

Garch Forecasting Performance under Different Distribution Assumptions

ANDERS WILHELMSSON*

*Swedish School of Economics and Business Administration,
Department of Finance and Statistics, Helsinki, Finland*

ABSTRACT

This paper investigates the forecasting performance of the Garch (1, 1) model when estimated with NINE different error distributions on Standard and Poor's 500 Index Future returns. By utilizing the theory of realized variance to construct an appropriate *ex post* measure of volatility from intra-day data it is shown that allowing for a leptokurtic error distribution leads to significant improvements in variance forecasts compared to using the normal distribution. This result holds for daily, weekly as well as monthly forecast horizons. It is also found that allowing for skewness and time variation in the higher moments of the distribution does not further improve forecasts. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS volatility forecasts; GARCH; intra-day data; realized volatility

INTRODUCTION

During the last few decades we have seen a multitude of different suggestions for how to model the second moment, often referred to as volatility, of financial returns. Among the models that have proven the most successful are the auto-regressive conditional heteroscedasticity (Arch) family of models introduced by Engle (1982) and the models of stochastic variance (SV) pioneered by Taylor (1986). During the last couple of years ARFIMA type modeling of high-frequency squared returns has proved very fruitful.

One of the main purposes of modeling variance is forecasting, which is crucial in many areas of finance such as option pricing, value at risk applications and portfolio selection. Therefore, the out-of-sample forecasting ability should be a natural model selection criterion for volatility models. At least 93 different articles, reviewed by Poon and Granger (2003), have addressed this question. The results are inconclusive. There are several reasons for this; the ranking of models is dependent on, at least, the choice of sample period, sample frequency, forecast horizon, loss functions used and, importantly, the proxy used for the *ex post* variance. The vast majority of variance forecasting articles has used squared daily returns as the proxy for *ex post* variance.

This is, as shown by Andersen and Bollerslev (1998), an unbiased but exceedingly noisy estimator.

* Correspondence to: Anders Wilhelmsson, Swedish School of Economics and Business Administration, Department of Finance and Statistics, PO Box 479, FIN-00101 Helsinki, Finland. E-mail: Anders.wilhelmsson@hanken.fi

While the literature that examines competing variance models is abundant,¹ very little work has been done comparing different distribution assumptions, with the noticeable exceptions of Hamilton and Susmel (1994), Lopez (2001) as well as Franses and Ghijssels (1999). However, none of these three papers has explicitly focused on evaluating Garch forecasting using different error distributions.

This paper adds to the literature in several important directions. Firstly, a very flexible error distribution that nests several well-known distributions including the normal, Student's t and the non-central Student's t distribution, is used. Secondly, a much less noisy proxy of the *ex post* variance is constructed using high-frequency data from Standard and Poor's 500 index future to facilitate meaningful comparison of the forecast results. In addition, the forecast horizon is extended to include 1-, 5- and 20-day forecasts. Furthermore, to gauge whether the forecasting performance of the models is statistically different Diebold and Mariano's (1995) test for equal predictive accuracy will be employed. This test has also been used by Lopez (2001).

In the model estimation, it is shown that allowing for a leptokurtic and skewed return distribution significantly improves the fit of the model. In terms of out-of-sample forecasting performance, allowing for excess kurtosis leads to significant improvements over the normal distribution, whereas allowing for non-centrality does not further enhance forecasts. Nor are the forecasts improved by allowing the kurtosis and skewness to be time varying.

The rest of this article is structured as follows. Prior research is covered in the next section and the methodology for volatility forecasting is discussed in the third section. In the fourth section the return distribution is presented. The data and model used in the empirical study are exhibited in the fifth section. The sixth section displays the results and, lastly, the seventh section concludes.

PRIOR RESEARCH

Previous research on the effects of distribution assumptions on the variance forecasting performance of Garch models is scarce. Articles that do address this question are Hamilton and Susmel (1994), Lopez (2001) and Franses and Ghijssels (1999).

The main contribution of Hamilton and Susmel is to allow for different variance regimes by using a Markov switching Garch model, but the authors also allow the error term to be distributed according to a normal, Student's t or generalized error distribution. Weekly stock return data are used for both in-sample estimation and the construction of *ex post* variance. Forecasts are made for 1, 4 and 8 weeks. When the mean square error (MSE) loss criterion is used, no improvement is found over using constant variance.

However, when logarithmic loss (LL) and absolute LL are used, the Garch model with a t -distribution performs best, followed by the generalized error distribution on the 1-week horizon. Results are not reported for the 4- and 8-week comparison. The reason for this is probably the focus on the Markov switching models.

In the article by Lopez (2001) the main focus is on developing loss functions based on probability scoring rules but the performance of a GARCH (1, 1) model fitted with the normal, Student's t and generalized error distribution on four daily exchange rate series is also evaluated by more traditional loss functions. The results show that the relative performance of the models in-sample and

¹For a recent review, see Poon and Granger (2003).

out-of-sample are different. This highlights the importance of out-of-sample results as a model selection criterion. The forecasting results are mixed; model rankings change depending on data series and loss function. *Ex post* variance is not computed from high-frequency returns.

Franses and Ghijssels (1999) estimate three Garch models using weekly European stock market data. The first model has the effect of additive outliers removed, the second model has t -distributed errors and the third is a regular Garch model. Interestingly, the performance of the Garch model with t -distributed errors is worst out-of-sample. *Ex post* variance is computed from a weekly return frequency.

DISCUSSION ON FORECAST EVALUATION

Let $r_{(m),t} \equiv p_t - p_{t-1/m}$, where p is the natural logarithm of the assets' price, denote the continuously compounded $1/m$ period return and let the unit time period equal 1 day. Consider the (demeaned) return generating process $r_{(m),t} = \sigma_{(m),t} z_{(m),t}$ where $\sigma_{(m),t}$ is the latent standard deviation and $z_{(m),t}$ is distributed i.i.d. (0, 1). Ideally, forecast variance, denoted by $\hat{h}_{t,t+q}$, should be compared to the latent variance $\sigma_{(m),t,t+q}^2$. Since this is not possible $r_{(m),t}^2$ is often used to proxy for $\sigma_{(m),t}^2$ with the justification of being an unbiased estimator, since $E[r_{(m),t}^2] = E[\sigma_{(m),t}^2 z_{(m),t}^2] = E[\sigma_{(m),t}^2] E[z_{(m),t}^2] = \sigma_{(m),t}^2$. While unbiased, Andersen and Bollerslev (1998) show the variance in $z_{(m),t}$ to be several orders of magnitude larger than the variance in $\sigma_{(m),t}$ making daily squared returns a noisy estimate of the latent variance. A central result in Andersen and Bollerslev (1998) is that the latent variance can be closely approximated by summing up squared returns of a higher frequency than the forecast horizon.² The results also show that measurements of the true degree of predictability are crucially dependent on the proxy for *ex post* variance. However, not only is it important to use a proxy with low noise to measure the absolute degree of predictability but also the relative forecasting performance between models can be affected by the noise in the proxy, as shown in Hansen and Lunde (2006). It is therefore strongly suggested to construct an *ex post* variance that is as noise free as possible also in comparative studies.

Loss functions

Ideally, a loss function should be tailored to the individual person and problem at hand, but to get results that are more general standard statistical loss functions have to be used. There have been a large number of loss functions used previously in the literature and, since the relative forecasting performance is not invariant with respect to the used loss function, this unavoidably introduces a degree of arbitrariness into the results.

Since the variance is being forecast, traditional loss functions such as mean square forecast error will raise the return innovation to the fourth power and hence make the loss function very sensitive to outliers. For this reason the more robust mean absolute error is used.

$$\frac{1}{T} \sum_{t=1}^T |RV_{m,t,t+q} - \hat{h}_{t,t+q}| \quad \text{Mean-absolute error (MAE)} \quad (1)$$

²This method of approximating the latent variance is often called 'realized' variance. For a formal treatment of the subject derived from the theory of quadratic variation see Andersen *et al.* (2005).

$RV_{m,t,t+q} = \sum_{j=1}^{mq} r_{(m),t+j/m}^2$ is the q period realized variance computed with frequency m from time t to time $t+q$ used to proxy the unobservable *ex post* variance. $\hat{h}_{t,t+q}$ is the q period model variance forecast from time t to time $t+q$.

To account for the heteroscedasticity also a heteroscedasticity-adjusted mean-absolute error loss function is used following Andersen *et al.* (1999):

$$\frac{1}{T} \sum_{t=1}^T \left| 1 - \frac{\hat{h}_{t,t+q}}{RV_{m,t,t+q}} \right| \quad \text{Heteroscedasticity-adjusted MAE (HMAE)} \quad (2)$$

RETURN DISTRIBUTION

The inability of the Garch model estimated with a normal error distribution to account for all the excess kurtosis in financial returns is well known. Bollerslev (1987) had already addressed the problem by using a t -distribution and Nelson (1991) by using a generalized error distribution. Recently the problem has been highlighted by Malmsten and Teräsvirta (2004), who use moment results to show that the rather low but slowly decaying autocorrelation function of squared (or absolute) returns together with the high excess kurtosis typically found in financial data are not attainable using standard Garch models in conjunction with the normal distribution. It is also a well-known fact (documented by Glosten *et al.*, 1993, among others) that daily stock index returns exhibit negative skewness.

A probability density function that can account for both time-varying excess kurtosis and skewness is proposed in Hansen (1994). This density function has the additional benefit of having a parsimonious parameterization of its third and fourth moments. Therefore, this density function, presented in equation (3), will be the preferred choice in this paper:

$$g(z|\eta, \lambda) = \begin{cases} bc \left[1 + \frac{1}{\eta-2} \left(\frac{bz+a}{1-\lambda} \right)^2 \right]^{\frac{\eta+1}{2}}, & z < -\frac{a}{b} \\ bc \left[1 + \frac{1}{\eta-2} \left(\frac{bz+a}{1+\lambda} \right)^2 \right]^{\frac{\eta+1}{2}}, & z \geq -\frac{a}{b} \end{cases} \quad (3)$$

where a , b , c and z are defined as below:

$$a = 4\lambda c \left(\frac{\eta-2}{\eta-1} \right) \quad (4)$$

$$b^2 = 1 + 3\lambda^2 - a^2 \quad (5)$$

$$c = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\pi(\eta-2)}\Gamma\left(\frac{\eta}{2}\right)} \quad (6)$$

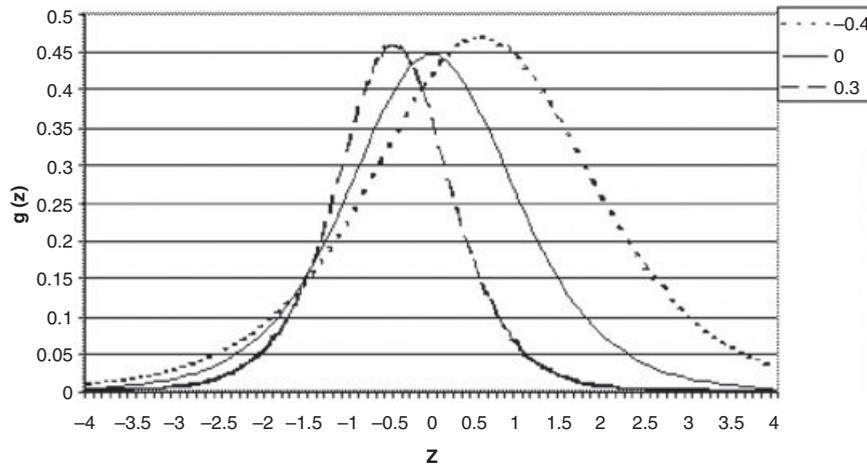


Figure 1. Density plot. This figure shows three different density distributions given by equation (3) with degrees of freedom parameter η , equal to 7 and the skewness parameter λ_i , equal to -0.4 , 0 and 0.3 respectively

$$z = \frac{\varepsilon_t}{\sigma_t} \quad (7)$$

The parameters η and λ are called the shape parameters of the distribution, λ controls the asymmetry of the distribution and η the tail fatness. More specifically the distribution has a single mode at $-\lambda$.

The shape parameters need to be restricted with $2 < \eta < \infty$ and $-1 < \lambda < 1$. Figure 1 shows the distribution for a few different values on the shape parameters.

The presented density function is very flexible since it allows the shape parameters to vary over time; actually, every observation is allowed to have its own values on the shape parameters. To isolate the effects of allowing for excess kurtosis, non-centrality and time variation in the shape parameters, restrictions are imposed on the specification in equation (3).

The restrictions decide if the shape parameters are to be given a value or if they are to be estimated and, in case they are estimated, whether they are allowed to be time varying or not. These restrictions result in the nine different specifications (including the unrestricted model) presented in Table I.

DATA AND ESTIMATION

The data used are intra-daily logarithmic returns from Standard and Poor's 500 index future from 2 January 1996 to 30 December 2002 obtained from Tickdata. The reason why the index future is preferred over the cash index is that the constituents of the SP500 index are not all traded at the same time and some are traded rather infrequently. This induces spurious autocorrelation into the cash index return, especially at higher frequencies.

The data are divided into a 5-year in-sample period (1996–2000) and a 2-year period (2001–2002) is reserved for out-of-sample forecasting evaluation. All in-sample estimations are conducted on a

Table I. Error distributions

Name	λ	η	λ Time-varying	η Time-varying
1. Normal distribution	0	∞	No	No
2. Skewed normal distribution	Est.	∞	No	No
3. Student's t distribution	0	Est.	No	No
4. Skewed Student's t distribution	Est.	Est.	No	No
5. Normal distribution with time-varying skewness	Est.	∞	Yes	No
6. Student's t -distribution with time-varying kurtosis	0	Est.	No	Yes
7. Student's t -distribution with time-varying skewness	Est.	Est.	Yes	No
8. Skewed Student's t distribution with time-varying kurtosis	Est.	Est.	No	Yes
9. Skewed Student's t distribution with time-varying skewness and kurtosis	Est.	Est.	Yes	Yes

Note: This table shows the nine different error distributions obtained by imposing restrictions on the shape parameters of the density function in equation (3).

Table II. Descriptive statistics of the Standard and Poor's 500 Future Index daily returns

SP500 Future returns	1996–2000	2001–2002
Number of observations	1262	500
Daily mean	0.060%	−0.083%
Yearly standard deviation	19.68%	25.19%
Maximum	5.67%	7.67%
Minimum	−7.71%	−6.19%
Skewness	−0.303	0.134
Skewness robust	−0.021	−0.026
Excess kurtosis	3.848	2.045
JB	<0.001	<0.001

Note: This table shows the descriptive statistics for the daily Standard and Poor's 500 future returns for the in-sample period 1996–2000 as well as for the out-of-sample period 2001–2002. Skewness robust is the skewness measure $\frac{q3+q1-2q2}{q3-q1}$, suggested in Kim and White (2004) because of its robustness against outliers. JB is the p -value from the Jarque and Bera (1987) test with the null hypothesis of normally distributed returns.

daily return frequency. The returns are sampled from the futures contract with the highest liquidity, which resulted in the contracts being rolled over around the eighth of every third month.³

The descriptive statistics in Table II show that the in-sample period has an average yearly standard deviation of 19.68%, whereas the out-of-sample period has a somewhat higher standard deviation of 25.19%.

Both periods exhibit excess kurtosis and skewness, though the skewness defined as the standardized third central moment is negative in the in-sample period and positive in the out-of-sample

³The future contracts expiry date is the third Friday in the months March, June, September and December.

period. When the more robust skewness measure $\frac{q3+q1-2q2}{q3-q1}$, where $q1$, $q2$ and $q3$ are the

quartiles of the returns, is used as suggested by Kim and White (2004) the skewness is small but negative in both periods. Normality is overwhelmingly rejected for both periods with p -values from the Jarque and Bera (1987) test less than 0.001.

The index future is traded on the Chicago Mercantile Exchange from 8:30 a.m. to 3.15 p.m., giving 405 minutes of daily trading time. Theoretically, the realized variance should be constructed from the highest possible return frequency since this gives the least noisy estimate of the true variance. However, market microstructure noise will distort this theoretical result.

In the absence of market microstructure effects the realized variance computed at any frequency will give an unbiased estimate of the variance; this fact can be used to select an appropriate return frequency to construct the realized variance series. In previous studies, 5-minute returns have often been used, for example, by Andersen and Bollerslev (1997) for the foreign exchange market and by Andersen *et al.* (2001) with stock market returns. Typically for ultra-high-frequency returns, the realized variance measure becomes severely biased. From the variance signature plot in Figure 2 it can be seen that the realized variance constructed from 1-minute returns is still unbiased.⁴

The daily realized variance in this study is thus constructed from 405 daily 1-minute returns. The 1-minute returns are constructed by using the tick price occurring closest after each 1-minute interval. The 500 trading days during 2001–2002 have 1,601,416 ticks recorded, which gives an average of 7.9 ticks per 1-minute interval. Since the S&P 500 Future is not traded 24 hours a day this measure

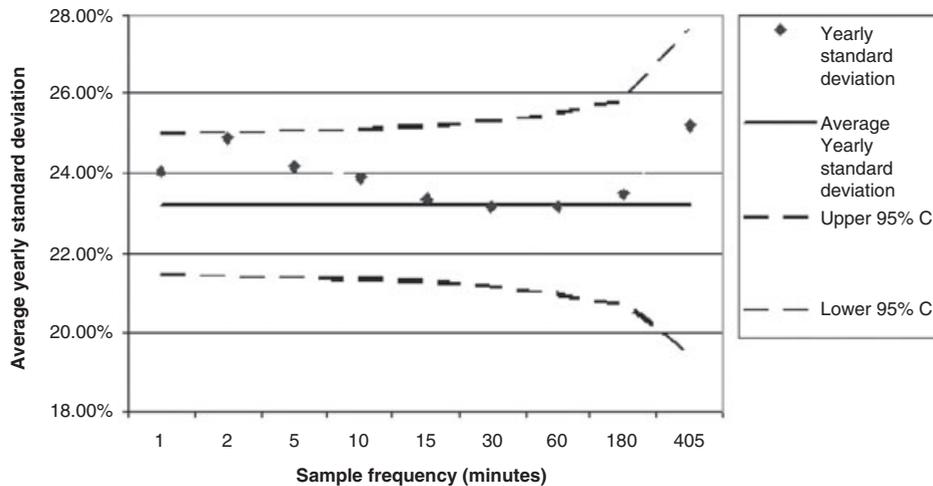


Figure 2. Variance signature plot. This figure shows the average yearly percentage standard deviation for the SP500 Index future during the 2 January 2001 to 30 December 2002 computed at different return frequencies. The solid line in the middle is the average yearly percentage standard deviation computed from 30-minute returns. The dashed lines are 95% confidence intervals computed according to Appendix A in Hansen and Lunde (2004)

⁴Of course, even higher frequencies could be examined but the noise reduction from using even higher frequencies is minuscule, as shown in Andersen *et al.* (2004).

will be downward biased. The reason is that the measure completely ignores the variance in the overnight return. One way to remedy this is to simply add the variance of the overnight return to the variance of the daily return.

This would, however, introduce a lot of noise in the measure and hence eliminate much of the gains from using *intra-day* data. Instead the scaling estimator presented in Hansen and Lunde (2005b) will be used. This method assumes the variance of the overnight return to be a constant fraction of the daily variance and each daily realized variance is simply scaled appropriately to eliminate the bias.⁵

The model estimated is the Garch (1, 1) model of Bollerslev (1986) given by the mean equation:

$$r_t = \omega + \varepsilon_t \quad (8)$$

and the conditional variance equation:

$$h_t = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 h_{t-1} \quad (9)$$

with β_0, β_1 and $\beta_2 > 0$ and $\beta_1 + \beta_2 < 1$. The mean equation could alternatively be specified as a function of the conditional variance giving the Garch-M model of Engle *et al.* (1987) or restricted to have zero mean as suggested by Figlewski (1997). However, Hansen and Lunde (2005a) show that these different specifications have virtually no impact on the out-of-sample variance forecasts of the model.

The restrictions on the density function in Table I are imposed to give nine different distributions for ε_t . When time varying, the shape parameters η and λ are made dependent on lagged residuals from the mean equation, according to

$$\eta_{t(\text{estimated})} = a + b\varepsilon_{t-1} + c\varepsilon_{t-1}^2 \quad (10)$$

and

$$\lambda_{t(\text{estimated})} = d + e\varepsilon_{t-1} + f\varepsilon_{t-1}^2 \quad (11)$$

Here a to f are parameters to be estimated and a, b and c are not to be confused with the a, b and c defined in equations (4)–(6). To restrict the shape parameters the logistic transformation

$$\text{Shape}_{t(\text{restricted})} = L + \frac{(U - L)}{1 + \exp(-\text{Shape}_{t(\text{estimated})})} \quad (12)$$

is used. U and L are the upper and lower bounds for the restricted parameters. In the estimation the skewness parameter is restricted to]-0.9, 0.9[. By excluding zero the models are formally non-nested, which makes the asymptotic theory of the Diebold and Mariano (1995) test valid. This additional restriction is never close to being binding in the estimation. The tail thickness parameter is restricted to the open interval]2.1, 30[.

⁵The yearly standard deviation including the overnight return is 25.19% and excluding the overnight return 21.22%. Each daily realized variance is hence scaled with $(0.2519/0.2122)^2 = 1.409$ to eliminate the bias.

The mean, variance and shape parameter equations are estimated jointly by numerical maximization of the log-likelihood function with the BFGS algorithm. Model convergence is not sensitive to the selection of starting values for the parameters. Forecasts are computed for all nine models; also, a naïve benchmark is introduced that simply uses average variance calculated as squared daily returns over the last 30 days.

Daily returns are used to facilitate a fair comparison with the Garch models that do not utilize *intra-day* returns. The forecast horizons are $q = 1, 5$ and 20 days. The first forecasts are made on $t = 31$ December 2000 and then additional forecast are made on each consecutive day for the whole out-of-sample period, resulting in 500 daily forecasts, 496 5-day forecasts and 481 20-day forecasts. The Garch parameter estimates are held constant during the entire forecasting period in accordance with Andersen *et al.* (1999). The q -day-ahead variance forecasts are given by $\hat{h}_{t,t+q} = (\hat{\beta}_1 + \hat{\beta}_2)^q (\hat{h}_t - v) + v$ with $v = \frac{\hat{\beta}_0}{1 - \hat{\beta}_1 - \hat{\beta}_2}$, which is the unconditional variance. For the first forecast ε_t^2 and h_t are given by the estimated model and for the consecutive forecasts the prior periods forecast \hat{h}_t replaces h_t . The forecasting performance for each model is examined by the loss functions in equations (1) and (2).

As mentioned in Poon and Granger (2003) it is important to assess whether the forecasting performance of competing models is statistically different. This poses some econometric problems since the forecast errors from two competing models are likely to be both contemporaneously and serially correlated, making standard tests invalid. Diebold and Mariano (1995) propose a test for equal predictive ability that allows for serial and contemporaneous correlation, arbitrary loss functions and non-Gaussian forecast errors. Limitations of the test are that the asymptotic distribution results derived are not valid when comparing nested models. A further limitation is that models can be compared only pairwise.

Following the recommendations in Diebold and Mariano (1995) a Dirichlet spectral window is used with a truncation lag equal to one minus the forecast horizon when the test is implemented.

RESULTS

In-sample results

The level of volatility persistence in the models as measured by the sum of $\hat{\beta}_1$ and $\hat{\beta}_2$ (Table III) is well below one for all models with model nine, giving the highest persistence of 0.9675. The unconditional standard deviation given by $\sqrt{\frac{\hat{\beta}_0}{1 - (\hat{\beta}_1 + \hat{\beta}_2)}}$ varies from 19.75% for model seven to

24.92% given by model nine. The lambda parameter that models the skewness of the distribution is highly significant in all the specifications and the skewness found in the standardized residuals is around -0.65 for all the models that do not allow for skewness in the distribution. This is in agreement with previous studies. For example, Glosten *et al.* (1993) find a skewness value of -0.78 in their standard Garch specification and -0.70 when they allow for asymmetric responses to variance shocks. It is important to notice that modeling asymmetric responses in variance to return shocks does not by far account for the skewness in the return distribution.

When model one is estimated with a GJR-Garch specification (estimation results not reported) the skewness in the residuals changes only marginally to -0.59 . In addition, the lambda parameter in model two, which models the skewness of the distribution, has a value of -0.224 , which is highly significant. To show that the lambda parameter does not simply compensate for possible

Table III. Results from in-sample estimation

Estimate	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
ω	0.0814 (0.0282)	0.0675 (0.02566)	0.1074 (0.0292)	0.0733 (0.0302)	0.0783 (0.0209)	0.1059 (0.0270)	0.0772 (0.0298)	0.0611 (0.0249)	0.0696 (0.0279)
β_0	0.0939 (0.0365)	0.0730 (0.0227)	0.0835 (0.0269)	0.0771 (0.0262)	0.0739 (0.0013)	0.0888 (0.0199)	0.0751 (0.0293)	0.0830 (0.0289)	0.0807 (0.0266)
β_1	0.1316 (0.0229)	0.1278 (0.0271)	0.1131 (0.0274)	0.1184 (0.0256)	0.1278 (0.0015)	0.1446 (0.0031)	0.1154 (0.0293)	0.1509 (0.0306)	0.1560 (0.0068)
β_2	0.8115 (0.0468)	0.8272 (0.0321)	0.8344 (0.0350)	0.8328 (0.0353)	0.8263 (0.0075)	0.8156 (0.0114)	0.8365 (0.0389)	0.8129 (0.0364)	0.8115 (0.0187)
λ	*	-0.2238 (0.0371)	*	-0.1781 (0.0356)	Vector	*	Vector	-0.1929 (0.0388)	Vector
d.f.	*	*	6.9320 (1.2810)	7.9069 (1.3844)	*	Vector	8.0780 (1.430)	Vector	Vector
a	*	*	*	*	*	-1.7565 (0.3023)	*	-1.4651 (0.3540)	-1.3407 (0.4931)
b	*	*	*	*	*	-0.5364 (0.1607)	*	-0.6631 (0.1670)	-1.0508 (0.2301)
c	*	*	*	*	*	-0.0527 (0.0219)	*	-0.0603 (0.0212)	0.0723 (0.0262)
d	*	*	*	*	-0.5144 (0.0721)	*	-0.4213 (0.0784)	*	-0.4811 (0.0905)
e	*	*	*	*	0.1633 (0.0454)	*	0.0691 (0.0660)	*	0.1358 (0.0653)
f	*	*	*	*	0.0238 (0.0096)	*	0.0153 (0.0088)	*	0.0173 (0.0111)

Note: This table shows the quasi maximum likelihood parameter estimates with robust standard errors in parentheses from the estimation of models one to nine given by equations (8)–(12) and the restrictions from Table I. The models are estimated on daily SP500 Index future returns from 2 January 1996 to 30 December 2000.

misspecification in the variance equation (positive and negative innovations forced to have the same effect) a GJR-Garch model is also estimated (estimation results unreported) with the distribution of model two. This results in the lambda parameter changing only slightly, to -0.236 , and remaining highly significant. This clearly shows that the asymmetry in the return distribution and the asymmetric response of variance to return innovations need to be modeled separately. This points to the need for an asymmetric error distribution, which is an issue that has previously received only minor attention in the Garch forecasting literature. However, it has been used for in-sample Garch estimation by, for example, Hansen (1994), Brännäs and Nordman (2003) and Harvey and Siddique (1999).

Furthermore it can be seen from the e and f parameters in models five, seven and nine that the skewness in one period is positively related to the return and squared return innovation in the previous period, whereas the relationship is negative for the degrees of freedom parameters, b and c in models six, eight and nine (with the exception of the c parameter in model nine). Previously Hansen (1994) found both the degrees of freedom and skewness parameters to be negatively dependent on lagged return innovations for the weekly dollar/Swiss franc exchange rate.

The model diagnostics in Table IV show that only the full specification of model nine can account for both the skewness and leptokurtosis in the residuals.⁶ As usual, the Garch model does a good job

⁶The residuals are standardized with the conditional skewness and kurtosis produced by the model using the moment results in Jondeau and Rockinger (2003).

Table IV. Model diagnostics

Model	Log-likelihood	Kurtosis	Skewness	Q10
1	-1982.62	1.7792	-0.64145	11.3474
2	-1957.79	1.8980	-0.3334	11.6466
3	-1950.25	-0.3138	-0.6468	12.1653
4	-1940.35	0.1649	-0.1888	11.9576
5	-1955.18	1.8889	-0.3559	11.5838
6	-1946.40	-1.2265	-0.5927	11.2891
7	-1939.39	0.2640	-0.2038	12.1202
8	-1934.00	-1.1728	-0.3448	11.5641
9	-1933.69	0.1258	0.0331	11.6480

Note: This table shows the log-likelihood from the estimation of models one to nine from Table III on the SP500 Index future returns from 2 January 1996 to 30 December 2000. Kurtosis (in excess of 3) and skewness is for the standardized residuals from each model and Q10 is the Lagrange multiplier test for autocorrelation in squared residuals. To standardize the residuals with the conditional skewness and kurtosis produced by the model, the moment results in Jondeau and Rockinger (2003) are used.

at accounting for the dependence in squared returns, and the Q10 statistic shows no significant autocorrelation in squared residuals for any of the models.

As can be seen in Table IV, the two most drastic improvements in log-likelihood by adding a single parameter are to allow for skewness and leptokurtosis, respectively. By replacing the normal distribution with a t -distribution the log-likelihood increases from -1982.62 to -1950.25, which is significant according to a standard likelihood ratio test with a p -value of less than 0.001.

Allowing for skewness in the distribution, as is done in model two, increases the log likelihood compared to the normal from -1982.62 to -1957.79, which is a significant increase with a p -value of less than 0.001. By jointly allowing for both skewness and leptokurtosis the log-likelihood is further improved to -1940.35. It is also interesting to notice that the skewness parameter, lambda, increases from -0.224 to -0.178 and that the degrees of freedom parameter increases from 6.93 to 7.91. This shows the apparent importance of allowing for both skewness and leptokurtosis. This can also be seen from the negative excess kurtosis of the residuals produced by the models that allow for leptokurtosis but not skewness. The lambda parameter will partly model the empirical excess kurtosis as skewness, and the degrees of freedom parameter will partly model the empirical skewness as kurtosis, leading to a bias from under-specification.

Model six shows the effect of further allowing the degrees of freedom parameter to be time varying while the lambda parameter is restricted to zero; the improvement appears to be quite small, with an increase in likelihood from -1950.25 to -1946.40. This increase is not significant, with a p -value of 0.146.

The gain by allowing the skewness to be time varying is seen by comparing models two and five, resulting in an increase in the log-likelihood from -1957.79 to -1955.18, which is also insignificant, with a p -value of 0.272. Nevertheless, these results indicate that there is persistence in skewness and kurtosis as seen by the parameters d , e and f in model five that are all significant, and parameters a , b and c in model six that also show predictable time variation in the degrees of freedom parameter. In comparing models five and seven, the effect of allowing for a t -distribution when the skewness is time varying is shown; this increases the log-likelihood from -1955.18 to -1939.39, which is a significant improvement. By comparing models seven and eight it is seen that allowing for time

variation in the degrees of freedom parameter is more important than allowing for time variation in the skewness, judged by the higher likelihood of model eight.

Model nine, which is the full unrestricted model, shows only a very minor and insignificant improvement over model eight in terms of log-likelihood, indicating that there is not much gain in allowing the lambda parameter to be time varying.

Figures 3 and 4 show the time variation in skewness and degrees of freedom produced by model nine. The skewness is mostly negative but has a few sharp peaks, with the highest reaching 0.55. The degrees of freedom oscillate around seven, with a few peaks close to 30. The constraints imposed by the logistic transformation are not close to being binding for the lambda parameter. However, they are almost binding at 30 for the degrees of freedom parameter on a few occasions, showing that the conditional distribution is at times close to normal.

Forecast results

Forecasting performance

In terms of mean absolute forecast error, model three performs the best for all three forecast horizons (see Tables V and VI), closely followed by model seven. However, the difference is not significant for any of the three horizons, with the lowest p -values being 6.9% from the Diebold–Mariano (1995) test. Also, when the criterion is heteroscedasticity adjusted, model three performs the best on the 1-day and 5-day horizon; for the 20-day horizon the ranking is reversed, with model seven having the lowest error, followed by model three. The difference is not significant here either.

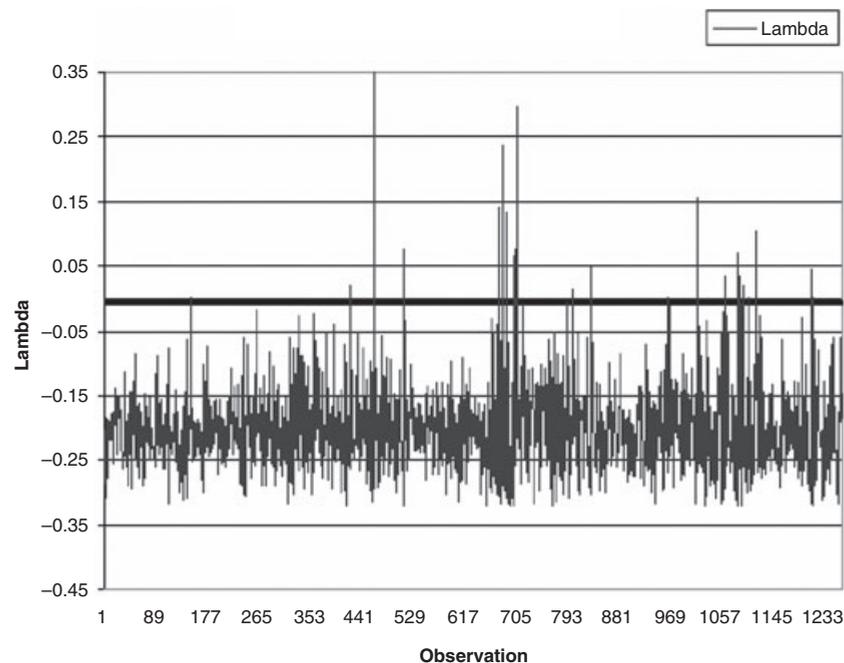


Figure 3. Conditional skewness over time. This figure shows the logistically transformed skewness parameter λ_t given by equations (11) and (12) from model nine during 2 January 1996 to 30 December 2000

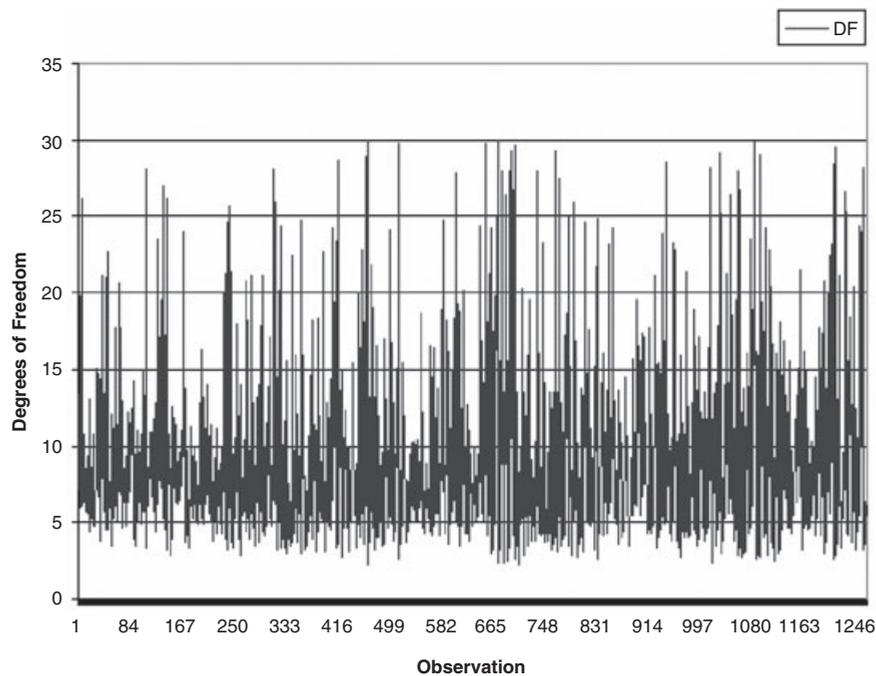


Figure 4. Conditional kurtosis over time. This figure shows the logarithmically transformed degrees of freedom parameter η_t given by equations (10) and (12) from model nine during 2 January 1996 to 30 December 2000

Table V. Absolute performance: 1-, 5- and 20-day forecast horizons

Criteria	1	2	3	4	5	6	7	8	9	10
<i>1 day</i>										
MAE	0.91216	0.91832	0.90567	0.90995	0.91759	0.94782	0.90892	0.9570	0.97009	1.25024
HMAE	0.33252	0.33607	0.32869	0.33045	0.33512	0.36718	0.32946	0.36924	0.37648	0.51643
<i>5 days</i>										
MAE	4.35047	4.35833	4.26345	4.29785	4.35602	4.49369	4.28593	4.55555	4.63354	6.19096
HMAE	0.29009	0.29407	0.28346	0.28710	0.29325	0.32339	0.28597	0.32743	0.33650	0.49010
<i>20 days</i>										
MAE	18.5813	18.5860	18.4528	18.4774	18.5724	19.7020	18.4649	19.9100	20.4031	28.3153
HMAE	0.30928	0.30969	0.30153	0.30253	0.30818	0.38413	0.30087	0.38833	0.40721	0.55958

Note: This table reports the forecast errors for the nine models from Table III plus the forecast errors from a 30-day rolling variance model. Forecasts are made on the SP500 Futures index from 2 January 2001 to 30 December 2002. Best-performing models are shown in bold.

In conclusion, of the six different rankings, model three is preferred five times and model seven once. The unrestricted model, which is model number nine, clearly performs the worst out-of-sample on both the MAE and HMAE criteria, suggesting possible over-fitting when both the skewness and degrees of freedom are allowed to be time varying. All the Garch models convincingly beat the benchmark model, which uses a 30-day moving average as predictor of variance. This is in

Table VI. Relative performance: 1-, 5- and 20-day forecast horizons

Criteria	1	2	3	4	5	6	7	8	9	10
<i>1 day</i>										
MAE	0.00649	0.01265	0	0.00428	0.01192	0.04215	0.00325	0.05133	0.06442	0.34457
HMAE	0.00383	0.00738	0	0.00176	0.00643	0.03849	0.00077	0.04055	0.04779	0.18774
<i>5 days</i>										
MAE	0.08702	0.09488	0	0.03440	0.09257	0.23024	0.02248	0.29210	0.37009	1.92751
HMAE	0.00663	0.01061	0	0.00364	0.00979	0.03993	0.00251	0.04397	0.05304	0.20664
<i>20 days</i>										
MAE	0.1285	0.13316	0	0.02463	0.11957	1.24924	0.01205	1.46715	1.95033	9.86246
HMAE	0.00841	0.00882	0.00066	0.00166	0.00731	0.08326	0	0.08746	0.10634	0.25871

Note: This table shows the relative performance of the models when the loss function from the best model is subtracted from each model, meaning that zero indicates the best-performing model. Forecasts are made on the SP500 Futures index from 2 January 2001 to 30 December 2002. Best-performing models are shown in bold.

Table VII. Model significance

Criteria	1 day		5 days		20 days	
	Best model	Sig. better than #	Best model	Sig. better than #	Best model	Sig. better than #
MAE	3	6	3	2	3	1
HMAE	3	7	3	7	7	7

Note: This table reports the best-performing models for the different forecast horizons and loss functions. It also shows the number of models that were significantly worse than the best model.

agreement with McMillan and Speight (2004), who examine Garch forecasting performance compared to moving average models for 17 exchange rate series.

Significance

As can be seen from Table VII, for the 1-day horizon model three is significantly better, on the 5% level, than six of its nine competing models when evaluated by the MAE criterion. Model three is significantly better than seven of the other models for the HMAE criterion. For the 5-day horizon model three is significantly better than two other models in terms of MAE, and significantly better than seven of the nine competing models in terms of HMAE. For the 20-day horizon model three is only significantly better than the 30-day moving average model for the MAE loss function. For the HMAE criterion model three is significantly better than seven of the competing nine models.

Looking generally at significance, 45 pairwise model comparisons are computed for each loss function and each forecast horizon. As can be seen from Table VIII, there is a substantial difference between the MAE and HMAE loss functions if a model comparison can be considered significant or not. The heteroscedasticity adjusted mean absolute error criterion finds the most significant differences: 95.56% of the comparisons are significant at the 1-day forecast horizon, while the corresponding figure for the MAE criterion is 77.78%.

When the forecast horizon is extended to 5 and 20 days the heteroscedasticity-adjusted mean absolute error has more significant differences, especially on the 20-day horizon, when 73.33% of the comparisons is significant compared to 13.33% for the MAE. Since the Diebold–Mariano (1995)

Table VIII. Percentage of comparisons significant at 5%

Horizon	Criterion		
	MAE	HMAE	Average
1	77.78%	95.56%	86.67%
5	33.33%	86.67%	60.00%
20	13.33%	73.33%	43.33%
Average	41.48%	85.19%	63.33%

Note: This table shows the percentage of comparisons significant at the 5% level when all models are compared pairwise with the Diebold and Mariano (1995) test.

test adjusts the standard errors for heteroscedasticity it seems a good idea, in terms of attaining significant results for longer forecast horizons, to adjust for the heteroscedasticity in the loss function. Averaged across the two loss functions the percentage of significant comparisons between models is 86.67% for 1-day forecasts, 60.00% for the 5-day horizon and drops to 43.33% for the 20-day horizon.

Properties that improve the forecasts

In order to discover the possible causes for the model performance reported above, the model specifications will now be looked at in more detail. The three models that allow for time-varying degrees of freedom are models six, eight and nine. These three models perform poorly on both the MAE and HMAE loss criteria.

Models three and seven, which had the best forecasting performance, both allow for a t -distribution with an estimated degrees of freedom parameter. This is in close agreement with the in-sample results, where the largest increase in log-likelihood came from allowing for fat tails. These results partly agree with Hansen and Lunde (2005a), who found Garch models estimated with t -distributions performing on average better for the DM/\$ exchange rate but worse than the Gaussian models on IBM stock return data.

By comparing models one and two we see that there is no gain in forecasting performance by allowing for skewness in the distribution. In fact, model two performs worse than model one on all six comparisons. Model five further allows for time-varying skewness, which makes a slight improvement over model two, but the model is still worse than the regular Garch model estimated with a normal distribution. Since skewness was found significant in the in-sample estimation, this is somewhat puzzling. A possible explanation is that a few outliers cause the observed skewness since it almost disappears when assessed by a skewness measure that is more robust to outliers (see Table I). These outliers have a large positive effect on the log-likelihood of the models that allow for skewness.

By comparing models one and three the effect of allowing for a leptokurtic but symmetric distribution is examined. There seems to be substantial improvement over the normal distribution, with model three outperforming model one on all six measures.

Model six further allows the degrees of freedom to be time varying, which does not lead to an improvement; on all forecast horizons and for both loss functions the model with time-varying degrees of freedom is worse than the regular t -distribution.

The joint effect of allowing for non-time-varying skewness and kurtosis is determined by comparing models one and four. Model four performs better on all measures, but worse than model three,

showing that there is no improvement in also allowing for skewness when excess kurtosis is already modeled. By further allowing for time variation in the skewness parameter, model seven is obtained. Model seven outperforms both model one and four on all the measures, showing an improvement in allowing the skewness to be time varying. By allowing the kurtosis to be time varying and the skewness to be constant model eight is obtained. Model eight, though, performs worse than both models one and four on all measures. This shows that the forecasting performance is actually worsened by allowing for time variation in the kurtosis. When both kurtosis and skewness are allowed to be time varying model nine is obtained; this model is worse than model one in all comparisons.

CONCLUSIONS

Model three, the Garch model estimated with Student's t distribution, must be considered the best performing model, since it performs the best on both the MAE and HMAE loss functions for both the 1- and 5-day forecasts. In addition, model three has the lowest MAE for the 20-day forecasts and is a close second on the HMAE criterion.

Other models that performed very well are models number seven and four; these two models have in common with model three that they allow for non-time-varying leptokurtosis. Overall, the Garch models clearly outperform the moving average forecasts of model 10 on both criteria and for all forecast horizons.

The above results are in partial agreement with Hamilton and Susmel (1994), who found the Garch model with a t -distribution to perform the best when the logarithmic loss criterion was used. However, Hamilton and Susmel also found that the Garch models gave no improvements over a model with constant variance when using the MSE loss criterion. In contradiction to the results of this study, Franses and Ghijssels found the out-of-sample performance for the Garch model with a t -distribution worse than for a Garch model estimated with the normal distribution. The contradictory results can probably be explained partly by the noisy *ex post* variance proxy (squared weekly returns) used by both Hamilton and Susmel as well as Franses and Ghijssels and partly by the MSE criterion's sensitivity to large return innovations.

Allowing for non-time-varying kurtosis improves the forecasts; however, also allowing non-time-varying skewness or time-varying kurtosis does not improve the forecasts further. When the models were estimated, the two clearest improvements in log-likelihood came from allowing for a skewed distribution and to allow for excess kurtosis. The out-of-sample results show that only allowing for excess kurtosis improves forecasts; allowing for skewness does not lead to any improvements over the normal distribution.

This shows that additional important insights can be gained by using forecasting performance as a model selection criterion and thus not by solely relying on in-sample estimation results. Allowing for time variation gives insignificant increases in log-likelihood for the estimated models; this is confirmed by the forecasting performance, which shows no increase in forecasting ability for the models with time-varying skewness and/or kurtosis. Of course, the skewness present in financial returns cannot be ignored only because including it in the model does not improve the variance forecasts. The implications of a forecast variance, e.g. for option pricing and risk management, are quite different depending on the degree of skewness present in the distribution. Since the shape parameters of the distribution were significantly dependent on lagged return innovations, future work on forecasting the entire return distribution seems promising, and is currently being investigated by the author.

ACKNOWLEDGEMENTS

Financial support from Hanken's grant for young doctoral students, grant number 04-2-96-1-8, is gratefully acknowledged. The author would like to thank Timo Teräsvirta, Anders Ekholm, Peter Nybergand and Anette Björkman for helpful comments and suggestions.

REFERENCES

- Andersen TG, Bollerslev T. 1997. Heterogeneous information arrivals and return variance dynamics: uncovering the long-run in high frequency returns. *Journal of Finance* **52**: 975–1005.
- Andersen TG, Bollerslev T. 1998. Answering the skeptics: yes standard variance models do provide accurate forecasts. *International Economic Review* **39**: 885–905.
- Andersen TG, Bollerslev T, Lange S. 1999. Forecasting financial market variance: sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance* **6**: 457–477.
- Andersen TG, Bollerslev T, Diebold FX, Ebens H. 2001. The distribution of stock return variance. *Journal of Financial Economics* **61**: 43–76.
- Andersen TG, Bollerslev T, Meddahi N. 2004. Analytic evaluation of variance forecasts. *International Economic Review* **45**: 1079–1110.
- Andersen TG, Bollerslev T, Diebold FX. 2005. Parametric and nonparametric volatility measurement. In *Handbook of Financial Econometrics*, Hansen LP, Ait-Sahalia Y (eds). North-Holland: Amsterdam (forthcoming).
- Bollerslev T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**: 307–327.
- Bollerslev T. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* **69**: 542–547.
- Brännäs K, Nordman N. 2003. An alternative conditional asymmetry specification for stock returns. *Applied Financial Economics* **13**: 537–541.
- Diebold FX, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.
- Engle R. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1008.
- Engle R, Lilien DM, Robins RP. 1987. Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica* **55**: 391–407.
- Figlewski S. 1997. Forecasting variance. *Financial Markets, Institutions and Instruments* **6**: 1–88.
- Franses P, Ghijssels H. 1999. Additive outliers, Garch and forecasting variance. *International Journal of Forecasting* **15**: 1–9.
- Glosten L, Jagannathan R, Runkle D. 1993. On the relation between the expected value and the variance of the nominal excess returns on stocks. *Journal of Finance* **48**: 1779–1801.
- Hamilton J, Susmel R. 1994. Autoregressive conditional heteroscedasticity and changes in regime. *Journal of Econometrics* **64**: 307–333.
- Hansen B. 1994. Autoregressive conditional density estimation. *International Economic Review* **35**: 705–730.
- Hansen P, Lunde A. 2004. An unbiased measure of realized variance. Working paper, Brown University and Aarhus School of Business. <http://www.stanford.edu/~prhansen/> [May 2006].
- Hansen P, Lunde A. 2005a. A forecast comparison of variance models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics* **20**: 873–889.
- Hansen P, Lunde A. 2005b. A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics* **3**: 525–554.
- Hansen P, Lunde A. 2006. Consistent ranking of variance models. *Journal of Econometrics* **131**: 97–121.
- Harvey CR, Siddique A. 1999. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* **34**: 465–488.
- Jarque CM, Bera AK. 1987. A test for normality of observations and regression residuals. *International Statistical Review* **55**: 163–172.
- Jondeau E, Rockinger M. 2003. Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control* **27**: 1699–1737.

- Kim T, White H. 2004. On more robust estimation of skewness and kurtosis: simulation and application to the S&P500 index. *Finance Research Letters* **1**: 56–73.
- Lopez J. 2001. Evaluating the predictive accuracy of variance models. *Journal of Forecasting* **20**: 87–109.
- Malmsten H, Teräsvirta T. 2004. Stylized facts of financial time series and three popular models of variance. *SSE/EFI Working Paper Series in Economics and Finance no 563*.
- McMillan D, Speight A. 2004. Daily volatility forecasts: reassessing the performance of Garch models. *Journal of Forecasting* **23**: 449–460.
- Nelson DB. 1991. Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica* **59**: 347–370.
- Poon S, Granger C. 2003. Forecasting variance in financial markets: a review. *Journal of Economic Literature* **41**: 478–540.
- Taylor S. 1986. *Modelling Financial Time Series*. Wiley: Chichester.

Author's biography:

Anders Wilhelmsson received his Master's degree in Economics from the University of Lund, Sweden, in 2002. He is currently a visiting researcher at the Aarhus School of Business, Denmark, and a doctoral student at the Swedish School of Economics and Business Administration in Helsinki, Finland.

Author's address:

Swedish School of Economics and Business Administration, Department of Finance and Statistics, PO Box 479, FIN-00101 Helsinki, Finland.