2]Let $\{u_t^+\}$ and $\{u_t^-\}$ be two sequences with a threshold level r instead of 0. Set the means of u_t^+ and u_t^- to μ_r^+ and μ_r^-, respectively. For $r = 0$ the means are μ_0^+ and μ_0^-. We may write $u_t^{+0} = u_t^+ + \mu_0^+ - \mu_r^+$ and $u_t^{-0} = u_t^- + \mu_0^- - \mu_r^-$. Substitution into (2) shows that a constant term is sufficient to replace r by 0 in the $\{u_t^+\}$ and $\{u_t^-\}$ sequences.

regimes of different linear AR polynomials. In particular, let $\theta(k) = \theta^+ I(u_k \geqslant 0) + \theta^- I(u_k < 0)$ ($k > 1$). Then for arbitrary t the resulting model has the form

$$Y_t = \theta(t-1)Y_{t-1} - \theta(t-1)\theta(t-2)Y_{t-2} + \ldots - \theta(t-1)\cdots\theta(1)Y_1 + u_t.$$

As $\theta(t-1)\cdots\theta(t-k) = (\theta^+)^j (\theta^-)^{k-j}$ ($j = 0, \ldots, k$) arises with binomial probability $\Pr(j) = [k!/j!(k-j)!]2^k$ for symmetric u_t, the probability for each possible outcome in the autoregression can be calculated. The density function $\Pr(j)$ is symmetric, and the expected value of the parameter combinations is given by $\sum_{j=0}^{k} \Pr(j)(\theta^+)^j (\theta^-)^{k-j}$. Brännäs and Ohlsson (1999) give a numerical illustration.

Various models have been proposed to represent the conditional heteroskedasticity $h_t^2 = V(Y_t|Y_{t-1}, Y_{t-2}, \ldots)$ in (1). Sentana (1995) introduced the QGARCH(p,p) model in which the conditional variance is given by

$$h_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i u_{t-i} + \sum_{i=1}^{p} \beta_i u_{t-i}^2 + \sum_{i=1}^{p} \gamma_i h_{t-i}^2. \tag{3}$$

The second term on the right-hand side describes the asymmetry in the conditional variance. Of course, this term may also cause a problem with the positivity of h_t^2 in (3) unless parameters are constrained. In (3) positive shocks have a different effect than negative shocks. The response of the process is parabolic, though not symmetric around zero.

An alternative parametrization of the conditional variance, similar in structure to (3) and which allows for asymmetry in an even more flexible way, can be defined as

$$\begin{aligned} h_t^2 &= \alpha_0 + \sum_{i=1}^{Q}(\alpha_i^+ u_{t-i}^+ + \alpha_i^- u_{t-i}^-) + \sum_{i=1}^{p} \beta_i u_{t-i}^2 + \sum_{i=1}^{p} \gamma_i h_{t-i}^2 \\ &= \alpha_0 + \sum_{i=1}^{Q} \alpha_i^- u_{t-i} + \sum_{i=1}^{Q} \delta_i I(u_{t-i} \geqslant 0)u_{t-i} + \sum_{i=1}^{p} \beta_i u_{t-i}^2 + \sum_{i=1}^{p} \gamma_i h_{t-i}^2 \end{aligned} \tag{4}$$

where $\delta_i = \alpha_i^+ - \alpha_i^-$. Thus, by introducing u_{t-i}^+ and u_{t-i}^- each with their own coefficients, we allow that the response of the process to positive and negative shocks in volatility can be different from a parabolic function. We will refer to (1) jointly with (4) as the Asymmetric Quadratic Generalized ARCH (asQGARCH) model of order $(Q;p,p)$. Clearly, if $\delta_i = 0$ for all i ($i = 1, \ldots, Q$) and $Q = p$, (4) reduces to (3). We also see that in the case $\alpha_i^+ = \alpha_i^- = 0$, equations (1) and (4) simplify to the GARCH model of order (p,p) introduced by Bollerslev (1986). Note, however, that in the case $Q = p = 1$, (4) differs from the so-called Asymmetric Threshold GARCH (asTGARCH) of order (1;1,1) of Koutmos (1999) which is an asymmetric analogue of the TGARCH(1,1) model of Zakoïan (1994).

3 Mean, variance, and autocovariances

To highlight the differences between the asMA(q) model and the asMA(q)-asQGARCH(1;1,1) model it is important to derive some of the moment properties of the latter model. Denote the ith moment of $\{\varepsilon_t^+\}$ by $\nu_i^+ = E[(\varepsilon_t^+)^i]$. Then, using the fact that $\nu_i^+ = -\nu_i^- = -E[(\varepsilon_t^-)^i]$ when $\{\varepsilon_t\}$ is symmetrically distributed, it follows immediately from (2) and (1) that the mean of $\{Y_t\}$

is given by

$$\mu = E[Y_t] = \sum_{i=1}^{q}(\theta_i^+ E[u_t^+] + \theta_i^- E[u_t^-]) = \sum_{i=1}^{q}(\theta_i^+ - \theta_i^-)\nu_1^+ E[h_t]. \tag{5}$$

After some tedious but rather straightforward algebra it can be shown that the lag k ($k = 0, 1, \ldots$) autocovariance γ_k of (2) and (1) is given by

$$\gamma_k = \begin{cases} \begin{aligned} & E[u_t u_{t-k}] + \sum_{i=1}^{q}(\theta_i^+ + \theta_i^-)E[u_t^+ u_{t-k+i}] + E[u_t^+ u_{t-k}^+]\sum_{i=1}^{q}\{(\theta_i^+)^2 + (\theta_i^-)^2\} \\ & + 2E[u_t^- u_{t-k}^+]\sum_{i=1}^{q}\theta_i^+\theta_i^- \\ & + \sum_{i=1}^{q-1}\sum_{j=i+1}^{q}\Big((\theta_i^+\theta_j^+ + \theta_i^-\theta_j^-)\{E[u_t^+ u_{t-k+(j-i)}^+] + E[u_t^+ u_{t-k-(j-i)}^+]\} \\ & + (\theta_i^-\theta_j^+ + \theta_i^+\theta_j^-)\{E[u_t^- u_{t-k+(j-i)}^+] + E[u_t^- u_{t-k-(j-i)}^+]\}\Big) - \mu^2, \quad \text{if} \quad k \leqslant q \\[2ex] & \{\sum_{i=1}^{q}(\theta_i^+ E[u_t^+ u_{t-k}^+] + \theta_i^- E[u_t^- u_{t-k}^-])\}^2 - \mu^2, \quad \text{if} \quad k > q \end{aligned} \end{cases} \tag{6}$$

where

$$E[u_t^{\pm} u_{t-i}^{\pm}] = \begin{cases} (\nu_1^+)^2 E[h_t h_{t-i}], & i \neq 0 \\ \nu_2^+ E[h_t^2], & i = 0 \end{cases} \qquad E[u_t^+ u_{t+i}] = \begin{cases} 0, & i \neq 0 \\ \nu_2^+ E[h_t^2], & i = 0 \end{cases}$$

$$E[u_t^+ u_{t\pm\{k\pm(j-i)\}}^-] = \begin{cases} -(\nu_1^+)^2 E[h_t h_{t\pm\{k\pm(j-i)\}}], & k \neq |j-i| \\ 0, & k = |j-i| \end{cases}$$

and where an explicit expression for $E[h_t^2]$ is given in Appendix A.

Clearly, if $u_t \equiv \varepsilon_t$ with $\{\varepsilon_t\}$ a sequence of iid $(0,1)$ distributed random variables (6) becomes the autocovariance of the pure asMA(q) model.[3] In that case $\gamma_k = 0$, for $k > q$. Hence, in contrast to the pure asMA(q) and the symmetric MA(q) model, the asMA(q)-asQGARCH(1;1,1) model does not have autocovariances equal to zero beyond lags $k > q$. In other words, the asymmetry in the conditional variance added through the innovations clearly changes the correlation structure of the process. Consequently it may, at least on the basis of the theoretical autocovariances and autocorrelations, be possible to empirically discriminate between a pure asMA(q) and an asMA(q)-asQGARCH(1;1,1) model.

By setting $k = 0$ in (6) it follows directly that the asMA-asQGARCH model has a variance given by

$$\begin{aligned} \gamma_0 &= E[u_t^2] + E[(u_t^+)^2]\sum_{i=1}^{q}\{(\theta_i^+)^2 + (\theta_i^-)^2\} + 2\sum_{i=1}^{q-1}\sum_{j=i+1}^{q}\Big((\theta_i^+\theta_j^+ + \theta_i^-\theta_j^-)E[u_t^+ u_{t+(j-i)}^+] \\ &\quad + (\theta_i^-\theta_j^+ + \theta_i^+\theta_j^-)E[u_t^- u_{t+(j-i)}^+]\Big) - \mu^2 \\ &= \Big(\nu_2 + \nu_2^+\sum_{i=1}^{q}\{(\theta_i^+)^2 + (\theta_i^-)^2\}\Big)E[h_t^2] \\ &\quad + 2(\nu_1^+)^2\sum_{i=1}^{q-1}\sum_{j=i+1}^{q}(\theta_i^+ - \theta_i^+)(\theta_j^- - \theta_j^-)E[h_t h_{t+(j-i)}] - \mu^2 \end{aligned} \tag{7}$$

where $\nu_i = E[|\varepsilon_t|^i]$ ($i = 1, 2, \ldots$). By setting $\theta_i^+ = \theta_i^- \equiv \theta_i$ ($i = 1, \ldots, q$), (7) reduces to $\gamma_0 = \{\nu_2 + 2\nu_2^+\sum_{i=1}^{q}\theta_i^2\}E[h_t^2]$, i.e. the variance of an MA(q)-asQGARCH(1;1,1) model. When

---

[3] Wecker's (1981) formula (A.1) for the autocovariance at lag $k \leqslant q$ of an asMA(q) model with iid $N(0, \sigma^2)$ innovations is incorrect.
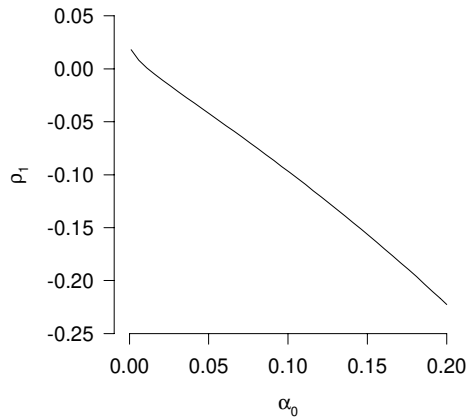
Figure 2: Lag one autocorrelation coefficient, ρ_1, against α_0. Unknown expectations in (6)-(7) are estimated by sample averages from generated series of length $T = 100\,000$.

again $u_t = \varepsilon_t$, this latter expression becomes equal to $\nu_2(1 + \sum_{i=1}^{q} \theta_i^2)$, i.e. the well-known formula for the variance of an MA(q) process.

Depending on the signs of the $\theta_i^{\pm}$ $(i = 1, \ldots, q)$ the model can capture an inverse relationship between autocorrelation and volatility (cf. LeBaron, 1992). Using estimated parameters (see Section 5, below) a time series of length $T = 100\,000$ is generated. From the generated series the unknown expected values are estimated as sample averages, and used in (6)-(7). In Figure 2 the lag one autocorrelation coefficient, $\rho_1 = \gamma_1/\gamma_0$, is graphed against α_0 (a larger α_0 corresponds to a larger $E(h_t^2)$). The expectations $E(Y_t^i)$ $(i = 1, 2)$ are increasing with α_0.

4 Estimation and Testing

Conditional on $Y_{t-1} = (y_1, \ldots, y_{t-1})$ the prediction error

$$e_t = Y_t - \theta_0 - \sum_{i=1}^{q} \theta_i^+ u_{t-i}^+ - \sum_{i=1}^{q} \theta_i^- u_{t-i}^-$$

has the distribution of $\varepsilon_t h_t$. Note that the included constant term θ_0 catches the mean return but also any nonzero threshold in the $\{u_t^+\}$ and $\{u_t^-\}$ sequences (cf. footnote 2). We assume $\{\varepsilon_t\}$ to be normally distributed so that the conditional density of Y_t given Y_{t-1} is normal with mean 0 and variance h_t^2. The log-likelihood function is then

$$\ell = \sum_{t=r}^{T} \ln L_t \propto -\sum_{t=r}^{T} \left(\ln h_t + \frac{1}{2} \frac{e_t^2}{h_t^2} \right)$$

where $r = \max(q, Q, p) + 1$. Since, we have moving average terms in the conditional mean and the conditional variance, the likelihood equations are more complicated than in the absence of

moving average terms. The gradient for observation t is

$$\frac{\partial \ln L_t}{\partial \boldsymbol{\theta}} = -\frac{e_t^\theta}{h_t^2} e_t + \frac{h_t^\theta}{h_t^3}(e_t^2 - h_t^2) \tag{8}$$

$$\frac{\partial \ln L_t}{\partial \boldsymbol{\psi}} = -\frac{e_t^\psi}{h_t^2} e_t + \frac{h_t^\psi}{h_t^3}(e_t^2 - h_t^2) \tag{9}$$

where $e_t^\vartheta = \partial e_t / \partial \boldsymbol{\vartheta}$ and $h_t^\vartheta = \partial h_t / \partial \boldsymbol{\vartheta}$ are vectors of dimensions corresponding to the numbers of elements in the $\boldsymbol{\theta} = (\theta_0; \theta_1^+, \ldots, \theta_q^+; \theta_1^-, \ldots, \theta_q^-)'$ and $\boldsymbol{\psi} = (\alpha_0; \alpha_1^+, \ldots, \alpha_Q^+; \alpha_1^+, \ldots, \alpha_Q^-;$ $\beta_1, \ldots, \beta_p; \gamma_1, \ldots, \gamma_p)'$ vectors. The gradient expressions are obviously weighted averages of the residuals for the conditional mean and conditional variance. It follows directly that practically simplifying, stepwise estimation procedures based on the block-diagonality (for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$) of the information matrix are not available in this case.

An estimator of the covariance matrix is most easily obtained from the outer product of the gradients, i.e.

$$Cov(\hat{\boldsymbol{\eta}}) = \left[\sum_{t=r}^T \frac{\partial \ln L_t}{\partial \boldsymbol{\eta}} \frac{\partial \ln L_t}{\partial \boldsymbol{\eta}'} \right]^{-1} \tag{10}$$

where $\boldsymbol{\eta} = (\boldsymbol{\theta}', \boldsymbol{\psi}')'$.

Hypotheses of symmetric responses in the conditional mean (cf. Brännäs and De Gooijer, 1994), the conditional variance, or in both jointly may be formulated as linear restrictions on the $\boldsymbol{\eta}$ vector, i.e. as $\boldsymbol{\eta}_0 = \mathbf{R}\boldsymbol{\eta} = \mathbf{0}$. Likelihood ratio tests are easy to apply in practice. Given the estimates and the covariance matrix estimator Wald testing is also quite straightforward. An LM test against asymmetry in the mean function was given and evaluated by Brännäs, De Gooijer and Teräsvirta (1998). Here, we give a new LM test against conditional heteroskedasticity of the asQGARCH(1;1,1) type. The conditional mean specification is the asMA(q) model in (2). Under H_0 of no conditional heteroskedasticity, i.e. when $\boldsymbol{\eta}_0' = \boldsymbol{\psi}_{*0}' = (\alpha^+ = 0, \alpha^- = 0, \beta = 0,$ $\gamma = 0)$ so that $h_t^2 = \alpha_0$, the (4×1) final set of likelihood equations takes the form

$$\ell_{\psi*0} = \frac{\partial}{\partial \boldsymbol{\psi}_*} \sum_{t=q+1}^T \ln L_{t|H_0} = \frac{2}{\alpha_0} \sum_{t=q+1}^T \left(\frac{1}{2\alpha_0} \mathbf{x}_t'(e_t^2 - \alpha_0) - \frac{\partial e_t}{\partial \boldsymbol{\psi}_{*0}} e_t \right)$$

where $\mathbf{x}_t' = \partial h_t^2 / \partial \boldsymbol{\psi}_{*0}$ with $\mathbf{x}_t = (u_{t-1}^+, u_{t-1}^-, u_{t-1}^2, \alpha_0)$ evaluated under H_0. The α_0 is replaced by $\hat{\sigma}^2$ based on the asMA model. The variance of $\ell_{\psi*0}$ is obtained as the (4×4) lower part of the diagonal block of the inverted information matrix, i.e. from $\mathbf{I}_{\psi\psi}^{-1} = (\mathbf{I}_{\psi\psi} - \mathbf{I}_{\psi\theta} \mathbf{I}_{\theta\theta}^{-1} \mathbf{I}_{\theta\psi})^{-1}$. The general expression for the test statistic is then

$$\mathrm{LM} = \ell_{\psi*0}' \mathbf{I}_{\psi_*\psi_*}^{-1} \ell_{\psi*0} \overset{a}{\sim} \chi^2(4).$$

In Appendix B we give the details for evaluating the blocks of the information matrix.

5 Empirical results

For our empirical analysis, we use the daily returns introduced in Section 1. In the next four subsections we discuss various issues related to the estimation, evaluation, interpretation, and forecasting of the models fitted to this series.

Table 1: Sample autocorrelation functions for the series Y_t and Y_t^2 (LB is the Ljung-Box statistic at ten lags).

Series					Lag						LB
	1	2	3	4	5	6	7	8	9	10	
Y_t	0.051	-0.019	-0.031	-0.018	0.030	-0.002	-0.024	-0.009	-0.003	0.006	29.31
Y_t^2	0.091	0.161	0.080	0.020	0.145	0.029	0.011	0.045	0.038	0.009	331.4

5.1 Estimation and evaluation

To model the series $\{Y_t\}$ in terms of the conditional mean and variance, the adopted strategy is to start with the conditional mean function and choose the "best" fitted asMA(q) model on the basis of the minimum value of AIC subject to the condition that the residuals are not serially correlated. Next, in a second step, the asMA(q) model is augmented with an asQGARCH(1;1,1) model for conditional heteroskedasticity.

Table 1 shows values of the sample autocorrelation function (sacf) for both the Y_t and Y_t^2 series. On comparing the sacf of Y_t with the 95% asymptotic confidence bands $\pm 1.96/\sqrt{4956} = \pm 0.028$ the autocorrelations of Y_t at lags 1 and 5 stand out as large in absolute value. The Ljung-Box (LB) statistic for ten lags indicates jointly, statistically significant autocorrelations. Wecker (1981) has shown that the asMA(q) model can generate series with no manifest autocorrelation structure. Hence, it is reasonable to include also higher order lags in the asMA model. The LB statistic for the Y_t^2 series indicates that higher order moment, temporal dependencies are also present in the series. Note, however, that the LB statistic is incapable of detecting asymmetries in neither the conditional mean nor the conditional variances.

Table 2 summarizes the maximum likelihood estimates of the various asMA-asQGARCH models fitted to the series of returns. The first column gives estimation results for the best fitted pure asMA model whereas the remaining columns are for the combined asMA-asQGARCH model. One sees from column one that all parameters in the asMA model are significantly different from zero. Specifically, there is strong evidence for asymmetric responses to negative and positive shocks at the same lags. Also there are different lags coming into the positive and negative lag MA polynomials. The first two parameters (θ_1^+ and θ_2^+) are positive so that positive shocks during the previous two days have incremental impacts on returns. A positive shock three days ago (θ_3^+) has a negative and rather large effect on returns. There can only be a positive response in returns to shocks at $t-2$, since θ_2^- is negative and θ_2^+ is positive. Shocks at $t-3$ can only have a negative impact as θ_3^- is positive. A Wald test of symmetry rejects the hypothesis ($p < 0.001$) so that a symmetric response to shocks at the included lags can be rejected.

On the basis of the above findings the average response lags to shocks (good or bad news) can be calculated. For the asMA(5) model, giving rise to 2^5 parameter combinations, the average response lag is defined as $\sum_{i=0}^{5} i w_i / \sum_{i=0}^{5} w_i$, where $w_i = |\theta_i^+|$ or $w_i = |\theta_i^-|$ with $\theta_0^\pm = 1$. The values of this quantity ranges between 0.32 and 0.82 days, i.e. the returns are on average affected by within the day news. Note that the calculated LB values indicate that the residuals, $\hat{u}_t$, are uncorrelated up to ten lags. However, the LB statistic is clearly significant for the squared residuals, $\hat{u}_t^2$, and their sacf fails to die out quickly. A Dickey-Fuller test is employed to test a
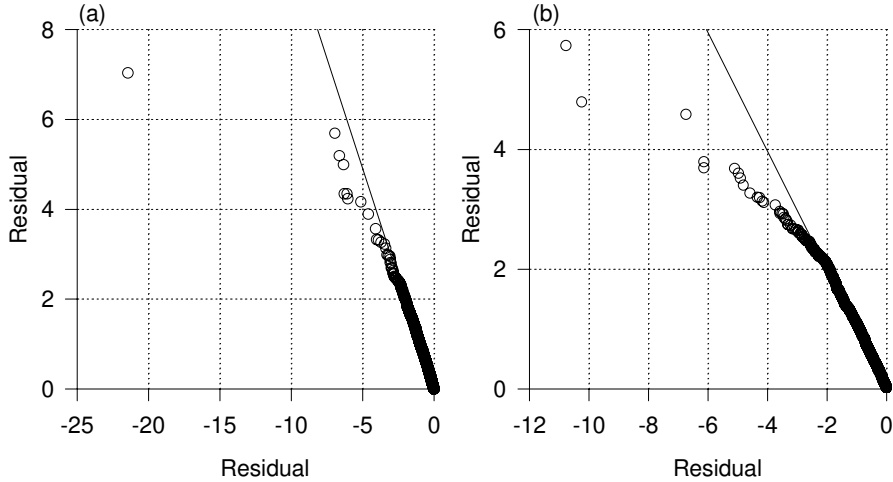
Figure 3: Graphical symmetry check of the residual distribution for the asMA model. The graphs gives size ordered residuals above median (descending ones along y-axis) versus smallest residuals below median (ascending ones along x-axis). Symmetry is supported if observations fall on the indicated line with minus one slope. In (a) residuals for asMA model and in (b) residuals for asMA-asQGARCH model.

unit root hypothesis. This hypothesis is clearly rejected ($p < 0.001$), providing some indication that the parameter γ in the asQGARCH(1;1,1) model is likely to be less than one. The asMA residuals are negatively skewed (-2.91) and leptokurtic (68.79). To give some additional insight into the nature of the asymmetry in the residual distribution, Figure 3.a shows a plot of the largest residual against the smallest, the second largest against the next to smallest, and so on. For a symmetric distribution we expect points to fall on a line of unit negative slope. For most values of the residual this holds quite well, and asymmetry is therefore due to only a few residuals, and these are all confined to the tails of the distribution. Thus a model that is successful in fitting the most extreme observations could well bring the residual distribution closer to a symmetric distribution with thinner tails. The LM test against conditional variance of the asQGARCH(1;1,1) type clearly rejects a constant conditional variance (LM = 345.5, $p = 0$).

Table 2 columns 2-6, show parameter estimates of the considered asMA-asQGARCH(1;1,1) models. From the values of the log-likelihood function, $\ln L$, it is apparent that the pure asMA model can be rejected against any of the reported asMA-asQGARCH specifications. Moreover, we see that the specification containing all lagged u_{t-1} variables cannot be rejected against any of the rival and nested specifications. The preferred model, given in the final column of Table 2, contains an additional asMA parameter θ_1^-. The standardized residuals, $\hat{u}_t/\hat{h}_t$, of this model have skewness and kurtosis equal to respectively -0.68 and 6.93. Both measures are substantially smaller than the corresponding values for the skewness and kurtosis for the asMA model residuals (see also Figure 3.b and note the differences in scales). In general, it may be

9

Table 2: Parameter estimates for the conditional mean (asMA) and the joint conditional mean and variance [asMA-asQGARCH(1;1,1)] models (standard errors in parentheses).

Parameter		Model-type					
	asMA	asMA-asQGARCH					
θ_1^+	0.113	0.116	0.130	0.127	0.123	0.122	0.109
	(0.010)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.014)
θ_2^+	0.072	0.048	0.052	0.054	0.052	0.057	0.056
	(0.011)	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)
θ_3^+	-0.128	-0.060	-0.039	-0.043	-0.050	-0.044	-0.045
	(0.011)	(0.013)	(0.014)	(0.014)	(0.013)	(0.012)	(0.014)
θ_1^-							0.033
							(0.014)
θ_2^-	-0.081	0.015	0.009	0.013	0.015		
	(0.003)	(0.012)	(0.012)	(0.012)	(0.012)		
θ_3^-	0.075	0.019	0.004	0.012	0.017		
	(0.006)	(0.015)	(0.015)	(0.015)	(0.015)		
θ_4^-	-0.035	-0.051	-0.050	-0.046	-0.047	-0.049	-0.051
	(0.010)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
θ_5^-	0.074	-0.022	-0.024	-0.021	-0.020		
	(0.004)	(0.012)	(0.012)	(0.012)	(0.012)		
θ_0	0.031	0.005	-0.004	-0.005	-0.001	-0.007	0.004
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.008)	(0.009)
α_0	-0.199	0.003	0.034	0.019	0.010	0.010	0.010
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
α^+			-0.106			-0.032	-0.033	-0.032
			(0.005)	-0.064		(0.006)	(0.006)	(0.006)
α^-			-0.113		(0.003)	-0.095	-0.094	-0.096
			(0.005)			(0.006)	(0.006)	(0.006)
β		0.046	0.109	0.076	0.059	0.059	0.059
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
γ		0.906	0.892	0.901	0.905	0.905	0.905
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\sigma}^2$	0.820	0.828	0.829	0.0828	0.828	0.827	0.827
$\ln L$	-3964.2	-2511.9	-2530.0	-2511.8	-2508.4	-2509.8	-2508.3
$\mathrm{LB}_{10}(\hat{u}_t/\hat{h}_t)$	4.81	8.28	6.92	6.81	7.25	7.95	5.90
$\mathrm{LB}_{10}(\hat{u}_t^2/\hat{h}_t^2)$	184.8	7.11	4.96	5.56	6.24	6.07	6.23
Skewness	-2.91	-0.71	-0.59	-0.65	-0.68	-0.67	-0.68
Kurtosis	68.79	7.26	6.01	6.51	6.86	6.89	6.93

Note: $\hat{\sigma}^2$ is the variance estimator of the unstandardized residuals.

noted that all fitted asMA-asQGARCH specifications have residual distributions which are much closer to the normal distribution as compared to the distribution of the residuals of the pure asMA model. Wald tests for symmetry in the fitted asMA-asQGARCH models rejected the null hypothesis in all cases (p-values < 0.001). An LR test of the QGARCH model of column four rejects this model against the models corresponding to columns 5-7. The standard errors of the table are based on the covariance matrix in (10). Using the consistent covariance matrix estimator of Bollerslev and Wooldridge (1992), which is appropriate for non-normal $\{\varepsilon_t\}$, gives almost identical standard errors.

Comparing column 1 with columns 2-7 in Table 2, some effects due to the augmentation of the asMA model with the asQGARCH(1;1,1) specification are worth mentioning. First, most fitted asMA-asQGARCH models yield substantially smaller parameter estimates at most time lags. The θ_1^+ and θ_4^- estimates are the only ones to be of the same or larger sizes. Further, for the best asMA-asQGARCH model in column seven the average response lag ranges between 0.03 and 0.44 days, which is substantially smaller than the range of averaged response lags for the asMA model. The largest mean lag is found for positive shocks during lags one to three and a negative shock at lag four. The smallest mean lag arises when the signs of the shocks are reversed. Finally, except for the parameters dropped from the fitted models, all standard errors are rather constant across different models.

5.2 Interpretation

A characteristic feature of the fitted asMA-asQGARCH model in columns 2-3 and 5-7 of Table 2 is that a positive shock has a negative effect on conditional heteroskedasticity through the negative values of the parameter estimates of α^+. Note, however, that the squared shock has a larger parameter estimate so that for positive shocks larger than about 0.5 the total effect is an enhancing one. A negative shock has a risk enhancing effect for all specifications. In Figure 4 the lag one responses are plotted for the model in column seven. As is obvious the quadratic term dominates for larger absolute values of u_{t-1}. The plots also indicate that h_t^2 is always positive though no restrictions were employed on parameters during estimation. The asymmetric effects are manifest both in the sense that the curves are not centered at zero and in the sense of larger values for negative values on u_{t-1}. For the model with all lagged u_{t-1} included the negative innovation increases volatility about 1.3 times more than a positive innovation of the same size.

Another way of depicting the effect of asymmetry is through a dynamic simulation of the conditional heteroskedasticity specification. Since the estimates of γ are about 0.9 a shock will have a long lasting effect on h_t^2. In Figure 5 we see that a negative shock prolongs the period with enhanced risk for the asMA-asQGARCH model in the last column of Table 2, while a positive shock though prolonging has a much smaller effect.

In Figure 6 the observations for September 1988 – December 1999 are displayed with point-wise 95 per cent confidence bands. In common with other conditional variance models the present one quickly adjusts to a higher variance level, with a subsequent slow return to the average variance level of about 0.8. A close view suggests that a large negative observation is followed by a larger variance in the next period, than would be the case for an equally large positive observation. Note also that the the limits of the confidence intervals are rather symmetrically located around zero. This indicates the small fitted values of the model (following from the small parameter estimates).
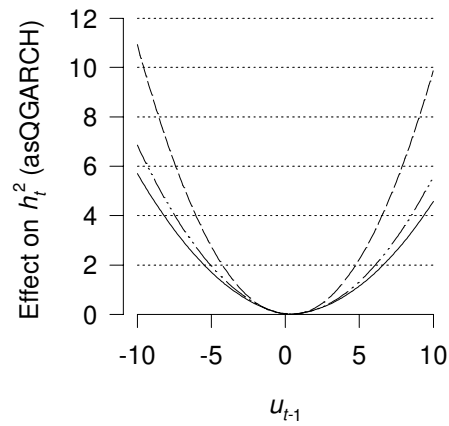
Figure 4: Conditional variance responses to shocks in u_{t-1}. Model without u_{t-1}^{+} (solid line), model without u_{t-1}^{-} (dashed line), and model with all lagged u_{t-1} included (dot-dashed line).
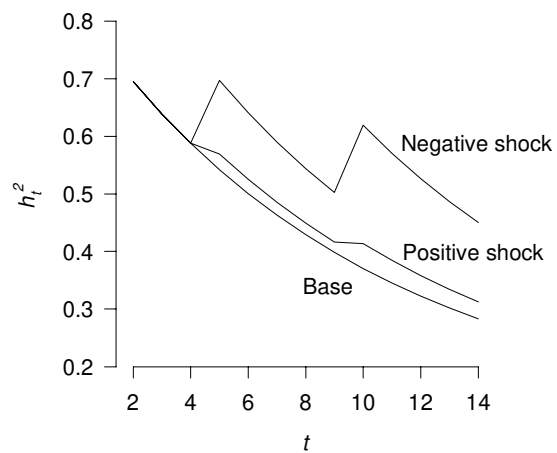

Figure 5: Effects of a unit positive shock at time $t = 4$ and $t = 11$ (Positive shock), a negative shock at the same time points, and a base case with $u_{t-1} = 0$, for all t.
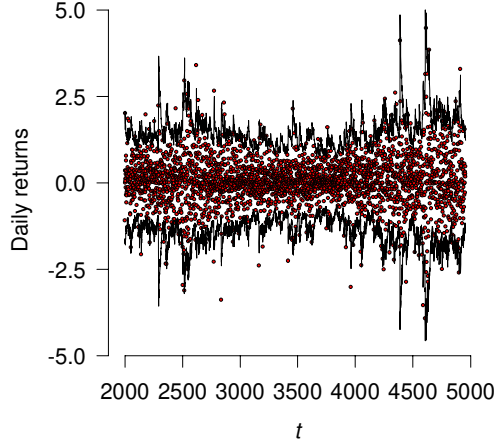
Figure 6: Observed series (dots) for September 1988 – December 1999 with pointwise confidence intervals, $\hat{Y}_t \pm 1.96 h_t$.

5.3 Forecasting

In this subsection we evaluate the forecasting performance of the best asMA-asQGARCH model, both in terms of conditional mean forecasting and in terms of risk forecasting. The H-step-ahead predictor for asMA(q), $Y_{T+H|T} = E(Y_{T+H}|Y_T, Y_{T-1}, \ldots)$, given observations up to and including time T, is given by

$$
\begin{aligned}
Y_{T+H|T} &= \theta_0 + \sum_{i=H-1}^{q} (\theta_{i+1}^+ u_{T+2-i}^+ + \theta_{i+1}^- u_{T+2-i}^-) \\
&\quad + \sum_{i=1}^{H-1} [\theta_i^+ E(u_{T+i}^+|Y_T, Y_{T-1}, \ldots) + \theta_i^- E(u_{T+i}^-|Y_T, Y_{T-1}, \ldots)], \qquad \text{if} \quad H \leqslant q \\
Y_{T+H|T} &= \theta_0 + \sum_{i=1}^{H-1} [\theta_i^+ E(u_{T+i}^+|Y_T, Y_{T-1}, \ldots) + \theta_i^- E(u_{T+i}^-|Y_T, Y_{T-1}, \ldots)], \qquad \text{if} \quad H > q.
\end{aligned}
$$

Since $E(u_{T+i}^\pm|Y_T, Y_{t-1}, \ldots) = E(\varepsilon_{T+i}^\pm)E(h_{T+i}|Y_T, Y_{T-1}, \ldots) \neq 0$ forecast expressions close to the conventional MA results do not emerge. The values of $E(\varepsilon_{T+i}^\pm)$, which are both time invariant, may be estimated from the standardized residual by calculating the sample averages of values falling above zero (for ε_t^+) or below zero (for ε_t^-). An analytical expression for $E(h_{T+i}|Y_T, Y_{T-1}, \ldots)$ is difficult to obtain as h_t is obviously a square root function. An approximative variant is to use the h_t^2 specification to forecast future conditional variances, and then to simply take their square roots. In the prediction of future values of h_t^2 we face the similar problem of having to find $E(u_{T+i}^\pm|Y_T, Y_{t-1}, \ldots)$ as well as $E(u_{T+i}^2|Y_T, Y_{t-1}, \ldots) = E(\varepsilon_{T+i}^2)E(h_{T+i-1}^2|Y_T, Y_{t-1}, \ldots) = E(h_{T+i-1}^2|Y_T, Y_{t-1}, \ldots)$. For the former we use the same averages as for the conditional mean predictor, while for the latter we use the predicted h_{T+i-1}^2

13

rather than its expectation. We will introduce the label "Naive" to indicate a forecast where all future $\varepsilon_t^{\pm}$, ε_t and ε_t^2 are set equal to zero. The label "Approx" identifies the approximative predictor outlined above. For the latter forecast, we estimate the values of $E(\varepsilon_{T+i}^+)$ and $E(\varepsilon_{T+i}^-)$ to 0.74 and -0.71, respectively, and $E(h_{T+i-1}^2)$ to 0.80.

The conditional mean forecast performance is compared to that of the pure asMA model and to a no-change forecast. The models are re-estimated for the daily returns covering the period January 1, 1981–October 8 (Friday), 1999 (4896 observations). Subsequently the Naive and Approx forecasts four days ahead are calculated. Next, the forecast values are compared to the actual observations. Adding one observation at a time, re-estimating and forecasting we continue to the end of the series, i.e. we employ a rolling forecasting framework. In total, this approach gives rise to $n = 60$ one-step-ahead forecasts, $n = 59$ two-step-ahead forecasts, and so on. Before reporting the forecasting results it is worth mentioning that for an efficient financial market one should not expect to beat the no-change forecast. If it was possible this indicates an inefficient market and would suggest that the evaluated model is a money-maker. Equally strong arguments do not apply to the risk forecast.

Table 3 gives the performance measures for conditional mean forecasting. There are small differences in terms of mean performance. On comparing forecast variances, the no-change forecast has the by far largest one. The variance of the asQGARCH model is marginally smaller than that of the asMA model for all forecast horizons H.

Figure 7 gives the one-step-ahead risk or conditional heteroskedasticity forecasts over the evaluation period. The mean of the forecasts is 0.92 and the median is 0.78, based on $n = 60$ observations. This may be compared to the variance of the time series for the whole period 1981-1999 of about 0.83 or to the variance in the period of 0.89. As the risk distribution is positively skewed (0.69) the mean of the risk forecasts is not far off target. The risk forecasts vary between 0.38 (end of December, 1999) and 1.79 (October, 1999). Over the complete period the risk forecasts have a significant negative trend. Also included in Figure 7 are two alternative risk measures formed as the moving variance over the previous 10 and 15 days, respectively. The curves are positively correlated ($r \geqslant 0.9$), though the risk forecast series varies much less than the other two.

5.4   Discussion

To investigate the robustness of the reported estimates several alternative specifications for capturing asymmetry in the conditional variance were fitted to the data. Here some of the estimation results are briefly discussed.[4] For instance, we adapted the Glosten *et al.* (1993) specification by replacing the term βu_{t-1}^2 in the asQGARCH(1;1,1) part of the asMA-asQGARCH model by the term $\beta_1 u_{t-1}^2 + \beta_2 (u_{t-1}^+)^2$. Estimation gave a significant (positive) additional parameter value for β_2. However, at the 5% critical level, it also resulted in a positive though insignificant value of α^+. More interestingly, the introduction of asymmetry in the quadratric term has hardly no effect on the LB-test for residual autocorrelation nor on the distributional properties of the residuals.

In trying to expand the asQGARCH part of our model with higher order lags, we found empirically that most parameters would require restrictions to be introduced during the estimation phase to render h_t^2 positive for all t. Only for u_{t-2}^- did we find a significant contribution

---

[4]Detailed results are available upon request from the authors.

Table 3: Conditional mean forecasting performance of competing models based on forecast errors.

| Model	Measure	Horizon H			
		1	2	3	4
asQGARCH	Mean	0.04	0.07	0.12	0.14
(Naive)	Median	0.01	0.07	0.08	0.06
	Correct sign (%)	52	59	64	54
	Variance	0.87	0.90	0.86	0.81
asQGARCH	Mean	0.04	0.02	0.03	0.08
(Approx)	Median	0.01	0.01	-0.02	-0.01
	Correct sign (%)	53	56	53	54
	Variance	0.87	0.90	0.85	0.80
asMA	Mean	0.06	0.09	0.14	0.07
	Median	-0.02	0.04	0.09	0.02
	Correct sign (%)	45	47	48	49
	Variance	0.88	0.90	0.94	0.94
No-change	Mean	-0.01	-0.01	0.03	0.06
	Median	-0.03	0.04	0.03	0.23
	Correct sign (%)	60	47	48	49
	Variance	1.53	1.88	2.25	1.58
	Nr of forecasts (n)	60	59	58	57
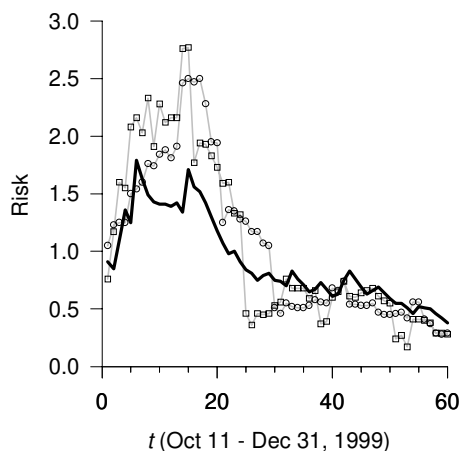
Figure 7: One-step-ahead risk forecasts (black line) with lines for variances over the previous 10 days (grey line, circle) and previous 15 days (grey line, square).

without employing restrictions, but again little effect was found on the values of the parameters and the test statistics reported in Table 2. Nelson (1991) and others have used conditional moment tests to test aspects of their model specifications. Our reported LB statistics and significant skewness and kurtosis statistics are alternative statistics answering the same questions. Bollerslev and Wooldridge (1992) showed that the covariance matrix estimator (10) is not consistent if normality can be rejected. Empirically we found that their estimator and that in (10) gave almost identical results. In addition to asQGARCH specifications we have also estimated some asymmetric versions of the EGARCH model (cf. Nelson, 1991). Using the AIC or $\ln L$ to discriminate between models, all EGARCH specifications are inferior to the asQGARCH specifications of Table 2.

The conditional mean of the NYSE index returns responds more slowly to positive shocks, though the response is stronger than for negative shocks. For the conditional variance the impact of negative shocks is greater. The asymmetric and positive responses in volatility to past shocks is in line with, e.g., Nelson (1991). To study the previously found inverse relationship between autocorrelation and volatility (LeBaron, 1992), consider Figure 2 which is based on estimated versions of (6)-(7) to obtain the autocorrelation coefficient at lag one, $\rho_1 = \gamma_1/\gamma_0$. Variation in h_t^2 is obtained by varying the constant term α_0 ($E(h_t^2)$ is increasing from 0.47 to 7.31). Figure 2 is based on a generated series of length $T = 100\,000$ to estimate the unknown expectations. As the graph manifests a negative slope, the estimated model provides additional support to the previously found inverse relationship.

16

6   Some Concluding Remarks

In this article we presented a general methodology for simultaneously modeling and estimating asymmetries in the conditional mean and the conditional variance of observed time series. We also introduced a framework for testing asymmetries in the conditional mean and the conditional variance, separately or jointly. Finally, we derived some theoretical properties of the proposed asMA-asQGARCH model and showed how the model can be used in a forecasting context. Using these results, potential asymmetries in (economic) time series data can be investigated in a unified and coherent way. In particular, we explored asymmetries in the NYSE composite daily returns. The empirical evidence suggests that both the conditional mean and the conditional variance respond asymmetrically to past information. Specifically, the daily returns are on average affected by within the day news. Furthermore, we noted that positive shocks (good news) have a negative effect on conditional heteroskedasticity. On the other hand negative shocks (bad news) increase volatility by about 1.3 times more than positive innovations of the same magnitude. This behaviour is consistent with a partial adjustment price model where bad news are incorporated faster into current market prices than good news. It implies that the cost of failing to adjust prices downward is higher.

In a recent paper of Harvey and Siddique (1999) asymmetries introduced through a skewness mechanism were found. Such a model may reduce the importance of conditional variance asymmetry. For an extension of the model of this paper a density that incorporates, at least, three non-zero moments is required.

Appendix A: Moments of h_t^2

We begin by introducing notation. Let $\gamma_{i,j}^{bh,c} = E[(b_t h_t)^i (c_t)^j]$, $\gamma_i^{bh} = E[(b_t h_t)^i]$, and $\gamma_i^{ch^2} = E[(c_t h_t^2)^i]$ with i and j positive integers. Consider the following, slightly more general, specification of the asQGARCH(1;1,1) model

$$h_t^2 = \alpha_0 + b(\varepsilon_{t-1})h_{t-1} + c(\varepsilon_{t-1})h_{t-1}^2 \tag{11}$$

where $b_t = b(\varepsilon_t)$ and $c_t = c(\varepsilon_t)$ are well-defined functions of ε_t.

Lemma: *Consider the* asQGARCH *model (11) and (4). Assume that* $\gamma_{(c)^m} < 1$. *Then the* m-*th* $(m = 1, 2, \ldots)$ *unconditional moment of* h_t^2 *can be expressed recursively as*

$$
\begin{aligned}
E[h_t^{2m}] &= \frac{1}{1 - \gamma_m^c} \sum_{j=1}^{m} \binom{m}{j} \left\{ \left(\gamma_{j,m-j}^{bh,c} + \frac{j!}{(j-1)!} \alpha_0 \gamma_{j-1,m-j}^{bh,c} \right) E[h_t^{2(m-j)}] \right. \\
&\quad \left. + \sum_{i=2}^{j} \binom{j}{i} \alpha_0^i \gamma_{j-i}^{bh} \gamma_{m-j}^{ch^2} \right\}.
\end{aligned}
$$

Proof: Raising (11) to the power m yields

$$h_t^{2m} = (c_{t-1} h_{t-1}^2)^m + \sum_{j=1}^{m} \binom{m}{j} \left\{ (b_{t-1} h_{t-1})^j + \sum_{i=1}^{j} \binom{j}{i} \alpha_0^i (b_{t-1} h_{t-1})^{j-i} \right\} (c_{t-1} h_{t-1}^2)^{m-j}.$$

Taking the unconditional expectation of h_t^2 yields

$$
\begin{aligned}
E[h_t^{2m}] &= E[c_{t-1}^m] E[(h_{t-1}^2)^m] + \sum_{j=1}^{m} \binom{m}{j} \left\{ E[(b_{t-1} h_{t-1})^j c_{t-1}^{m-j}] \right. \\
&\quad \left. + \frac{j!}{(j-1)!} \alpha_0 E[(b_{t-1} h_{t-1})^{j-1} c_{t-1}^{m-j}] \right\} E[(h_{t-1}^2)^{m-j}] \\
&\quad + \sum_{j=1}^{m} \binom{m}{j} \sum_{i=2}^{j} \binom{j}{i} \alpha_0^i E[(b_{t-1} h_{t-1})^{j-i}] E[(c_{t-1} h_{t-1}^2)^{m-j}]. \tag{A.1}
\end{aligned}
$$

Applying (11) to the first term on the right-hand side of (A.1) one obtains

$$
\begin{aligned}
E[h_t^{2m}] &= E[c_{t-1}^m] E\left\{ (c_{t-2} h_{t-2}^2)^m + \sum_{j=1}^{m} \binom{m}{j} \left\{ (b_{t-2} h_{t-2})^j \right. \right. \\
&\quad \left. \left. + \sum_{i=1}^{j} \binom{j}{i} \alpha_0^i (b_{t-2} h_{t-2})^{j-i} \right\} (c_{t-2} h_{t-2}^2)^{m-j} \right\} \\
&\quad + \sum_{j=1}^{m} \binom{m}{j} \left\{ E[(b_{t-1} h_{t-1})^j c_{t-1}^{m-j}] \right. \\
&\quad \left. + \frac{j!}{(j-1)!} \alpha_0 E[(b_{t-1} h_{t-1})^{j-1} c_{t-1}^{m-j}] \right\} E[(h_{t-1}^2)^{m-j}] \\
&\quad + \sum_{j=1}^{m} \binom{m}{j} \sum_{i=2}^{j} \binom{j}{i} \alpha_0^i E[(b_{t-1} h_{t-1})^{j-i}] E[(c_{t-1} h_{t-1}^2)^{m-j}].
\end{aligned}
$$

Further recursion gives the unconditional expectation of h_t^2

$$
\begin{aligned}
E[h_t^{2m}] &= (\gamma_m^c)^n E[(h_{t-n}^{2m}] + \sum_{j=1}^{m} \binom{m}{j} \Big\{ E[(b_{t-n}h_{t-n})^j c_{t-n}^{m-j}] \\
&\quad + \frac{j!}{(j-1)!}\alpha_0 E[(b_{t-n}h_{t-n})^{j-1} c_{t-n}^{m-j}] \Big\} E[(h_{t-n}^2)^{m-j}] \\
&\quad + \sum_{j=1}^{m} \binom{m}{j} \sum_{i=2}^{j} \binom{j}{i} \alpha_0^i E[(b_{t-n}h_{t-n})^{j-i}] E[(c_{t-n}h_{t-n}^2)^{m-j}] \Big\} \prod_{u=0}^{n-1} (E[c_t^m])^u. \ (\text{A.2})
\end{aligned}
$$

Assume that the process started at some finite value infinitely many periods ago. Then letting $n \to \infty$ in (A.2) yields

$$
\begin{aligned}
E[h_t^{2m}] &= \frac{1}{1-\gamma_m^c} \sum_{j=1}^{m} \binom{m}{j} \Big\{ \Big(\gamma_{j,m-j}^{bh,c} + \frac{j!}{(j-1)!}\alpha_0 \gamma_{j-1,m-j}^{bh,c} \Big) \\
&\quad \times E[h_t^{2(m-j)}] + \sum_{i=2}^{j} \binom{j}{i} \alpha_0^i \gamma_{j-i}^{bh} \gamma_{m-j}^{ch^2} \Big\}
\end{aligned}
$$

if and only if $\gamma_m^c < 1$.

Remark: Explicit expressions for $E[h_t^m]$ in the case of asQGARCH(1;1,1) model follow directly from the above Lemma. In particular, setting $b_{t-1} = \alpha^+ \varepsilon_{t-1}^+ + \alpha^- \varepsilon_{t-1}^-$ and $c_{t-1} = \gamma + \beta \varepsilon_{t-1}^2$ in the (11). Then, assuming that $\gamma_m^c = (\gamma + \beta \nu_2)^m < 1$ ($m = 1, 2$), the first two moments of h_t^2 are respectively given by

$$
\begin{aligned}
E[h_t^2] &= \frac{1}{1-\gamma_1^c} \{ \alpha_0 + \gamma_1^{bh} \} \\
E[h_t^4] &= \frac{1}{1-\gamma_2^c} \Big\{ \alpha_0^2 + \gamma_2^{bh} + 2\alpha_0 \gamma_1^{bh} + 2(\gamma_{1,1}^{bh,c} + \alpha_0 \gamma_1^c) E[h_t^2] \Big\}
\end{aligned}
$$

where $\gamma_m^{bh} = E[(\alpha^+ u_t^+ + \alpha^- u_t^-)^m]$, $\gamma_{1,1}^{bh,c} = E[(\alpha^+ u_t^+ + \alpha^- u_t^-)(\gamma + \beta \varepsilon_t)]$, and where we dropped the subscripts in the parameters of the asQGARCH(1;1,1) model.

Appendix B: Partial derivatives for the LM test statistic

The partial derivatives are given for evaluating the LM test under the null of an asMA(q) conditional mean specification is the model in (2) against the asMA-asQGARCH(1;1,1) as given in (2)-(4). Under H_0 of no conditional heteroskedasticity, $\boldsymbol{\psi}_0' = (0,0,0,0)$ and $h_t^2 = \alpha_0$, and the (4×1) score vector is $(2/\alpha_0) \sum_{t=q+1}^{T} (\mathbf{x}_t'(e_t^2 - \alpha_0)/2\alpha_0 - e_t \partial e_t / \partial \boldsymbol{\psi}_0)$, where $\mathbf{x}_t' = \partial h_t^2 / \partial \boldsymbol{\psi}_0$ with $\mathbf{x}_t = (u_{t-1}^+, u_{t-1}^-, u_{t-1}^2, \alpha_0)$. The variance of the score is taken from the appropriate diagonal block of the inverted information matrix, i.e. $\mathbf{I}_{\psi\psi}^{-1} = (\mathbf{I}_{\psi\psi} - \mathbf{I}_{\psi\theta}\mathbf{I}_{\theta\theta}^{-1}\mathbf{I}_{\theta\psi})^{-1}$. We use the outer product of gradients for the matrix blocks of $\mathbf{I}$, which then are estimated using

$$
\begin{aligned}
\hat{\mathbf{I}}_{\theta\theta} &= \frac{4}{\alpha_0^2} \sum_{t=q+1}^{T} e_t^2 \frac{\partial e_t}{\partial \boldsymbol{\theta}} \frac{\partial e_t}{\partial \boldsymbol{\theta}'} \\
\hat{\mathbf{I}}_{\theta\psi} &= \sum_{t=q+1}^{T} \Big(\frac{4}{\alpha_0^2} e_t^2 \frac{\partial e_t}{\partial \boldsymbol{\theta}} \frac{\partial e_t}{\partial \boldsymbol{\psi}'} - \frac{2}{\alpha_0^3} \frac{\partial e_t}{\partial \boldsymbol{\theta}} \mathbf{x}_t e_t (e_t^2 - \alpha_0) \Big) \\
\hat{\mathbf{I}}_{\psi\psi} &= \sum_{t=q+1}^{T} \Big(\frac{4}{\alpha_0^2} e_t^2 \frac{\partial e_t}{\partial \boldsymbol{\psi}} \frac{\partial e_t}{\partial \boldsymbol{\psi}'} - \frac{2 e_t (e_t^2 - \alpha_0)}{\alpha_0^3} \Big(\frac{\partial e_t}{\partial \boldsymbol{\psi}} \mathbf{x}_t + \mathbf{x}_t' \frac{\partial e_t}{\partial \boldsymbol{\psi}'} \Big) + \frac{(e_t^2 - \alpha_0)^2}{\alpha_0^4} \mathbf{x}_t' \mathbf{x}_t \Big)
\end{aligned}
$$

and evaluated under H_0.

The required first order partial derivatives are, for $k = 1, \ldots q$, calculated recursively by

$$
\begin{aligned}
\frac{\partial e_t}{\partial \theta_k^+} &= \frac{\partial u_t}{\partial \theta_k^+} = -u_{t-k}^+ - \sum_{i=1}^q \left(\theta_i^+ \frac{\partial u_{t-i}^+}{\partial \theta_k^+} + \theta_i^- \frac{\partial u_{t-k}^-}{\partial \theta_k^+} \right) \\
\frac{\partial e_t}{\partial \theta_k^-} &= \frac{\partial u_t}{\partial \theta_k^-} = -u_{t-k}^- - \sum_{i=1}^q \left(\theta_i^+ \frac{\partial u_{t-i}^+}{\partial \theta_k^-} + \theta_i^- \frac{\partial u_{t-k}^-}{\partial \theta_k^-} \right) \\
\frac{\partial e_t}{\partial \theta_0} &= \frac{\partial u_t}{\partial \theta_0} = -1 - \sum_{i=1}^q \left(\theta_i^+ \frac{\partial u_{t-i}^+}{\partial \theta_0} + \theta_i^- \frac{\partial u_{t-k}^-}{\partial \theta_0} \right) \\
\frac{\partial e_t}{\partial \psi_k} &= \frac{\partial u_t}{\partial \psi_k} = - \sum_{i=1}^q \left(\theta_i^+ \frac{\partial u_{t-i}^+}{\partial \psi_k} + \theta_i^- \frac{\partial u_{t-k}^-}{\partial \psi_k} \right)
\end{aligned}
$$

where ψ_k is equal to respectively $\alpha_0, \alpha^+, \alpha^-, \beta$, and γ. In the recursions zeroes are used as initial values.

References

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. and Wooldridge, J.M. (1992), "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews*, 11, 143-172.

Brännäs, K. and De Gooijer, J.G. (1994). "Autoregressive-asymmetric moving average models for business cycle data", *Journal of Forecasting*, 13, 529-544.

Brännäs, K., De Gooijer, J.G. and Teräsvirta, T. (1998), "Testing linearity against nonlinear moving average models", *Communications in Statistics: Theory and Methods*, 27, 2025-2035.

Brännäs, K. and Ohlsson, H. (1999), "Asymmetric time series and temporal aggregation", *Review of Economics and Statistics*, 81, 341-344.

De Gooijer, J.G. and Brännäs, K. (1995), "Invertibility of non-linear time series models", *Communications in Statistics: Theory and Methods*, 24, 2701-2714.

Glosten, L., Jagannathan, R. and Runkle, D. (1993), "On the relation between expected value and the volatility of the nominal excess return on stocks", *Journal of Finance*, 48, 1779-1801.

Harvey, C.R. and Siddique, A. (1999), "Autoregressive conditional skewness", *Journal of Financial and Quantitative Analysis*, 34, 465-487.

Hentschel, L. (1995), "All in the family: Nesting symmetric and asymmetric GARCH models", *Journal of Financial Economics*, 39, 71-104.

Koutmos, G. (1999), "Asymmetric index stock returns: evidence from the G-7", *Applied Economics Letters*, 6, 817-820.

LeBaron, B. (1992), "Some relations between volatility and serial correaltions in stock market returns", *Journal of Business*, 65, 199-219.

Lee, Y.N. and Li, W.K. (1998), "On smooth transition double threshold models", Dept. of Statistics, University of Hong Kong, Research Report No. 198.

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

Lundbergh, S. and Teräsvirta, T. (1998), "Modelling economic high-frequency time series with STAR-STGARCH models", Working paper 291, Stockholm School of Economics.

Nelson, D. (1991), "Conditional heteroskedasticity in asset returns: A new approach", *Econometrica*, 59, 347-370.

Pagan, A. (1996), "The econometrics of financial markets", *Journal of Empirical Finance*, 3, 15-102.

Sentana, E. (1995), "Quadratic ARCH models", *Review of Economic Studies*, 62, 639-661.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical Systems Approach*, Oxford: Oxford University Press.

Wecker, W.E. (1981), "Asymmetric time series", *Journal of the American Statistical Association*, 76, 16-21.

Zakoïan, J.-M. (1994), "Threshold heteroskedastic models", *Journal of Economic Dynamics and Control*, 18, 931-955.