GARCH Models for Daily Stock Returns: Impact of Estimation Frequency on Value−at−Risk and Expected Shortfall Forecasts

David Ardia[a]
Lennart Hoogerheide[b]

[a] Université Laval, Québec (Québec), Canada;
[b] VU University Amsterdam, The Netherlands, and Tinbergen Institute.

GARCH models for daily stock returns:
Impact of estimation frequency on Value-at-Risk and Expected Shortfall forecasts[☆]

David Ardia[a], Lennart F. Hoogerheide[b,∗]

[a]*Département de finance, assurance et immobilier, Université Laval, Québec (Québec), Canada*
[b]*Department of Econometrics and Tinbergen Institute, Vrije Universiteit Amsterdam, The Netherlands*

Abstract

We analyze the impact of the estimation frequency — updating parameter estimates on a daily, weekly, monthly or quarterly basis — for commonly used GARCH models in a large-scale study, using more than twelve years (2000-2012) of daily returns for constituents of the S&P 500 index. We assess the implication for one-day ahead 95% and 99% Value-at-Risk (VaR) forecasts with the test for correct conditional coverage of Christoffersen (1998) and for Expected Shortfall (ES) forecasts with the block-bootstrap test of ES violations of Jalal and Rockinger (2008). Using the false discovery rate methodology of Storey (2002) to estimate the percentage of stocks for which the model yields correct VaR and ES forecasts, we reach the following conclusions. First, updating the parameter estimates of the GARCH equation on a daily frequency improves only marginally the performance of the model, compared with weekly, monthly or even quarterly updates. The 90% confidence bands overlap, reflecting that the performance is not significantly different. Second, the asymmetric GARCH model with non-parametric kernel density estimate performs well; it yields correct VaR and ES forecasts for an estimated 90% to 95% of the S&P 500 constituents. Third, specifying a Student-t (or Gaussian) innovations' density yields substantially and significantly worse forecasts, especially for ES. In sum, the somewhat more advanced model with infrequently updated parameter estimates yields much better VaR and ES forecasts than simpler models with daily updated parameter estimates.

JEL classification: C12, C22, C58, G17, G32.

Keywords: GARCH, Value-at-Risk, Expected Shortfall, equity, frequency, false discovery rate

---

1. Introduction

In this paper, we investigate the impact of the estimation frequency of commonly used GARCH models on one-day ahead forecasts of Value-at-Risk (VaR) and Expected Shortfall (ES), where the latter is sometimes called Conditional Value-at-Risk. This is of substantial importance for large risk management systems, which involve thousands of models which need to be estimated and updated. We perform a large-scale study, using more than twelve years (2000-2012) of daily returns for constituents of the S&P 500 index. We rely on a rolling-window estimation approach and forecast the VaR and ES at the 95% and 99% confidence levels. For the models, we rely on symmetric and asymmetric specifications for the variance equation (i.e., without and with a so-called *leverage* effect) and consider Gaussian, Student-t and kernel-based distributions for the errors. We observe that updating the parameter estimates of the GARCH equation on a daily frequency improves only marginally and not significantly the forecasting performance of the model, compared with weekly, monthly or even quarterly updates. On the other hand, the differences between the model specifications are substantial and significant. An asymmetric GARCH model with non-parametric kernel density estimate of the error distribution performs well, yielding correct VaR and ES forecasts for an estimated 90% to 95% of the S&P 500 constituents, whereas specifying a Student-t (or Gaussian) innovations' density yields much worse forecasts, especially for ES. Therefore, if one needs to reduce the computational burden of a large risk management system involving thousands of models, then it seems much worse to choose a simpler model (with daily updated parameter estimates) than to infrequently update the parameter estimates of a more advanced model. The latter substantially reduces the computational efforts without seriously harming the models' performance.

The remainder of this article is organized as follows. In section 2 we present the model specifications, the testing and introduce the false discovery rate method. In section 3 we present and discuss the empirical results. Section 4 concludes.

2. Model specification, testing and false discovery rate method

As in McNeil and Frey (2000), each model considered starts with an AR(1) component in order to filter a possible autoregressive part of the equity log-returns. The models differ in the way the volatility of the error terms is specified. For that purpose, we rely on the sym-

metric GARCH(1,1) model of Bollerslev (1986) and on the asymmetric GJR(1,1) model by Glosten et al. (1993). The latter accounts for the asymmetric effect of positive and negative stock returns on next period's stock return's variance, known as the leverage effect in the literature (Black, 1976).

Both models have a long empirical history and have proved to be successful in volatility modeling in several markets (Bollerslev et al., 1992). They are simple yet powerful GARCH-type models. More specifically, in the AR(1)-GJR(1,1) model the log-returns r_t are expressed as:

$$
\begin{aligned}
r_t &= \mu + \rho\, r_{t-1} + u_t \quad (t = 1, \ldots, T), \\
u_t &= \sigma_t\, \varepsilon_t \quad \varepsilon_t \sim iid\, f_\varepsilon, \\
\sigma_t^2 &= \beta_1 + (\beta_2 + \beta_3\, 1\{u_{t-1} \leq 0\})\, u_{t-1}^2 + \beta_4\, \sigma_{t-1}^2,
\end{aligned}
\tag{1}
$$

where we require $\beta_1 > 0$ and $\beta_2, \beta_3, \beta_4 \geq 0$ to ensure a positive conditional variance σ_t^2. $1\{\}$ denotes the indicator function, whose value is one if the constraint holds and zero otherwise. The covariance stationarity constraint $\beta_2 + \beta_3/2 + \beta_4 < 1$ is imposed in the estimation. The symmetric GARCH model results by imposing $\beta_3 = 0$. For the distribution f_ε, we consider the simple Gaussian and Student-t distributions, together with a non-parametric Gaussian kernel estimator. The Student-t distribution is probably the most commonly used alternative to the Gaussian for modeling stock returns and allows modeling fatter tails than the Gaussian. The kernel approach gives a non-parametric alternative which can deal with skewness and fat tails in a convenient manner. In addition to the standard GARCH models, we consider the Exponentially Weighted Moving Average (EWMA) as advocated by RiskMetrics (RiskMetrics Group, 1996); this amounts to a GARCH model with Gaussian errors and fixed parameter values $\beta_1 = \beta_3 = 0$, $\beta_2 = 0.94$ and $\beta_4 = 0.06$.

Models are fitted by quasi maximum likelihood. For the non-parametric Gaussian kernel estimator, the bandwidth is selected by the rule-of-thumb of Silverman (1986) on the residuals of the quasi maximum likelihood fit – i.e., on the estimated u_t. Alternative bandwidth choices lead to similar results. We rely on the rolling-window approach where 1000 log-returns – i.e., approximately four trading years – are used to estimate the models. Similar results were obtained for windows of 750 and 1500 observations. Then, the next log-return is used as a forecasting

window. At each time point, the one-day ahead Value-at-Risk (VaR) and Expected Shortfall (ES) forecasts are obtained for the different models. VaR represents the risk from market movements as one number: the maximum loss expected on an investment, over a given time period at a specific level of confidence. It is nowadays a standard risk measure of downside risk. In our study, the VaR is a negative percentage; in the literature it is sometimes quoted as a positive percentage (i.e., a percentile of the distribution for the negative of the return) or an amount of dollars. In a similar fashion, the ES is the conditional expectation of the loss given that the loss exceeds the VaR; in other words, the ES is the conditional expectation of the log-return given that log-return is more negative than the VaR. The ES is a coherent measure of risk in the sense of Artzner et al. (2000).

We consider several estimation frequencies for the scedastic function parameters: daily, weekly (5 days), monthly (20 days) and quarterly (60 days) updates. Our goal is to measure to which extent the updating process of the parameters has an impact on the forecasting performance of the VaR and ES. This is of substantial importance for large risk management systems, which involve thousands of models which need to be estimated and updated.

To test the ability of our models to capture the true VaR, we compare the realization of the returns r_t with the one-day ahead VaR forecasts ($\text{VaR}_{t|t-1}$) at 95% and 99% risk levels. To that aim, we adopt the backtesting methodology proposed by Christoffersen (1998) which has become the standard practice in financial risk management. This approach is based on the study of the random sequence V_t where $V_t \doteq 1\{r_t < \text{VaR}_{t|t-1}\}$. A sequence of VaR forecasts at confidence level $(1 - \alpha)$ has correct conditional coverage (CC) if the V_t form an independent and identically distributed sequence of Bernoulli random variables with parameter α. The test will reject the null of correct CC if the fraction of VaR violations ($V_t = 1$) is much lower or much higher than α, reflecting that the VaR forecasts are generally (in an absolute sense) too large or too small, or if the VaR violations occur in clusters, reflecting that the VaR forecasts are too small in specific sub-periods.

To test the ability of our models to capture the true Expected Shortfall, we analyze the *standardized residuals* corresponding to returns that exceed the predicted VaR:

$$e_t \doteq \frac{r_t - \text{ES}_{t|t-1}}{\sigma_t} \qquad \text{with } r_t < \text{VaR}_{t|t-1} . \tag{2}$$

4

If a model yields correct ES forecasts, then the (conditional) mean of these e_t (with $r_t <$ $\text{VaR}_{t|t-1}$) should be zero. However, even for a model that yields correct ES forecasts, the standardized residuals may be clustered, ruling out the i.i.d. hypothesis for these e_t. For this reason, we use the block-bootstrap test of Jalal and Rockinger (2008), which makes no assumption about the standardized residuals' underlying dynamics. The test extends the usual bootstrapped t-test from the algorithm of Efron and Tibshirani (1993, page 224) to the circular blocks bootstrap (Politis and Romano, 1992; Shao and Yu, 1993). In the latter approach the data are wrapped around as a circle, so that each of the original observations has an equal chance of appearing in a simulated series and all the blocks have the same length. The choice of the block length is crucial in any block-bootstrap method: the block length should be large enough so that it includes most of the dependence structure, but not too large so that the number of blocks becomes insufficient. We select the block length following the algorithm based on the spectral density estimation as proposed by Politis and White (2004).

We will estimate the percentage of the time series for which a model provides correct VaR forecasts (in the sense of correct conditional coverage) and correct ES forecasts (in the sense that the standardized residuals e_t in (2) have mean zero). A naive way to estimate this percentage is to compute the percentage of time series for which the p-value is above a preset significance level, say 5%. However, this approach obviously suffers from Type I errors (rejection for approximately 5% of those time series for which the model performs correctly) and Type II errors (non-rejection for some – or possibly many – of those time series for which the model performs incorrectly). Therefore, the naive estimate may underestimate or overestimate the number of time series for which the model has correct performance, respectively. We therefore correct the percentage of non-rejections using the false discovery rate method of Storey (2002).

The key insight of the false discovery rate method is that in case of a model that delivers correct forecasts for a certain time series the p-value is uniformly distributed between zero and one (see, e.g. Barras et al., 2010). Otherwise, the p-value has an unknown distribution, which should be relatively close to zero. Let λ be the separating value such that for p-values above λ, it is almost certain that they correspond to the null of a correctly performing model. By the properties of the uniform distribution, we can therefore extrapolate the true number of correctly performing models from the p-values exceeding the λ threshold. We rely on the bootstrap

method proposed by Barras et al. (2010) to determine the optimal value of λ in a purely data driven way.

3. Results and discussion

We test the performance of the models on constituents (as of June 29, 2012) of the S&P 500 index for a period ranging from January 3, 2000, to June 29, 2012, thus representing more than twelve years of daily data. The data are then filtered for liquidity following Lesmond et al. (1999). In particular, we remove the time series with less than 1500 data points history, with more than 10% of zero returns and more than two trading weeks of constant price. This filtering approach reduces the database to 466 equities for which the adjusted daily closing prices are downloaded from Datastream.

Table 1 reports the forecasting results, from which we draw the following conclusions. First, the impact of the updating frequency (of the estimated model parameters) on the quality of the forecasts of VaR and ES is remarkably small. For each model and for each risk measure, we notice only marginal differences in performance as the updating frequency decreases from daily to weekly or monthly. Moreover, the confidence bands overlap, reflecting that the performance is not significantly different. This suggests that a risk system for equities based on a GARCH-type model could be updated every month only, without altering significantly the forecasting performance in terms of VaR and ES. Even quarterly parameter updates seem sufficient, except for the prediction of the 95% VaR using a GJR-Student model (with a substantial and significant drop from 70% to 46% of the time series for which the GJR-Student model yields correct 95% VaR forecasts).

Second, the asymmetric GARCH model with non-parametric kernel density estimate (GJR-Kernel) performs strikingly well; it yields correct VaR and ES forecasts for an estimated 90% to 95% of the S&P 500 constituents.

Third, specifying a Student-t (or Gaussian) innovations' density yields substantially and significantly worse forecasts. Especially for ES, the use of a flexible distribution for the innovations u_t in (1), such as a kernel density estimate, is crucial. These *empirical* results are in line with the interesting findings from the extensive analysis of *simulated* data sets by Jalal and Rockinger (2008). Jalal and Rockinger (2008) estimate GARCH models with Gaus-

6

Model	Estimation Frequency	95% VaR	99% VaR	95% ES	99% ES
GARCH-Normal	daily	30	5	1	1
		[23;37]	[2;8]	[0;3]	[0;2]
GARCH-Normal	weekly	31	4	1	1
		[24;39]	[1;6]	[0;3]	[0;3]
GARCH-Normal	monthly	30	4	2	1
		[23;37]	[1;6]	[0;3]	[0;3]
GARCH-Normal	quarterly	28	3	1	1
		[22;35]	[1;5]	[0;2]	[0;2]
GJR-Normal	daily	43	4	1	1
		[34;51]	[1;6]	[0;2]	[0;2]
GJR-Normal	weekly	42	3	1	1
		[34;50]	[1;5]	[0;2]	[0;2]
GJR-Normal	monthly	41	4	1	1
		[33;50]	[1;7]	[0;2]	[0;1]
GJR-Normal	quarterly	41	3	1	1
		[32;49]	[1;5]	[0;1]	[0;1]
GJR-Student	daily	70	75	3	33
		[64;75]	[69;80]	[1;5]	[26;41]
GJR-Student	weekly	70	75	4	31
		[64;75]	[70;81]	[1;6]	[24;39]
GJR-Student	monthly	65	71	3	35
		[60;70]	[66;77]	[1;5]	[27;43]
GJR-Student	quarterly	46	65	4	32
		[37;54]	[60;71]	[1;6]	[25;40]
GJR-Kernel	daily	91	94	96	95
		[86;96]	[89;99]	[91;100]	[90;100]
GJR-Kernel	weekly	90	95	96	97
		[85;96]	[89;100]	[91;100]	[92;100]
GJR-Kernel	monthly	89	94	92	99
		[84;95]	[89;99]	[86;97]	[94;100]
GJR-Kernel	quarterly	88	89	95	99
		[82;93]	[84;95]	[90;100]	[94;100]
EWMA		27	1	1	0
		[20;34]	[0;3]	[0;2]	[0;0]

Table 1: False discovery rate results: Estimated percentages of the time series for which the model provides correct Value-at-Risk forecasts (using the *conditional coverage* test of Christoffersen (1998)) and correct Expected Shortfall forecast (using the block-bootstrap test of ES violations of Jalal and Rockinger (2008)). Percentages are computed from the set of p-values for the series using the *false discovery rate* approach of Storey (2002). []: asymptotically valid 90% confidence bands derived in Barras et al. (2010). GARCH-Normal: symmetric GARCH(1,1) with Gaussian innovations; GJR-Normal: asymmetric GJR(1,1) with Gaussian innovations; GJR-Student: asymmetric GJR(1,1) with Student-t innovations; GJR-Kernel: asymmetric GJR(1,1) with non-parametric kernel density estimate. EWMA: Exponentially Weighted Moving Average of RiskMetrics.

Model	Estimation Frequency	95% VaR	99% VaR	95% ES	99% ES
GARCH-Normal	daily	66	22	3	4
		[62;69]	[19;25]	[2;5]	[3;6]
GARCH-Normal	weekly	64	22	4	3
		[60;68]	[19;25]	[2;5]	[2;5]
GARCH-Normal	monthly	64	20	4	4
		[60;68]	[17;23]	[2;5]	[3;6]
GARCH-Normal	quarterly	64	18	3	4
		[60;68]	[15;21]	[2;5]	[3;6]
GJR-Normal	daily	76	22	4	5
		[73;80]	[19;25]	[3;6]	[3;7]
GJR-Normal	weekly	74	22	5	5
		[70;77]	[19;26]	[3;6]	[3;7]
GJR-Normal	monthly	74	20	4	4
		[70;77]	[17;23]	[3;6]	[3;6]
GJR-Normal	quarterly	74	17	4	4
		[71;77]	[14;20]	[2;5]	[2;5]
GJR-Student	daily	84	82	5	54
		[82;87]	[79;85]	[3;7]	[50;58]
GJR-Student	weekly	84	83	6	52
		[82;87]	[80;86]	[4;7]	[48;56]
GJR-Student	monthly	84	81	6	51
		[81;87]	[78;84]	[4;7]	[47;55]
GJR-Student	quarterly	81	77	6	51
		[78;84]	[74;80]	[4;8]	[47;55]
GJR-Kernel	daily	91	93	98	96
		[89;93]	[91;95]	[97;99]	[94;97]
GJR-Kernel	weekly	89	93	98	96
		[87;91]	[91;95]	[97;99]	[94;97]
GJR-Kernel	monthly	88	92	98	95
		[86;90]	[90;94]	[97;99]	[93;97]
GJR-Kernel	quarterly	86	91	96	94
		[83;89]	[89;93]	[94;97]	[92;96]
EWMA		75	3	2	3
		[72;78]	[2;4]	[1;3]	[2;5]

Table 2: Naive results: Percentages of non-rejections at 5% significance level for the *conditional coverage* test of Christoffersen (1998) for the 95% and 99% Value-at-Risk and percentages of non-rejections at 5% significance level of the block-bootstrap test of ES violations of Jalal and Rockinger (2008). []: asymptotically valid 90% confidence bands. GARCH-Normal: symmetric GARCH(1,1) with Gaussian innovations; GJR-Normal: asymmetric GJR(1,1) with Gaussian innovations; GJR-Student: asymmetric GJR(1,1) with Student-*t* innovations; GJR-Kernel: asymmetric GJR(1,1) with non-parametric kernel density estimate. EWMA: Exponentially Weighted Moving Average of RiskMetrics.

sian and Student-t errors, and apply the conditional Extreme Value Theory (EVT) approach of McNeil and Frey (2000). They consider huge numbers of simulated data sets from several models such as the EGARCH model of Nelson (1991), a stochastic volatility with jumps model, a regime-switching model, and a pure jumps model. For several of these data generating processes, the conditional EVT approach – where a generalized Pareto distribution (GPD) with two parameters is estimated for the relevant tail of the distribution of u_t – yields substantially better ES forecasts than the less flexible GARCH model with Student-t errors, in which merely one parameter (degrees-of-freedom) is estimated to fit the whole distribution of u_t. Figure 1 and Figure 2 show histograms of the 466 averages of the ratio $y_t/\text{ES}_{t|t-1}$ (at risk levels 95% and 99%, respectively) for observations with $y_t < \text{VaR}_{t|t-1}$ for the GJR-Student model (top) and GJR-Kernel model (bottom). For the GJR-Student model the average ratios are mostly smaller than one: the GJR-Student model fails due to its *over*estimation (in an absolute sense) of the ES. This result is also found in the simulation experiments of Jalal and Rockinger (2008). For the GJR-Kernel model the average ratios lie mostly around one. For the VaR, the difference in forecast quality is much smaller. Note the similarity of the conditional EVT approach and the GJR model with kernel density estimate: in both approaches the relevant tail of the distribution of u_t is estimated on the basis of only observations from this tail. The kernel density estimate has a similar (or arguably even larger) flexibility than the GPD. Further, the kernel density estimate has the advantage that it does not contain parameters which need to be estimated.

Fourth, the Exponentially Weighted Moving Average (EWMA) of RiskMetrics performs worst. In sum, the GJR-Kernel model with monthly or quarterly updated parameter estimates yields much better VaR and ES forecasts than simpler models with daily updated parameter estimates or the EWMA. Further, it seems intuitively clear that the test for correct ES has more power than the test for correct VaR, since the first takes into account how large excesses over the VaR are, whereas the latter is merely based on whether the VaR is exceeded.

Table 2 shows the naive estimate of the percentage of stocks for which a model yields correct VaR or ES forecasts, at the 5% significance level. For several models and risk measures, the influence of Type II errors is substantial. For example, the 95% VaR forecasts from EWMA are not rejected for 75% of the time series, whereas the FDR analysis indicates that the VaR is
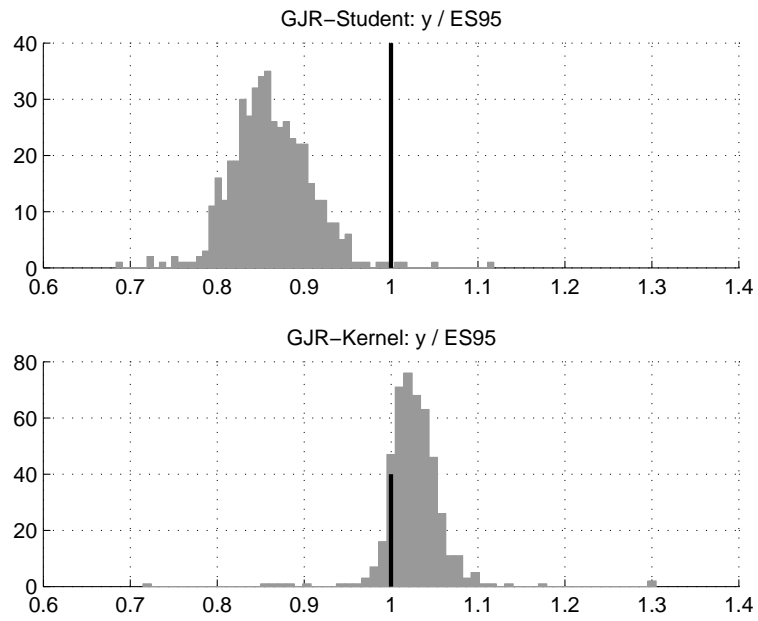
9

Figure 1: Histogram of the 466 averages of the ratio $y_t/\text{ES}_{t|t-1}$ (at risk level 95%) for observations with $y_t < \text{VaR}_{t|t-1}$ for the GJR-Student model (top) and GJR-Kernel model (bottom).
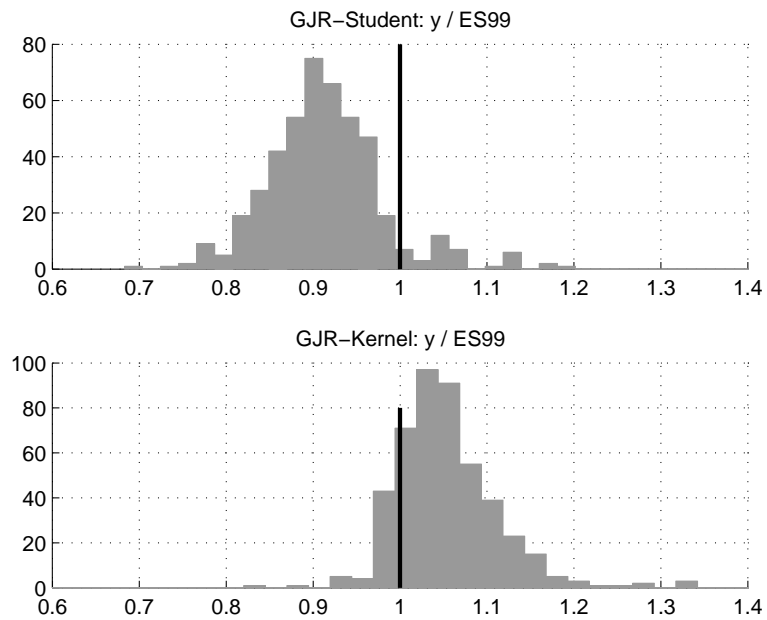

Figure 2: Histogram of the 466 averages of the ratio $y_t/\text{ES}_{t|t-1}$ (at risk level 99%) for observations with $y_t < \text{VaR}_{t|t-1}$ for the GJR-Student model (top) and GJR-Kernel model (bottom).

only correct for approximately 25% of the stocks. This suggests that a Type II error occurs for roughly half of the time series of EWMA 95% VaR forecasts. Also the quality of VaR forecasts from models with Gaussian or Student-t errors would be seriously overestimated, if one would base one's conclusions on Table 2. The differences between Table 1 and Table 2 reflect the relatively low power of the test for correct conditional coverage of the VaR, but also the usefulness of the false discovery rate methodology, especially when faced with a test suffering from low power. On the other hand, for the GJR-Kernel model the estimated percentages in Table 1 and Table 2 are close: here the effect of Type II errors is relatively small and seems to be approximately canceled by the effect of Type I errors.

Figure 3 and Figure 4 graphically illustrate the differences between the results from the FDR approach and the naive results. Figure 3 shows that for the GJR-Student model the hypothesis of correct 99% ES forecasts is rejected for 46% of the stocks (at a 5% significance level), leading to a naive estimate that the model yields correct ES forecasts for 54% of the time series. However, the FDR approach extrapolates the number of p-values above λ, where the optimized value of λ is approximately 0.7 in this case; that is, the number of time series with correct 99% ES forecasts is estimated as the number of p-values above λ multiplied by $\frac{1}{1-\lambda}$. This leads to the substantially lower estimate of 33% of the stocks. Especially, many of the p-values between 0.05 and 0.20 are 'classified' as p-values corresponding to time series for which the GJR-Student model yields incorrect 99% ES forecasts. On the other hand, Figure 4 for the GJR-Kernel model shows a histogram of the 466 p-values that is much closer to a uniform distribution. The FDR and naive approaches yield approximately the same estimates for the percentage of time series for which the GJR-Kernel model yields correct 99% ES forecasts. The naive approach seems to suffer equally from Type I errors and Type II errors – both for approximately 5% of the stocks.
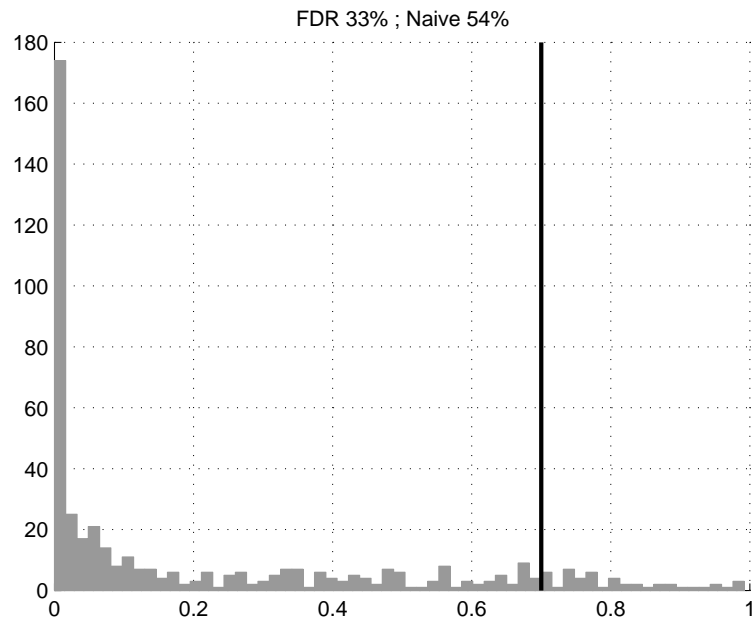
Figure 3: Histogram of the 466 p-values of the 99% ES test for the GJR-Student model. The vertical line indicates the optimal λ obtained by the method proposed by Barras et al. (2010). The percentage stocks for which we do not have rejection of the test at the 5% level is 33% for the FDR approach and 54% for the naive approach.

Figure 4: Histogram of the 466 p-values of the 99% ES test for the GJR-Kernel model. The vertical line indicates the optimal λ obtained by the method proposed by Barras et al. (2010). The percentage stocks for which we do not have rejection of the test at the 5% level is 95% for the FDR approach and 96% for the naive approach.
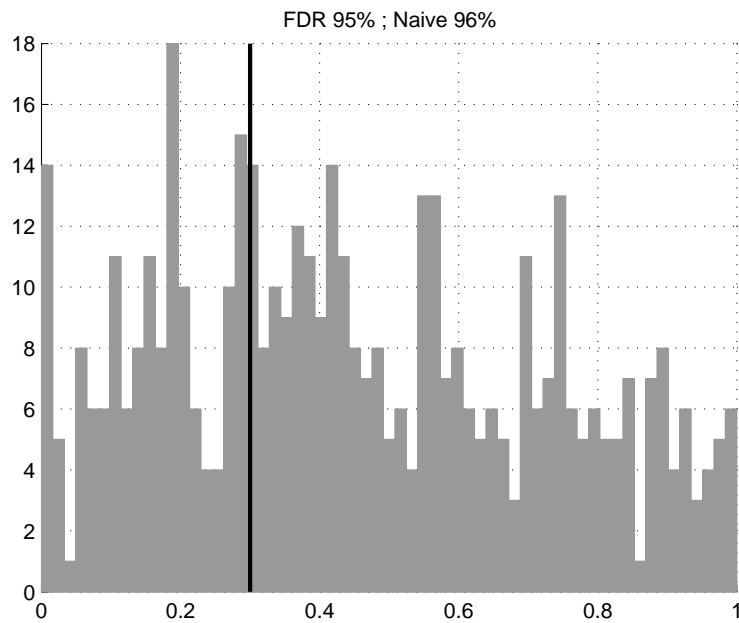
4. Final remarks

The following possible extensions are left as topics for further research. First, different model specifications can be considered. The symmetric Student-t distribution can be replaced by the skewed Student t distribution of Hansen (1994) or the generalized error distribution. Further, the performance of the conditional Extreme Value Theory (EVT) approach of McNeil and Frey (2000), which shows good results in the extensive simulation study of Jalal and Rockinger (2008), can also be assesses for our large empirical data set. Different GARCH-type models can be considered such as the EGARCH model of Nelson (1991) or the Markov-Switching GARCH Model (Haas et al., 2004; Ardia, 2008). Alternatively, the several types of the increasingly popular stochastic volatility models (see e.g., Harvey and Shephard (1996)) can be analyzed.

Second, different risk measures can be analyzed. For example, the VaR and ES can be predicted for higher confidence levels such as 99.5% or 99.9%. Especially for very high confidence levels (regarding the very *deep* tail) a comparison between the performance of the GJR model with non-parametric kernel density estimate and the conditional EVT approach may be interesting, as the latter approach – particularly designed as a tool to assess the likelihood of rare but large events – may be considered theoretically superior in such cases.

Third, stocks of different countries or continents can be considered instead of the S&P 500 constituents. In less mature markets, it may be necessary to update one's parameter estimates more frequently. Fourth, one can use different tests such as the recently developed forecast rationality test of Patton and Timmermann (2012), extended to the assessment of VaR forecasts by Hoogerheide et al. (2012c).

Fifth, the models can be estimated in the Bayesian framework (instead of the classical/ frequentist framework), see e.g., Hoogerheide and Van Dijk (2010) or Hoogerheide et al. (2012a). Then for each model parameter a whole posterior distribution of values will be simulated, replacing the single value of the quasi maximum likelihood estimate. The effect of the number of simulations or different simulation methods (see e.g., Hoogerheide et al. (2012b)) can then also be analyzed. The effect of estimation frequency on VaR and ES can also be assessed for *combinations* of multiple models. The Bayesian framework allows for Bayesian Model Averaging (BMA) using marginal likelihoods (Ardia and Hoogerheide, 2010; Ardia et al., 2012),

as well as other model combination techniques such as the robust time-varying weights of Hoogerheide et al. (2010). This is our current research focus.

References

Ardia, D., 2008. Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications. Vol. 612 of Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, Berlin, Germany.

Ardia, D., Baştürk, N., Hoogerheide, L. F., Van Dijk, H. K., 2012. A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood. Computational Statistics & Data Analysis 56 (11), 3398–3414.

Ardia, D., Hoogerheide, L. F., 2010. Efficient Bayesian estimation and combination of GARCH-type models. In: Böcker, K. (Ed.), Rethinking Risk Measurement and Reporting - Volume II: Examples and Applications from Finance. Risk Books, London, Ch. 1.

Artzner, P., Delbaen, F., Eber, J., Heath, D., 2000. Thinking coherently. Risk 10 (11), 68–71.

Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. Journal of Finance 65, 179–216.

Black, F., 1976. The pricing of commodity contracts. Journal of Financial Economics 3 (1–2), 167–179.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31 (3), 307–327.

Bollerslev, T., Chou, R. Y., Kroner, K., 1992. ARCH modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52 (1–2), 5–59.

Christoffersen, P. F., 1998. Evaluating interval forecasts. International Economic Review 39 (4), 841–862.

Efron, B., Tibshirani, R., 1993. An Introduction to Bootstrap. Chapman and Hall, New York, USA.

Glosten, L. R., Jaganathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48 (5), 1779–1801.

Haas, M., Mittnik, S., Paolella, M. S., 2004. A new approach to markov-switching garch models. Journal of Financial Econometrics 2 (4), 493–530.

Hansen, B. E., 1994. Autoregressive conditional density estimation. International Economic Review 35 (3), 705–730.

Harvey, A. C., Shephard, N., 1996. Estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business & Economic Statistics 14 (4), 429–434.

Hoogerheide, L. F., Ardia, D., Corré, N., 2012a. Density prediction of stock index returns using GARCH models: Frequentist or Bayesian estimation? Economics Letters 116 (3), 322–325.

Hoogerheide, L. F., Kleijn, R., Ravazzolo, F., Van Dijk, H. K., Verbeek, M., 2010. Forecast accuracy and economic gains from Bayesian Model Averaging using time varying weights. Journal of Forecasting 29, 251–269.

Hoogerheide, L. F., Opschoor, A., Van Dijk, H. K., 2012b. A class of adaptive importance sampling weighted EM algorithms for efficient and robust posterior and predictive simulation. Journal of Econometrics 171 (2), 101–120.

Hoogerheide, L. F., Ravazzolo, F., Van Dijk, H. K., 2012c. Comment on forecast rationality tests based on multi-horizon bounds. Journal of Business & Economic Statistics 30 (1), 30–33.

Hoogerheide, L. F., Van Dijk, H. K., 2010. Bayesian forecasting of value at risk and expected shortfall using adaptive importance sampling. International Journal of Forecasting 26 (2), 231–247.

Jalal, A., Rockinger, M., 2008. Predicting tail-related risk measures: The consequences of using GARCH filters for non-GARCH data. Journal of Empirical Finance 15, 868–877.

Lesmond, D. A., Ogden, Joseph, P., Trzinka, C. A., 1999. A new estimate of transaction costs. The Review of Financial Studies 12 (5), 1113–1141.

McNeil, A. J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance 7 (3–4), 271–300.

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59 (2), 347–370.

Patton, A. J., Timmermann, A., 2012. Forecast rationality tests based on multi-horizon bounds. Journal of Business & Economic Statistics 30 (1), 1–17.

Politis, D. N., Romano, J. P., 1992. A circular block-resampling procedure for stationary data. In: LePage, R., Billard, L. (Eds.), Exploring the Limits of Bootstrap. John Wiley, New York, USA, pp. 263–270.

Politis, D. N., White, H., 2004. Automatic block-length selection for the dependent bootstrap. Econometric Reviews 23, 53–70.

RiskMetrics Group, 1996. RiskMetrics Technical Document. J. P. Morgan/Reuters, 4th Edition.

Shao, Q.-M., Yu, H., 1993. Bootstrapping the sample means for stationary mixing sequences. Stochastic Processes and their Applications 48 (1), 175–190.

Silverman, B. W., 1986. Density Estimation for Statistics and Data Analysis, 1st Edition. Chapman and Hall, New York, USA.

Storey, J., 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society B 64, 479–498.