

TI 2014-090/III  
Tinbergen Institute Discussion Paper



# Improving Density Forecasts and Value-at-Risk Estimates by Combining Densities

*Anne Opschoor*<sup>1,3</sup>

*Dick van Dijk*<sup>2,3</sup>

*Michel van der Wel*<sup>2,3</sup>

<sup>1</sup> Faculty of Economics and Business Administration, VU University Amsterdam, the Netherlands;

<sup>2</sup> Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands;

<sup>3</sup> Tinbergen Institute, the Netherlands.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

# Improving Density Forecasts and Value-at-Risk Estimates by Combining Densities

Anne Opschoor\*      Dick van Dijk      Michel van der Wel

This version: July 15, 2014<sup>†</sup>

## Abstract

We investigate the added value of combining density forecasts for asset return prediction in a specific region of support. We develop a new technique that takes into account model uncertainty by assigning weights to individual predictive densities using a scoring rule based on the censored likelihood. We apply this approach in the context of recently developed univariate volatility models (including HEAVY and Realized GARCH models), using daily returns from the S&P 500, DJIA, FTSE and Nikkei stock market indexes from 2000 until 2013. The results show that combined density forecasts based on the censored likelihood scoring rule significantly outperform pooling based on the log scoring rule and individual density forecasts. The same result, albeit less strong, holds when compared to combined density forecasts based on equal weights. In addition, VaR estimates improve at the short horizon, in particular when compared to estimates based on equal weights or to the VaR estimates of the individual models.

**Keywords:** Density forecast evaluation, Volatility modeling, Censored likelihood, Value-at-Risk.

**JEL:** C53, C58, G17.

---

\*Corresponding author, e-mail address: a.opschoor@vu.nl. Postal address: VU University Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands. Phone number: +31(0)20-5982663.

<sup>†</sup>Anne Opschoor is from VU University Amsterdam and the Tinbergen Institute. Van Dijk is from Erasmus University Rotterdam, Tinbergen Institute and ERIM. Van der Wel is from Erasmus University Rotterdam, CREATES, Tinbergen Institute and ERIM. We appreciate the comments of participants at the 34th International Symposium on Forecasting (Rotterdam, July 2014), the 22nd Annual Society for Nonlinear Dynamics and Econometrics (SNDE) Symposium (New York, April 2014), the Sofie Workshop on Skewness, Heavy Tails, Market Crashes and Dynamics (Cambridge, April 2014) and the 4th Amsterdam-Bonn Workshop in Econometrics (Amsterdam, October 2013). Michel van der Wel is grateful to Netherlands Organisation for Scientific Research (NWO) for a Veni grant; and for support from CREATES, funded by the Danish National Research Foundation. We are responsible for all errors.

# 1 Introduction

Value-at-Risk (VaR) is a commonly used measure of downside risk for investments. Financial institutions are allowed by regulation (i.e. the Basel accords) to report VaR estimates for their asset portfolios obtained from their own “internal” model. An important related issue in this estimation is *model uncertainty*, as each model has its prespecified known form and takes no account of possible uncertainty regarding the model structure. In addition, given the availability of a considerable number of different risk-management methods, based on academic literature and/or his expertise, it is a difficult task for a decision-maker to choose the “best” model. Moreover, each model is an incomplete description of reality. Hence relying upon a single model is dangerous when constructing a VaR, i.e. a density forecast in the left tail, as any model is “wrong” in some sense.

In this paper, we investigate the usefulness of combining density forecasts with the focus on a particular region of the density. This is motivated in the first place by well known advantages of combining point or density forecasts.<sup>1</sup> We aim to obtain more accurate VaR estimates and density forecasts in the left tail. This motivates the investigation of combining density forecasts based on their behavior in the left tail as using the whole density does not necessarily lead to the same quality of forecasts as focusing purely on the left tail, which is the case for the VaR. Therefore, we develop a density forecast combination method that extends the method of Geweke and Amisano (2011), which uses the *whole* density, by considering the censored likelihood (*csl*) scoring rule of Diks *et al.* (2011) that focuses on a region of the densities’ support of particular interest, such as the left tail.

We use our novel methodology in an empirical application involving several recently developed univariate volatility models. Hence, as a second contribution to the literature, we make a comparison between these models with respect to their predictive ability in terms of density forecasts. In particular, beyond the traditional (Threshold) GARCH model (Bollerslev, 1986; Glosten *et al.*, 1993), we consider the Heavy model (Shephard and Sheppard, 2010) and the Realized GARCH model (Hansen *et al.*, 2012) that include realized measures, as well as the GAS model (Creal *et al.*, 2013). All models are applied to daily returns on the S&P 500, DJIA, FTSE and Nikkei stock market indexes from 2000 until 2013.

---

<sup>1</sup>We discuss this literature in more detail below.

We evaluate the added value of combining density forecasts both statistically and economically. First, we test equal predictive accuracy in the left tail of a combined density forecast based on our new method and three alternatives: (i) the method based on the whole density, (ii) a benchmark that consists of equal weights, and (iii) the density forecast of each individual model. Second, we compare 1- and 5-day VaR estimates based on these methods using the Unconditional Coverage (UC) test and the Independence test of Christoffersen (1998). In addition we test on equal accuracy with the use of an asymmetric tick-loss function and the test procedure of Giacomini and White (2006).

Our results show statistically that density forecasts in the tail are more accurate if we pool density forecasts using the *csl* scoring rule than using the log score based on the whole density, using equal weights or using the density forecast of any individual volatility model. This gain in accuracy holds in particular when the density forecasts of our new method are compared to the density forecasts based on the log score function; the gain with respect to the equal weights is only apparent in case of the FTSE and Nikkei index. In addition, the 95% one-day VaR estimates improve significantly compared to all univariate volatility models (except the HEAVY Skewed- $t$  model), such that either less violations are made and the unconditional coverage matches more closely the nominal value, or according to the asymmetric tick-loss function. Moreover, the new combination method outperforms the pooling method based on equal weights for most stock indexes (by means of the UC test or the tick-loss function) but outperforms the pooling method based on the log score function only in case of the DJIA returns. Finally, we show that the combination weights based on the *csl* scoring rule differ considerably from the weights obtained by using the whole density. Hence, a certain volatility model could get no or less weight based on the log scoring rule, but may be useful in our new method.

We contribute to the literature on combining forecasts. Starting with the seminal work of Bates and Granger (1969), combining point forecasts appears to be a successful forecasting strategy, improving upon individual forecasts. Timmermann (2006) surveys the literature on forecast combination and shows from a theoretical point of view why forecast combinations could work well. This is confirmed by numerous empirical applications in different areas including macroeconomic and financial forecasting. For example, forecasting output growth using individual predictors typically delivers forecasts that are unstable over time.

Combining forecasts offers more stable forecasts which improve upon autoregressive forecasts (Stock and Watson, 2004). Rapach *et al.* (2010) provide similar evidence in the context of equity premium prediction, by showing that combining forecasts leads to statistically and economically significant out-of-sample gains relative to the historical average return.

Although the literature shows the usefulness of combining point forecasts, point forecasts themselves are not very informative if there is no indication of their uncertainty (see Granger and Pesaran, 2000; Garratt *et al.*, 2003). This finding has led to a growing interest in *density* forecasts, which represent a full predictive distribution of a random variable and hence provide the most complete measure of this uncertainty. It is a natural step forward to bring together the concepts of forecast combinations and density forecasts. The literature on combining density forecasts is yet scarce, although the interest in this topic grows with applications in, for example, macro-economics (Jore *et al.*, 2010; Aastveit *et al.*, 2011).

One of the earliest examples of combining density forecasts is Wallis (2005), who considers a finite mixture distribution, taking a weighted linear combination of multiple density forecasts. Hall and Mitchell (2007) address the issue how to choose the weights assigned to each competing density. They propose a methodology with the aim to obtain the most accurate density forecast from a statistical point of view. This boils down to using the logarithmic scoring rule, which takes the log of the predictive density evaluated at the observed value of the variable of interest. Closely related is the work of Geweke and Amisano (2011), who use the logarithmic scoring rule to obtain weights to form optimal linear combinations of predictive densities. We extend this approach, by substituting the log score rule by the censored likelihood scoring rule. Krüger (2014) studies the added value of combining density forecasts by comparing various scoring rules from both a theoretical and empirical point of view, with an application to combined density forecasts of the UK inflation. Recently, the literature departs from linear density combinations to non-linear opinion pools (Gneiting and Ranjan, 2013; Fawcett *et al.*, 2014). Another extension is to specify a dynamic structure for the combination weights (Del Negro *et al.*, 2013; Billio *et al.*, 2013). Although this paper considers static weights, by means of re-optimizing the weights using a rolling window, we still capture the possible time-varying weight dynamics.

The remainder of this paper is organized as follows. Section 2 puts forward our methodology of combining density forecasts using the *csl* scoring rule. In Section 3, we provide

an overview of the univariate volatility models and the related assumed conditional density functions, which are used in the empirical application (Section 4). Section 5 concludes.

## 2 Combining density forecasts

Suppose a decision maker has  $n$  different models for a variable of interest  $y$ . Conditional on information available up to and including time  $t - 1$ , the predictive density corresponding with a particular model at time  $t$  is of the form  $p_t(y_t|I_{t-1}, \theta_{A_i}, A_i)$ , where  $I_{t-1}$  indicates the information set up to and including time  $t - 1$ ,  $A_i$  denotes the particular model  $i$ , ( $i = 1, \dots, n$ ) and  $\theta_{A_i}$  the estimated parameters of model  $A_i$  given  $I_{t-1}$ . Suppose further that the decision maker aims to choose the best predictive density at time  $T + 1$ , given the available density forecasts from time  $t = 1, \dots, T$ . An often used approach is to make use of scoring rules. A scoring rule measures the quality of density forecasts by assigning a numerical score. Typically, this rule is a objective function that depends on the density forecast and the actually observed value, such that a higher score is associated with a “better” density forecast. According to Gneiting and Raftery (2007), a scoring rule is *proper* if it satisfies the condition that incorrect density forecasts do not receive a higher average score than the true density. This property is important and a natural requirement for any rational decision maker.

A well founded scoring rule is the log score function (see Mitchell and Hall, 2005; Amisano and Giacomini, 2007). This function for a particular model  $A_i$  at a specific time  $t$  is defined as

$$S^l(y_t; A_i) = \log p_t(y_t|I_{t-1}, A_i), \tag{1}$$

with  $S^l$  the abbreviation of the log scoring rule, which simply takes the logarithm of the predictive density evaluated at  $y_t$ . This scoring rule is closely related to information theoretic goodness-of-fit measures such as the Kullback-Leibler Information Criterion (KLIC) associated with the density forecast  $p_t(y_t|I_{t-1}, A_i)$ . It can be shown that a higher value of the logarithmic score coincides with a lower value of the KLIC. Put differently, maximizing the logarithmic score is equivalent with minimizing the KLIC.

It is highly unlikely that one model is the true model for constructing a predictive density. Hence it might be beneficial to combine various densities. The literature dates back to Bacharach (1974), who considers linear combinations of (subjective) probability distributions, known as *linear opinion pools*. Hall and Mitchell (2007) and Geweke and Amisano (2011) also study linear opinion pools and introduce this idea into the econometric forecasting literature. In particular, Geweke and Amisano (2011) combine predictive densities using the log score function as stated in (1), and consider predictive densities of the form

$$\sum_{i=1}^n w_i p_t(y_t | I_{t-1}, A_i), \quad (2)$$

for  $i = 1, \dots, n$  and weights  $w_i$ , restricted such that they are positive and sum to one to ensure that (2) is a valid probability density function. It is natural to choose the weights in such a way that the log score function in (1) is maximized (and hence the KLIC is minimized):

$$S^l(Y_T, C) = \sum_{t=1}^T \log \left[ \sum_{i=1}^n w_i p_t(y_t | I_{t-1}, A_i) \right], \quad (3)$$

with  $Y_T = \{y_1, \dots, y_T\}$ , and  $C$  representing the fact that a combination of models is evaluated instead of a single model  $A_i$ .

The main idea of this paper is to extend the approach of Geweke and Amisano (2011) by focusing on a particular region of interest of the predictive density. In order to do so, we consider a scoring rule based on the censored likelihood (*csl*), advocated by Diks *et al.* (2011). They prove that this scoring rule is proper and show the usefulness of this scoring rule if one is interested in the accuracy of density forecasts in a specific region. In this study, the focus is on the left tail, which is important for risk management purposes. The *csl* score function for a specific region  $B_t$  for model  $A_i$  at time  $t$  reads

$$\begin{aligned} S^{csl}(y_t | A_i) &= I[y_t \in B_t] \log p_t(y_t | I_{t-1}, A_i) \\ &\quad + I[y_t \in B_t^c] \log \left( \int_{B_t^c} p_t(y | I_{t-1}, A_i) dy \right) \end{aligned} \quad (4)$$

with  $B_t^c$  the complement of  $B_t$  and  $I[\cdot]$  an indicator function that takes the value 1 if the



argument is true. The first part of this scoring rule focuses on the behavior of the density forecast in the region of interest  $B_t$ . The second part computes the cdf of the density in the region outside  $B_t$ . Hence any observation outside  $B_t$  ignores the shape of  $p_t(y_t|I_{t-1}, A_i)$  outside  $B_t$ . Note that (4) simplifies to the log scoring rule of (1) if  $B_t$  represents the full sample space.

The next step is combine the predictive densities based on the *csl* scoring rule. That is, we consider again predictive densities as defined in (2), however with the weights obtained by optimizing the corresponding censored likelihood score function over the values  $Y_T = \{y_1, \dots, y_T\}$ :

$$S^{csl}(Y_T, C) = \sum_{t=1}^T \log \left[ \sum_{i=1}^n w_i \left( I[y_t \in B_t] p_t(y_t|I_{t-1}, A_i) + I[y_t \in B_t^c] \int_{B_t^c} p_t(y|I_{t-1}, A_i) dy \right) \right]. \quad (5)$$

We end this section by a brief comment about the optimization of the weights  $w_t$  in (3) and (5). Although (numerical) constrained optimization techniques may be used, we consider the algorithm of Conflitti *et al.* (2012). This iterative algorithm is easy to implement works well even when the number of forecasts to combine gets large. See Appendix A for more details.

### 3 Models and distributions

This study focuses on density forecasting in the context of univariate volatility models. We consider several classes of models, including the standard GARCH model of Bollerslev (1986) and the Threshold GARCH extension of Glosten *et al.* (1993), the HEAVY model of Shephard and Sheppard (2010), the Realized GARCH model of Hansen *et al.* (2012) and the GAS model of Creal *et al.* (2013). All models are based on the following general specification for  $y_t$ , the return for a financial asset at day  $t$ :

$$y_t = \mu + \sqrt{h_t} z_t, \quad \text{with } z_t|I_{t-1} \sim D(0, 1), \quad (6)$$

where  $\mu$  denotes the conditional mean of the returns,  $h_t$  the conditional variance and  $z_t$  the standardized unexpected return following a certain conditional distribution  $D(\cdot)$  with mean zero and unit variance. For ease of exposition, we assume the conditional mean is constant over time, although it could easily be extended to a time-varying mean  $\mu_t$ . Further,  $I_t$  denotes the information set up to and including time  $t$ . The following subsections differentiate between various specifications for the dynamics of  $h_t$  and possible choices for the conditional return density function  $D(\cdot)$ .

### 3.1 Univariate volatility models

The first model we consider is the traditional GARCH(1,1) model (Bollerslev, 1986) for the conditional variance  $h_t$ :

$$h_t = \omega + \alpha(y_{t-1} - \mu)^2 + \beta h_{t-1}, \quad (7)$$

with  $\omega > 0$ ,  $\alpha > 0$  and  $\beta > 0$  to ensure a positive variance. The past squared demeaned return in this model is the innovation for the conditional variance. We also consider the Threshold GARCH model of Glosten *et al.* (1993), which is given by

$$h_t = \omega + \alpha(y_{t-1} - \mu)^2 + \gamma(y_{t-1} - \mu)^2 I[y_{t-1} - \mu < 0] + \beta h_{t-1}. \quad (8)$$

This model captures the so called ‘leverage effect’, which is often empirically found in stock returns: a negative return increases volatility more than a positive return of the same magnitude. Although many extensions of the GARCH model are proposed (e.g. the EGARCH model Nelson, 1991), we only consider this extension because of its popularity. We restrict also the other considered model classes in this study to the basis specification, although many variants/extensions are possible. The reason is that the main goal is to compare model *classes* combined with distributions, and not models *within* a specific class.

Creal *et al.* (2013) develop a broader set of models which also includes the GARCH model of (7), namely the Generalized Autoregressive Score (GAS) models. The key property of these models is that innovations for time-varying parameters are based on the score of the

probability density function at time  $t$ .<sup>2</sup> In terms of our univariate volatility models, the time-varying parameters are the conditional variances  $h_t$ . The GAS(1,1) model proposes the following structure for  $h_t$ :

$$\begin{aligned} h_t &= \omega + \alpha s_{t-1} + \beta h_{t-1}, \\ s_t &= Q_t \nabla_t, \\ \nabla_t &= \frac{\partial \log p(y_t | h_t, I_{t-1}, \theta)}{\partial h_t}, \end{aligned} \tag{9}$$

with  $p(y_t | h_t, I_{t-1}, \theta)$  the conditional return density,  $\theta$  the parameter vector,  $\nabla_t$  the score and  $Q_t$  a scale factor. We follow Creal *et al.* (2013) and define the scale factor as  $1/\mathbf{E}_{t-1}[\nabla_t^2]$ , where  $\mathbf{E}_t$  denotes the expectation with respect to the return density  $p(y_t | h_t, I_{t-1}, \theta)$ . For example, when the returns  $y_t$  follow a conditional Normal distribution, the GAS model corresponds exactly to the GARCH(1,1) model of (7).<sup>3</sup> In case of a fat-tailed Student- $t$  distribution for  $y_t$ , the score based volatility model reads

$$h_t = \omega + \alpha(1 + 3/\nu) \frac{\nu + 1}{(\nu - 2) + \frac{(y_{t-1} - \mu)^2}{h_{t-1}}} (y_{t-1} - \mu)^2 + \beta h_{t-1}, \tag{10}$$

and will be labeled as the GAS- $t$  model. The specification downweights the more extreme observations, in the sense that if the distribution is more heavy tailed, it is less likely that an extreme observation is due to an increase in volatility. Note that this is a function of  $\nu$ ; when  $\nu \rightarrow \infty$ , (10) converges to the GARCH(1,1) model of (7). We again impose  $\omega > 0, \alpha > 0$  and  $\beta > 0$  in the estimation of the parameters.

The third and fourth model classes in this study include realised measures to describe the dynamics of daily volatility. A realised measure is a high-frequency estimator of the variance of a particular asset return during the times the asset is traded on an exchange. For example, the realised variance (RV) for a particular day sums the squared returns during a specific intra-day period. The intuition is that realised measures are a more accurate estimate of daily volatility than the squared daily return, as used in the GARCH models

---

<sup>2</sup>Note that we use the term ‘score’ twice: (i) a number that is assigned to measure density forecasts and (ii) here in the GAS models to indicate the derivative of the logarithm of the density with respect to a certain parameter.

<sup>3</sup>When  $y_t \sim N(0, h_t)$ ,  $\nabla_t = -0.5h_t^{-1} + 0.5h_t^{-2}y_t^2$  and  $Q_t = 2h_t^2$ . Hence the GAS model becomes  $h_t = \omega + \alpha(y_t^2 - h_t) + \beta h_t$ , which is equivalent with the GARCH model of (7).

(see Andersen *et al.*, 2003).

A recently developed model that explicitly introduces high-frequency estimators in daily volatility models is the HEAVY model of Shephard and Sheppard (2010). In particular, this model assumes the following structure for the conditional variance  $h_t$  and the expectation of the realised measure  $\xi_t = \mathbb{E}[RM_t|I_{t-1}]$ :

$$h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1}, \quad (11)$$

$$\xi_t = \omega_R + \alpha_R RM_{t-1} + \beta_R \xi_{t-1}. \quad (12)$$

All parameters should be positive to avoid negative values of  $h_t$  and  $\xi_t$ . The Heavy model is seen to consist of a GARCH structure for both  $h_t$  and  $\xi_t$ , with  $RM_t$  as innovation term. One may also include the squared (demeaned) daily return in (11), however in practice the estimate of the corresponding parameter is generally close to zero and insignificant, as noted by Shephard and Sheppard (2010). Equation (12) “completes” the system, in the sense that without this equation one can only perform one-step ahead forecasts of the conditional variance  $h$  from (11) since future values of the realised measure are unknown at time  $t$ .

A second model that relates conditional volatility with realised measures is the Realized GARCH model (RGARCH) of Hansen *et al.* (2012). The basic specification is given by:

$$h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1}, \quad (13)$$

$$RM_t = \delta + \phi h_t + \tau(z_t) + u_t, \quad (14)$$

with  $\tau(z_t)$  the leverage function, defined in the basic form as  $\tau_1 z_t + \tau_2(z_t^2 - 1)$ . This function allows for the empirical finding that negative and positive shocks may have a different impact on the volatility. Except  $\tau_1$ , which is typically negative, all parameters are restricted to be positive. The dynamics for  $h_t$  are similar for both the HEAVY and RGARCH model, however the difference arises in the specification of (the expectation of)  $RM_t$ . The HEAVY model proposes a GARCH structure for  $\mathbb{E}[RM_t|I_{t-1}]$ , while the RGARCH model explicitly relates  $RM_t$  to the conditional variance at time  $t$  and additionally introduces a leverage component.

### 3.2 Conditional distributions

We consider four possible distributions  $D(\cdot)$  of  $z_t$  in (6), which corresponds with the conditional density of the returns  $y_t$ . The starting point is the conditional Normal distribution, since this distribution is simple and often used. However, to take into account possible conditional non-normality, skewness, and excess kurtosis, we also allow the return  $y_t$  to follow a Student- $t$  distribution with mean  $\mu$ , variance  $h_t$  and  $\nu$  degrees of freedom. That is,

$$f(y_t|\mu, h_t, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{h_t(\nu-2)}\pi} \left(1 + \frac{(y_t - \mu)^2}{h_t(\nu-2)}\right)^{-\frac{\nu+1}{2}}. \quad (15)$$

The degrees of freedom  $\nu$  is treated as an unknown parameter and is estimated together with the volatility parameters. In addition,  $\nu > 2$  is required to ensure a existing variance. The excess kurtosis of the Student- $t$  distribution is equal to  $6/(\nu - 4)$ , hence it is only defined if  $\nu > 4$ . In general, a lower value of  $\nu$  implies a more fat-tailed distribution. Third, we consider the Laplace distribution, which also exhibits fatter tails than the Normal distribution, but does not involve additional parameters:

$$f(y_t|\mu, h_t) = \frac{1}{\sqrt{2h_t}} \exp\left(-\sqrt{2}\frac{|y_t - \mu|}{\sqrt{h_t}}\right) \quad (16)$$

with again mean  $\mu$  and variance  $h_t$ . Finally the Skewed- $t$  distribution of Hansen (1994) enables returns to be distributed asymmetrically, in contrast to the three symmetric distributions discussed above. For a zero mean and unit variance variable  $z_t = (y_t - \mu)/\sqrt{h_t}$ , the distribution reads

$$f(z_t; \lambda, \nu) = \begin{cases} bc \left(1 + \frac{1}{\nu-2} \left(\frac{bz_t+a}{1-\lambda}\right)^2\right)^{-\frac{\nu+1}{2}}, & \text{if } z_t < -\frac{a}{b}, \\ bc \left(1 + \frac{1}{\nu-2} \left(\frac{bz_t+a}{1+\lambda}\right)^2\right)^{-\frac{\nu+1}{2}}, & \text{if } z_t \geq -\frac{a}{b}, \end{cases} \quad (17)$$

with

$$a = 4\lambda c \frac{\nu-2}{\nu-1}, \quad b^2 = 1 + 3\lambda^2 - a^2, \quad \text{and} \quad c = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})}$$

such that  $f(y_t|\mu, h_t, \nu, \lambda) = 1/h_t f(z_t; \lambda, \nu)$ . Further,  $\lambda$  is the skewness parameter and  $\nu$  again represents the degrees of freedom. A (positive) negative value of  $\lambda$  indicates (positive)

negative skewness.

Table 1 summarizes the various choices for the dynamics of the conditional variance  $h_t$  and a conditional distribution  $D(\cdot)$ , both defined in the general specification for the daily return of (6). For the GARCH(1,1), TGARCH(1,1), HEAVY and RGARCH models, we estimate their parameters in combination with the assumption of the four described conditional distributions of this section (i.e. Normal, Student- $t$ , Laplace and Skewed- $t$ ). Further, we assume a Student- $t$  and Laplace distribution for the GAS models. This delivers 18 models in total. We estimate all models by Maximum Likelihood. This is not a computationally involved step, since we are dealing with univariate models with a maximum of 8 parameters (RGARCH models) to be estimated. In addition, we can estimate the HEAVY parameters of (11) and (12) separately, see Shephard and Sheppard (2010) for more details.

**Table 1: Overview of volatility models and conditional distributions**

This table reports the various choices for the dynamics of the conditional variance  $h_t$  and the possible conditional distributions  $D(\cdot)$ , both apparent in the general specification of (6). An “x” (“-”) denotes that an particular specification together with a conditional distribution is (not) chosen.

	Normal	Student- $t$	Laplace	Skewed- $t$
GARCH(1,1)	x	x	x	x
TGARCH(1,1)	x	x	x	x
GAS(1,1)	x <sup>a</sup>	x	x	- <sup>b</sup>
Heavy	x	x	x	x
RGARCH	x	x	x	x

<sup>a</sup> The GAS(1,1) model with Normal distributed errors is the same as the GARCH(1,1) model with Normal errors.

<sup>b</sup> The GAS(1,1) with Skewed- $t$  distributed errors is not developed yet, and is we leave it therefore as a topic of further research.

## 4 Application

This section contains the application of our new method of combining density forecasts in the context of univariate volatility models. In the following subsections, we discuss the data and implementation details, the evaluation of the density forecasts and finally the results.

## 4.1 Data and implementation details

We apply the volatility models of Section 3 to daily returns from four major stock market indexes: S&P 500, DJIA, Nikkei and the FTSE. The sample period goes from January 3, 2000 until June 28, 2013. Daily returns as well as their corresponding realized measures are obtained from the Oxford-Man Institute’s ‘realised library’.<sup>4</sup> We follow Shephard and Sheppard (2010) and use the realised kernel (see Barndorff-Nielsen *et al.*, 2008) as the realised measure at time  $t$  ( $RM_t$ ). When the exchange is closed, days are deleted from the sample.<sup>5</sup> Figure 1 shows the dynamics of the S&P 500 index and Japanese Nikkei equity index, together with the square root of the realised kernel estimate of the daily variance. The dynamics of both indexes are quite similar, however the Nikkei index contains more downward spikes (e.g. the 2011 Tohoku earthquake). Nevertheless, both return graphs clearly show the presence of conditional heteroskedasticity, since calm periods and periods of high volatility occur in an alternating pattern. Similar alternating patterns are observed for the FTSE and DJIA indexes.

We apply a rolling window scheme to estimate the model parameters and construct density forecasts. More specifically, we use an estimation window of approximately 3 years ( $T_{est} = 750$  observations), estimate the model parameters and construct 1- until 5-step ahead forecasts of  $h_t$  at each time  $t$  ( $t = T_{est}, T_{est} + 1, \dots, T - 5$ ). Given these forecasts, we also construct the corresponding 1- and 5-step ahead density forecasts. After 750 subsequent density forecasts ( $T_w = 750$ ) have been obtained for each model, we optimize (3) and (5) to obtain  $w_t$ . In case of the *csl* score function, we define the region  $B_t$  as the left tail  $y_t < \hat{r}_t^\kappa$  with  $\hat{r}_t^\kappa$  the  $\kappa$ th quantile of the empirical CDF of the 750 returns corresponding with the estimation window  $T_{est}$ . We repeat also this optimization by means of a rolling window scheme with a window of  $T_w$  density forecasts evaluations at each time  $t$  ( $t = T_{est} + T_w + 1, T_{est} + T_w + 2, \dots, T - 5$ ). We choose  $\kappa$  equal to 0.15 and 0.25 respectively. We choose  $T_{est} = 750$  such that there is a sufficient number of observations for parameter estimation of the models. Further, we emphasize the trade-off in the choice of  $\kappa$ . Given our interest in the left tail, we should take a small value of  $\kappa$ . However, the corresponding

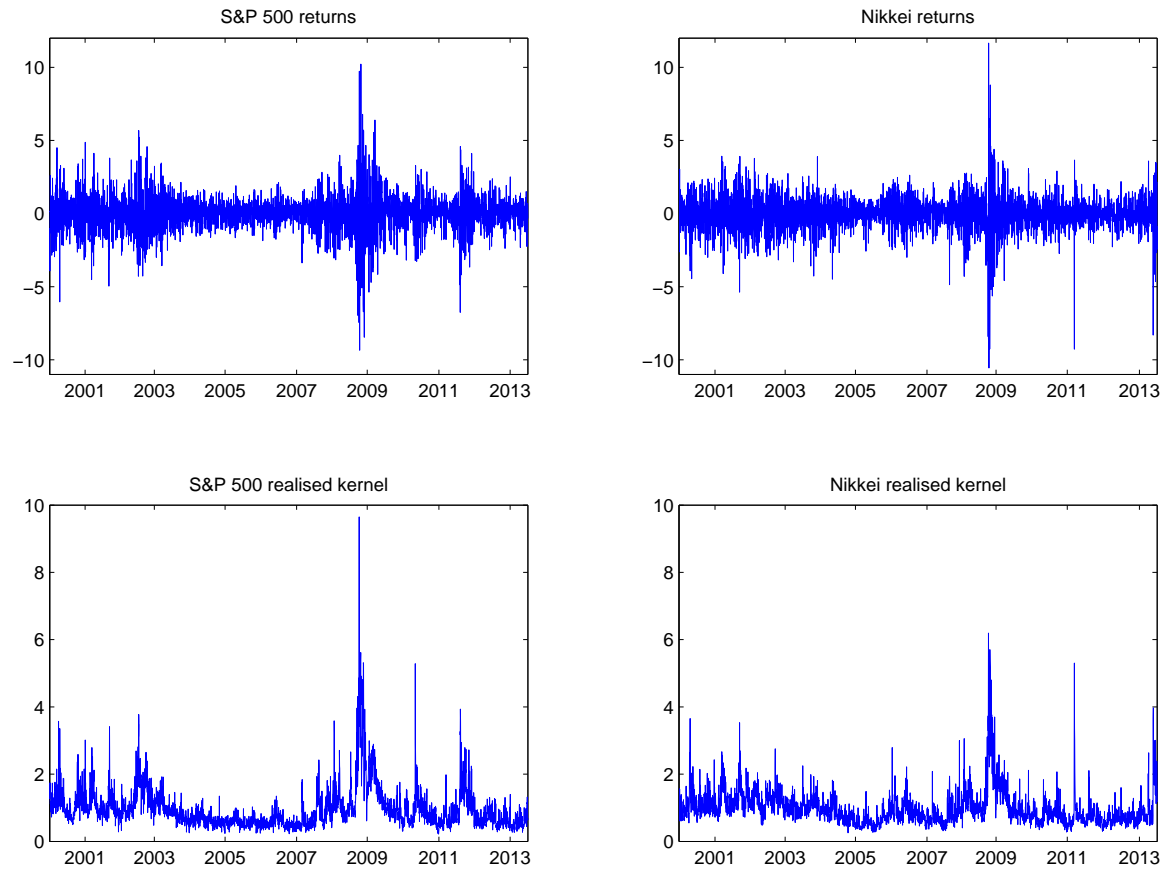
---

<sup>4</sup>See <http://realized.oxford-man.ox.ac.uk/>

<sup>5</sup>We have to delete 1-1.5% of the days on the S&P500, DJIA and FTSE index and 3% in case of the Nikkei index.

**Figure 1: Daily returns and realised measures**

This figure depicts the daily (close-to-close) returns on the S&P 500 index and the Nikkei index (upper part) and a realised kernel estimate of the corresponding daily (open-to-close) volatility (bottom part) from January 3, 2000, through June 27, 2013 (3,364 and 3,206 observations respectively). Both daily returns and volatilities are given in percentages.





number of observations in the region of interest becomes very low, such that the variation in the *csl* scores of the different models declines.<sup>6</sup> Similarly, there is a trade-off in the choice of  $T_w$ . On the one hand, one would choose  $T_w$  as high as possible in order to use the largest amount of available observations to compute the weights  $w_t$ . But on the other hand, if the relative performance of different models varies through time, one should take this into account and choose a smaller value of  $T_w$ . In addition,  $T_w$  and  $\kappa$  are related in the sense that a low value of  $\kappa$  combined with a small window results in a small amount of observations within the region  $B_t$ . Hence given these trade-offs and the relation between those two variables, we choose  $T_w$  and  $\kappa$  as 750 and 0.15 (0.25) such that there are 112 (187) observations in the left tail.

## 4.2 Evaluation

We assess the accuracy of our (combined) density forecasts in two ways. First, we focus purely on the predictive density in the left tail and investigate statistically whether pooling based on censored densities adds any value. Following Diks *et al.* (2011), we test the null hypothesis of equal performance of two density forecasts  $p_t(y_t; I_{t-1}, A_i)$  and  $p_t(y_t; I_{t-1}, A_j)$  based on the scoring rule of (4).<sup>7</sup> That is, given a sample of density forecasts and corresponding realizations for  $m$  periods, define the relative score  $d_t$  as

$$d_t = S^{csl}(y_t; A_i) - S^{csl}(y_t; A_j) \quad (18)$$

with corresponding null-hypothesis  $H_0 : \mathbb{E}[d_t] = 0$  for all  $m$  periods. The resulting Diebold and Mariano (1995) test-statistic is then given by

$$t_m = \frac{\bar{d}_m}{\sqrt{\hat{\sigma}_m^2/m}}, \quad (19)$$

with  $\bar{d}_m$  the sample average of the score differences and  $\hat{\sigma}_m^2$  a HAC-consistent variance estimator of the true variance  $\sigma_m^2$  of  $d_t$ . A positive value means that the density forecasts in

---

<sup>6</sup>Recall that if  $y_t$  is outside the region  $B_t$  with  $B_t$  the left tail  $y_t < \hat{r}_t^\kappa$ , the *csl* score is the cdf of  $y_t$  in the complement of the region.

<sup>7</sup>In case we consider density forecasts using combinations, the density forecast is given by  $\sum_{i=1}^n w_{it} p_t(y_t; I_{t-1}, A_i)$ .

the tail of model  $A_i$  are more accurate than the corresponding density forecasts of model  $A_j$ . This test allows for parameter estimation uncertainty and fits the framework of Giacomini and White (2006), who show that the use of a rolling window of  $m$  past observations for parameter estimation simplifies the asymptotic theory of tests of equal predictive accuracy. Moreover, the test allows to compare density forecasts of both nested and non-nested models.

The second way to explore the additional value of using censored densities in this study is based on 1- and 5-day Value-at-Risk (VaR) estimates. For the individual models considered in this study, the 1-day VaR estimate reads

$$VaR_t^{1-q} = \mu + z_q \sqrt{h_t}, \quad (20)$$

with  $\mu$  the estimated conditional mean return,  $h_t$  the (forecasted) conditional variance, and  $z_q$  represents the  $q$ -th quantile of the assumed cdf. However, we cannot apply (20) when our predictive distribution is a combination of individual distributions.<sup>8</sup> This also holds for the  $h$ -day ( $h \geq 2$ ) VaR estimates if the assumed distribution is non-Normal. We use simulation techniques to overcome this issue. That is, we simulate daily returns from each individual model/distribution according to the assigned weight (and conditional variance) to obtain the required quantile of the total distribution to compute the  $(1 - q)\%$  VaR.

Finally, we test the accuracy of the VaR estimates by focusing on two aspects. First, we assess the frequency of the VaR violations with the unconditional coverage (UC) of Kupiec (1995) and Christoffersen (1998). These tests compare the actual with the expected number of violations. In addition, we test whether the violations occur in clusters by means of the Independence test (Ind) of Christoffersen (1998). In order to apply both tests on the estimated 5-day VaRs, we create first 5 different sub-series to avoid any overlap. Thus, sub-series  $j$  contains the estimates  $\{VaR_j^{1-q}, VaR_{j+5}^{1-q}, VaR_{j+10}^{1-q}, \dots\}$  for  $j = 1, \dots, 5$ . According to the suggestion of Diebold *et al.* (1998), we use Bonferroni bounds for the 5 sub-series. That is, we assume that the VaR series has autocorrelation up to and including lag 4 (since we make 5-step ahead predictions of  $h_t$  on a daily basis), whereas each sub-series should have correct coverage and independent VaR violations. Hence we therefore backtest each sub-series separately with a size of  $\alpha/5$ , with  $\alpha$  the used significance level. Rejecting the

---

<sup>8</sup>The VaR of a mixture of densities is not equal to the weighted average of each individual VaR.

null hypothesis of unconditional coverage/independence occurs when the null is rejected for *any* of the 5 sub-series. Second, we compare the 1-day VaR estimates of two different methods/models using the following asymmetric linear (tick) loss function of order  $q$ , which is also used in the Conditional Predictive Ability (CPA) test of Giacomini and White (2006):

$$L_{A_i}^q(e_t) = (q - I[e_t < 0])e_t, \quad (21)$$

where  $q = 5\%$  and  $10\%$  and  $e_t = y_t - VaR_t^{1-q}$ . The loss function is asymmetric in the sense that if there occurs a violation (i.e.  $e_t < 0$ ) the negative number  $q - 1$  is multiplied by the magnitude of the violation  $e_t$ , resulting in a penalization of  $(1 - q) \times e_t$ . In contrast to this, if there is no violation, the loss is equal to  $q \times e_t$ , which is considerable lower.<sup>9</sup> Hence a model  $A_i$  is more penalized when a VaR violation is observed. The larger the magnitude of this violation, the larger the penalization. Similar to the density forecasts, we define the relative loss as

$$d_t^q = L_{A_i}^q(e_t) - L_{A_j}^q(e_t) \quad (22)$$

and consider again a Diebold and Mariano (1995) type statistic as given in (19). A negative value of the unconditional mean of  $d_t^q$  means that on average the VaR estimates of model  $A_i$  are better than the corresponding estimates of model  $A_j$ .

### 4.3 Results

In this subsection, we present both the statistical and economic results. In order to interpret these results, we first present the weights which are obtained by optimizing the log score function (3) and the *csl* score function as given in (5). Figure 2 shows the result of the iterative process of optimizing weights according to both score functions. The sub-graphs depict the dynamics of the weights using daily returns from the DJIA index according to the 18 models listed in Table 1.<sup>10</sup> The upper panel of the figure corresponds with the log score function, while the lower panel corresponds with the *csl* score function with  $\kappa = 0.25$ . The

---

<sup>9</sup>Suppose the 95% 1-day VaR of model A and B are equal to -5% and -8% respectively, while the actual return is -6%. The loss associated with model A is equal to  $(0.05 - 1)(-1) = 0.95$ , while the loss of model B is equal to  $(0.05 - 0)(-2) = 0.10$ .

<sup>10</sup>Figure B.1 in Appendix B provides weights corresponding to the other three stock market indexes.

upper panel shows that using the log score function results in a large weight for the Heavy model with Skewed- $t$  distributed errors until mid 2007. Subsequently, the weight of the Heavy Skewed- $t$  model declines to zero and there is room for the TGARCH and RGARCH models with skewed- $t$  errors. The years 2010-2012 combines almost four models, where the Heavy model with normal and Laplace distributed returns getting the most weights. Nevertheless, the Heavy Skewed- $t$  model appears again and dominates from 2012 onwards, with a subsequently minor role for the Threshold GARCH Laplace model.

A rather different dynamic pattern arises from the lower part of Figure 2, i.e. when the *csl* score function is optimized. Although the graph is similar in the sense that the Heavy Skewed- $t$  model dominates the other models during 2006 and since 2012, the years 2008-2012 show two main differences. First, the GARCH and RGARCH models with skewed- $t$  distributed errors have more impact in case of the *csl* score function during the period 2007-2009. For example, the GARCH skewed- $t$  model reaches a maximum weight of 0.49 at the start of 2007. Second, the HEAVY Lap model almost dominates from 2010 until mid 2012, while the upper figure attributes less weight to the Laplace distribution during 2011-2012.

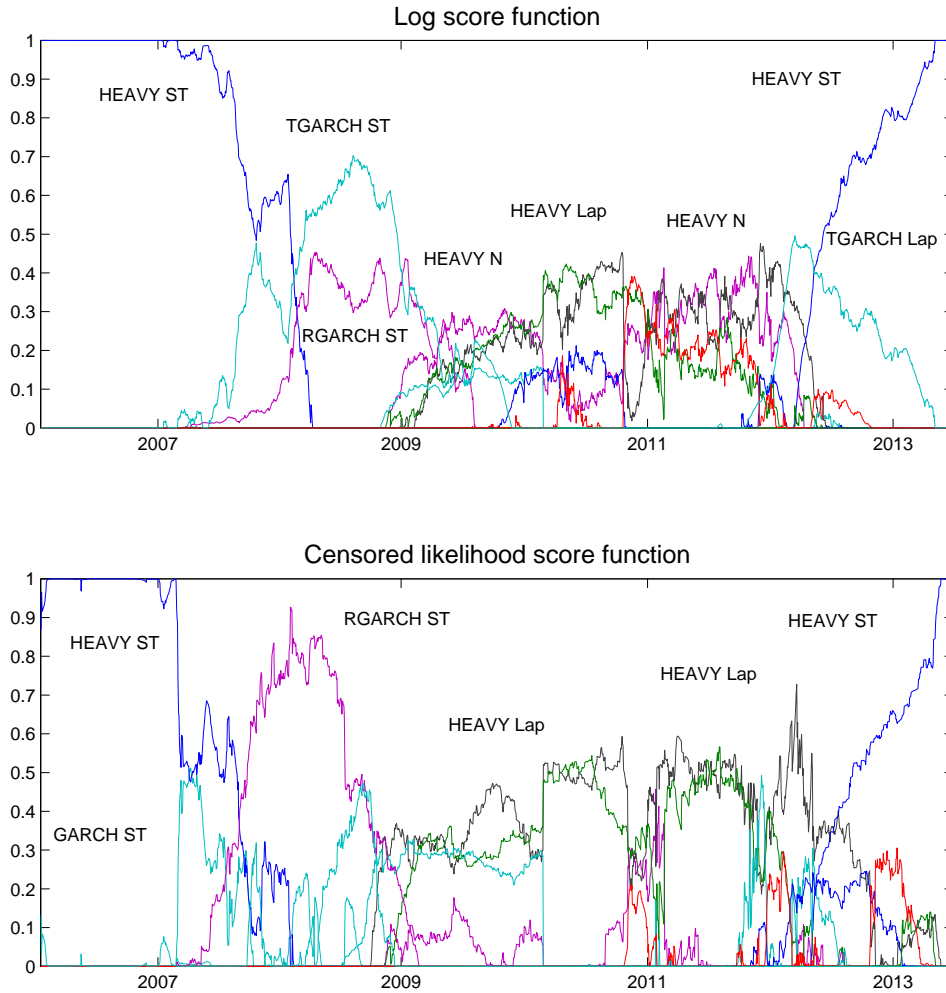
To ease the interpretation of this finding, Figure 3 sums up the weights according to each model class (upper part) and distribution (lower part), for both types of scoring rules. It seems that in case of the log score function, the Heavy model dominates the remaining models, with an important role for the Skewed- $t$  distribution. In contrast to this, focusing of the left tail of the distribution does lead to more influence of the RGARCH class of models. While Figure 2 indicates the most weight for the HEAVY class of models during 2010-2012, it seems that in aggregation the most weight goes to the RGARCH models. Furthermore, the Laplace distribution is more apparent during the years 2009-2012, with a climax at the start of 2012. Finally, both the Laplace and Normal distribution characterize 2012, while the log score function allocates the most weight to the Skewed- $t$  distribution in that year.

#### 4.3.1 Statistical results

Table 2 provides the importance of pooling of censored densities by showing results of the  $t$ -test on equal predictive accuracy, as provided in(19). We test equal accuracy of the combined density forecasts based on the *csl* score function and based on the log score function. In addition, we test the accuracy of the individual censored density of each

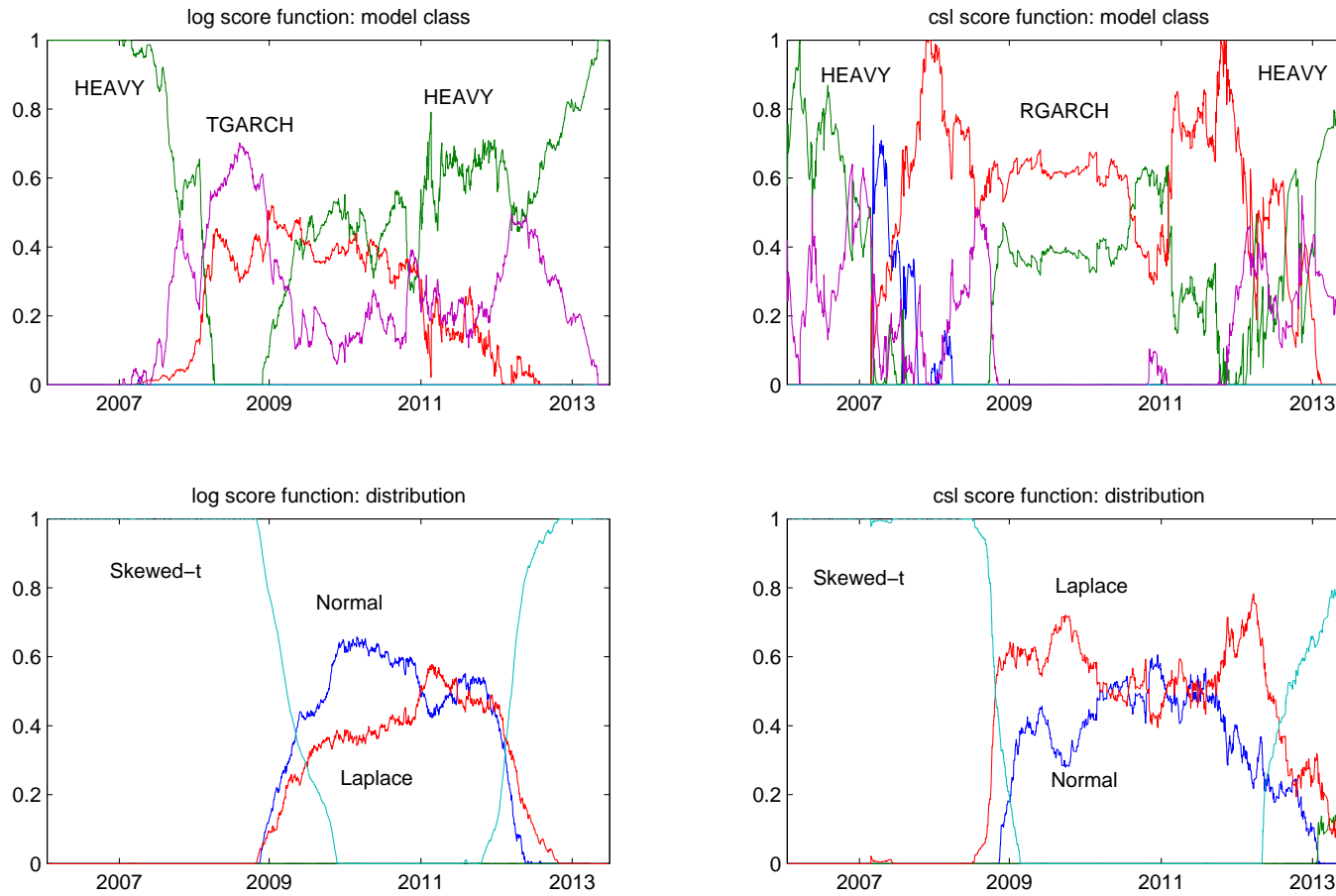
**Figure 2: Pooling weights DJIA index**

This figure depicts the evolution of weights based on optimizing the logarithmic score function (upper part) of (3) or the *csl* score function (bottom part) of (5) with a moving window of  $T = 750$  one-step ahead evaluated density forecasts using daily returns of the DJIA Index. In case of the *csl* score function,  $B_t$  represents the left tail  $y_t < \hat{r}^{0.25}$  with  $\hat{r}^{0.25}$  the 0.25th quantile of the empirical CDF of the moving estimation window of 750 returns. The labels refer to the models that have the highest weight at a given period. The abbreviations “ST”, “Lap” and “N” stand for Skewed-*t*, Laplace and Normal respectively.



**Figure 3: Pooling weights per model and distribution**

This figure sums up the optimized weights per model class (top panels) and distribution (bottom panels) based on optimizing the logarithmic score function (left part) of (3) or the *csl* score function of (5) (right part) with a moving window of  $T = 750$  one-step ahead evaluated density forecasts using daily returns of the DJIA Index. In case of the *csl* scoring function,  $B_t$  represents the left tail  $y_t < \hat{r}^{0.25}$  with  $\hat{r}^{0.25}$  the 0.25th quantile of the empirical CDF of the moving estimation window of 750 returns.



competing model. Panel A reports HAC-based  $t$ -statistics of the test of equal accuracy of density forecasts made by means of combination, using the *csl* or log score function of Section 2. As a benchmark, we consider also the case of equal weights assigned to each competing density. A positive number corresponds with more accurate density forecasts of the forecast method based on the *csl* score function. The table suggests that (except for the FTSE returns), combined forecasts based on the *csl* score function statistically outperform the density forecasts based on the log scoring rule, especially in case of one-step ahead forecasts. In addition, the combined density forecasts using equal weights are also improved, as indicated by the positive numbers. This improvement is statistically significant for the FTSE returns for both  $\kappa = 0.15$  and  $0.25$  (2.58, 3.07, 4.00 and 4.72) and the NIKKEI index when  $\kappa = 0.25$  (1.65). The final part of Panel A addresses the potential loss made when the accuracy is tested by means of the log score, while the weights are based on the *csl* score function. For the U.S. stock market indexes, the *csl* weights still outperform the weights based on the log score function, which in particular holds for the 5-step ahead forecasts. The opposite holds for the FTSE returns. All in all, there is a strong indication that (combined) density forecasts in the left tail significantly improves when using weights based on the *csl* score function compared to using weight based on the log score function. The improvement is less strong when compared to using equal weights. Finally, using only the left tail to forecast the whole density does not always imply less accurate density forecasts.

Panel B shows test results of using combined forecasts based on the left tail and the individual censored density forecasts. In general, using the *csl* score function results statistically in better density forecasts, as indicated by the positive numbers. This holds in particular when  $\kappa = 0.25$ .<sup>11</sup> Even if the null hypothesis cannot be rejected for a particular model, this result is not consistent for all data sets. For example, the Heavy Laplace model performs well in case of the U.S. stock market indexes, but is statistically beaten in case of the FTSE returns. In general, there is no striking difference between the results drawn from the 1-step and the 5-step ahead density forecasts. Interestingly, considering the S&P 500 and DJIA indexes in the upper part of Panel B, the RGARCH model with Normal or

---

<sup>11</sup>Table B.1 in Appendix B provides results where the weights are based on the log score function. The results indicate that the pooled density forecasts do not add any value in case of the S&P 500, DJIA and Nikkei indexes. Only in case of the FTSE index, there is evidence that the combined density forecasts statistically outperform the individual density forecasts.

**Table 2: Evaluation of 1- and 5-day ahead censored density forecasts**

This table reports results of testing equal predictive accuracy using the censored likelihood and log scoring rule of (4) and (1). In case of the *csl* score function,  $B_t$  represents the left tail  $y_t < \hat{r}^\kappa$  with  $\hat{r}^\kappa$  the  $\kappa$ th quantile of the empirical CDF of the in-sample returns. We set  $\kappa$  equal to 0.15 and 0.25 respectively. The weights are repeatedly optimized based on a moving window of 750 evaluated density forecasts. The test statistic is given in (19). Panel A compares combined density forecasts where the weights are based on the *csl* score function, log score function or each competing model gets the same weight. In Panel B, we test equal predictive accuracy of combined density forecasts and density forecasts of each competing model using the *csl* score function, where the weights of the combined density forecasting are also based on the *csl* score function. All models are estimated with a moving window of 750 daily returns from the S&P500, DJIA, FTSE and Nikkei index through the period January, 2000 - June, 2013. The test statistics are based on HAC-based standard errors and 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei) out-of-sample observations respectively.

Panel A: Comparison of combined forecasts								
	1-step ahead forecasts				5-step ahead forecasts			
	S&P500	DJIA	FTSE	Nikkei	S&P500	DJIA	FTSE	Nikkei
Loss function: <i>csl</i> score function								
	$\kappa = 0.15$							
csl vs log	3.46***	3.84***	0.31	2.81***	4.13***	4.25***	-0.65	2.08**
csl vs eqw	1.12	1.17	2.58***	1.03	0.67	1.04	3.07***	0.41
	$\kappa = 0.25$							
csl vs log	3.11***	3.66***	-0.71	2.17**	3.36***	3.62***	-0.41	1.61
csl vs eqw	1.33	1.54	4.00***	1.65*	0.43	0.79	4.72***	0.79
Loss function: log score function								
	$\kappa = 0.15$							
csl vs log	0.71	0.83	-2.84***	-0.20	3.31***	2.30**	-4.69***	-0.10
	$\kappa = 0.25$							
csl vs log	1.75*	2.55**	-3.05***	0.06	2.84***	2.11**	-3.21***	0.63
Panel B: Pooled ( <i>csl</i> score function) vs. individual								
	$\kappa = 0.15$							
GARCH N	3.23***	3.25***	4.25***	2.26**	2.28**	2.36**	3.83***	2.00**
GARCH T	3.26***	3.04***	4.36***	2.30**	1.58	1.55	4.47***	1.50
GARCH Lap	1.52	1.53	4.59***	2.72***	0.17	0.41	4.39***	1.14
GARCH ST	6.03***	5.77***	3.09***	3.72***	6.23***	5.95***	2.64***	3.41***
TGARCH N	2.15**	2.39**	4.48***	1.79*	1.89*	2.17**	4.14***	1.90*
TGARCH T	1.97**	2.09**	4.30***	2.67***	0.93	1.41	4.46***	1.71*
TGARCH Lap	0.63	0.60	4.38***	2.37**	-0.56	-0.14	4.18***	0.52
TGARCH ST	5.59***	5.36***	2.26**	3.81***	5.73***	5.64***	1.61	3.55***
HEAVY N	1.39	1.48	3.25***	1.34	1.54	1.46	3.63***	1.57
HEAVY T	0.85	1.14	3.08***	0.29	0.25	0.52	3.92***	1.23
HEAVY Lap	-0.30	-0.13	3.30***	0.26	-1.00	-0.72	3.48***	-0.02
HEAVY ST	5.33***	4.96***	0.22	3.31***	5.56***	4.98***	0.21	2.89***
RGARCH N	1.42	1.40	3.60***	2.20**	3.06***	2.89***	5.70***	3.31***
RGARCH T	1.11	1.02	3.68***	1.40	2.65***	3.06***	5.62***	2.85***
RGARCH Lap	0.12	0.19	3.79***	3.11***	1.50	1.87*	4.88***	3.57***
RGARCH ST	5.19***	4.91***	-1.29	5.35***	6.67***	6.69***	1.13	6.00***
GAS T	3.05***	2.77***	4.36***	2.09**	1.75*	1.31	4.53***	1.75*
GAS Lap	1.47	1.60	4.56***	2.58***	0.18	0.40	4.42***	1.07



(continued from previous page)

	1-step ahead forecasts				5-step ahead forecasts			
	S&P500	DJIA	FTSE	Nikkei	S&P500	DJIA	FTSE	Nikkei
	$\kappa = 0.25$							
GARCH N	3.74***	3.86***	5.83***	2.53**	2.48**	2.51**	4.86***	2.09**
GARCH T	3.71***	3.66***	5.94***	3.11***	1.51	1.53	5.91***	2.01**
GARCH Lap	1.90*	2.08**	5.58***	3.37***	0.04	0.34	5.50***	1.65*
GARCH ST	6.02***	5.62***	2.63***	3.25***	5.64***	5.38***	2.56**	3.08***
TGARCH N	2.93***	3.24***	6.26***	1.99**	2.28**	2.47**	5.28***	1.99**
TGARCH T	2.80***	3.08***	6.00***	3.36***	1.27	1.69*	5.91***	2.10**
TGARCH Lap	1.18	1.36	5.40***	3.07***	-0.50	-0.05	5.34***	1.08
TGARCH ST	5.31***	4.92***	1.90*	2.91***	4.97***	4.80***	1.66*	2.95***
HEAVY N	2.14**	2.35**	5.21***	1.52	2.06**	2.02**	5.11***	1.65*
HEAVY T	1.66*	2.34**	4.92***	1.28	0.83	1.26	5.70***	1.53
HEAVY Lap	0.20	0.75	4.45***	1.41	-0.86	-0.34	4.87***	0.81
HEAVY ST	5.04***	4.54***	-1.07	2.33**	5.25***	4.56***	-0.08	2.49**
RGARCH N	2.57**	2.67***	5.67***	2.52**	3.97***	3.55***	7.18***	3.55***
RGARCH T	2.26**	2.23**	5.53***	2.40**	3.46***	3.40***	7.32***	3.45***
RGARCH Lap	0.65	0.86	4.86***	3.81***	1.30	1.67*	6.12***	4.11***
RGARCH ST	5.60***	5.44***	-1.67*	7.41***	8.37***	8.13***	1.98**	8.39***
GAS T	3.45***	3.41***	5.92***	2.85***	1.70*	1.35	5.99***	2.19**
GAS Lap	1.82*	2.11**	5.54***	3.34***	0.01	0.28	5.55***	1.69*

Student- $t$  distributed errors produces accurate 1-step ahead density forecasts, while forecasting 5 steps ahead results in inaccurate forecasts compared to pooled density forecasts with weights based on the 15th quantile of the in-sample return distribution.

We take a closer look to the statistical results in Table 3, which lists results of equal predictive accuracy using the  $csl$  score function, but now across different model classes and distributions. All subsets are tested against the combined density forecasts using all 18 models. A positive number again means relatively better density forecasts by the combined method. Considering model classes,  $t$ -statistics corresponding with the model classes including realised measures produce accurate 1-step ahead density forecasts. This effect is stronger is somewhat stronger when  $\kappa = 0.15$ . The RGARCH model performs worse than the HEAVY model - especially when 5-step ahead forecasts are made - which could be due to more parameter uncertainty since 8 parameters has to be estimated, while the HEAVY models requires only 3 or 6 parameters.

Turning to the performance across various distributions, the Laplace distribution produces statistically the best forecasts, as suggested by the minus sign of the  $t$ -statistic. However, also the Normal and Student- $t$  distribution does a good job when  $\kappa = 0.15$ , since the combined method is not able to consistently beat these distribution. However, when

the 0.25th quantile is used to generate 1-step ahead density forecasts,  $t$ -statistics of 2.11, 1.91, 2.35 and 2.19 indicate that the Normal and Student- $t$  distribution are significantly outperformed. Only the Skewed- $t$  distribution is statistically outperformed by using all combinations of models and distributions for both values of  $\kappa$ , except for the FTSE returns. This is consistent with the figures shown earlier, as the skewed- $t$  disappears during the years 2008-2011.

Table 4 reports additional evidence of the added value of pooling using the  $csl$  score function, by providing the  $csl$  score over the out-of-sample period:

$$\sum_{t=1}^T \log \left[ \sum_{i=1}^n w_{i,t-1}^* \left( I[y_t \in B_t] \log p_t(y_t; I_{t-1}, A_i) + I[y_t \in B_t^c] \int_{B_t^c} p_t(y; I_{t-1}, A_i) dy \right) \right], \quad (23)$$

where  $w_{i,t-1}^*$  is the optimized weight for model  $A_i$  at the end of trading day  $t - 1$ , based on the evaluated density forecasts at time  $t - T_w$  through  $t - 1$ . In addition, we provide corresponding values of the individual models with bold numbers the  $csl$  scores which are exceeding the pooled  $csl$  score. The low amount of bold numbers indicates that the pooled  $csl$  scores are higher than the  $csl$  scores of most of the individual models. If this is not the case, the differences are small, with a maximum difference of 9 points (S&P 500, 5-step ahead forecasts with  $\kappa = 0.25$ ). Further, the  $csl$  score of our pooling method are higher than the scores of the remaining 17 models (16 in case of the U.S. stock market indexes), with differences that can be quite substantial. For example, if one favours the best performing individual model, i.e. the Heavy Laplace model, this results in a loss of 41 or 42 points with respect to the pooled  $csl$  score based with  $\kappa = 0.15$  in case of the FTSE data set. Finally, there is quite some positive difference between the  $csl$  scores of pooling with weights based on the  $csl$  score function and simply using equal weights. For the U.S. and Japanese indexes, the difference is on average around 8 points, however in case of the FTSE index, the difference increases to 20 ( $\kappa = 0.15$ ) or 45 ( $\kappa = 0.25$ ) points respectively. Note that the table relates to Table 2, in the sense that a negative  $t$ -stat of a particular model corresponds with a higher  $csl$  score of that model than the pooled  $csl$  score. We refer to Table B.2 in Appendix B for similar type of results regarding the log scores.

**Figure 4: Censored likelihood scores w.r.t. individual models**

This figure depicts the cumulative sum of the difference of the censored likelihood score corresponding with one-step ahead density forecasts of the pooled densities and the *csl* score of the three competing individual models according to Table 4. The weights of the pooled densities are based on maximizing the *csl* score function of (5) with a moving window of 750 evaluated density forecasts, using daily returns of the DJIA index. Further,  $B_t$  the left tail  $y_t < \hat{r}^{0.25}$  with  $\hat{r}^{0.25}$  the  $\kappa$ th quantile of the empirical CDF of the in-sample returns.

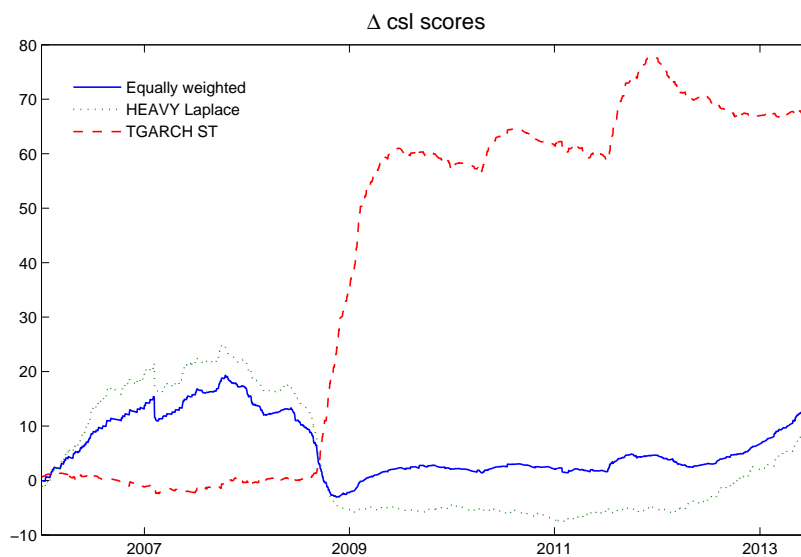


Figure 4 illustrates the evolution of the cumulative gain in the *csl* scores of Table 4 through time. In particular, it shows the cumulative difference of the *csl* scores corresponding with the combined density forecasts relative to the *csl* scores of the TGARCH Skewed-*t* model, the Heavy Laplace model and the combined density forecasts based on equal weights. The figure shows two different patterns. First, the gain of pooling with respect to the HEAVY Laplace model and the equal weighted combination method occurs mainly during the first years, decreases during the crisis period and increases slowly from 2009 onwards. Second, pooling does not add much value with respect to the Skewed-*t* distribution during the first years. However, the gain becomes striking at the end of 2008 and in 2011. This result is related with Figure 3, as the Skewed-*t* distribution dominates all the other distributions until the end of 2009.

#### 4.3.2 Economic results

Tables 5 and 6 shed light on the economic impact of pooling (censored) density forecasts in the context of VaR estimates. For each data set, we first compare the frequency and independence of the VaR violations corresponding with the combined densities based on pooling, either using the *csl* or log scoring rule or using equally weights. The latter can be seen as a benchmark. We report results of two approaches based on equal weights. The first approach takes into account that the VaR of a mixture of densities is not equal to the weighted VaR of the individual densities, while the second approach ignores this fact and simply takes the average of all individual VaR's, as also done in Giacomini and Komunjer (2005). We label these approaches as eqw(1) and eqw(2) respectively. Second, we show results of each individual model per data set. Furthermore, we compare the accuracy of the 1-day VaR estimates based on the *csl* scoring rule with VaR estimates from any other pooling method or from any individual model by applying a *t*-test based on the asymmetric tick loss function of (22). A negative number indicates that the pooled *csl* based VaR estimates are more accurate. Apart from this test, both tables contain the same type of results, although Table 5 focuses on the 1-day VaR estimates, while Table 6 provides results of the 5-day VaR estimates.

Three main conclusions are apparent from Table 5. First, the VaR estimates corresponding with our new proposed technique outperform the benchmark of using equal weights, both

**Table 3: Predictive accuracy per model class and distribution**

This table reports results of testing equal predictive accuracy using the censored likelihood scoring rule of (4), with  $B_t$  represents the left tail  $y_t < \hat{y}^\kappa$  with  $\hat{y}^\kappa$  the  $\kappa$ th quantile of the empirical CDF of the in-sample returns. We set  $\kappa$  equal to 0.15 and 0.25 respectively. The weights are repeatedly optimized based on a moving window of 750 evaluated density forecasts. The test statistic is given in (19). The table tests combined density forecasts based on each model class/distribution against density forecasts based on all models and distributions, where the weights of the each combined density forecast are also based on the *csf* score function. All models are estimated with a moving window of 750 daily returns from the S&P500, DJIA, FTSE and Nikkei index through the period January, 2000 - June, 2013. The test statistics are based on HAC-based standard errors and 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei) out-of-sample observations respectively.

	1-step ahead forecasts				5-step ahead forecasts			
	S&P500	DJIA	FTSE	Nikkei	S&P500	DJIA	FTSE	Nikkei
$\kappa = 0.15$								
GARCH	3.05***	2.89***	3.66***	2.71***	1.33	1.52	3.35***	1.06
TGARCH	2.11**	1.82*	3.17***	2.66***	-0.27	0.43	2.36**	0.24
HEAVY	0.19	0.09	1.26	-2.57**	-1.49	-0.71	0.70	-0.48
RGARCH	0.12	0.25	-0.08	1.56	2.33**	2.75***	1.68*	3.01***
GAS	2.78***	2.53**	4.21***	2.24**	1.25	1.00	4.35***	1.15
Normal	1.03	1.18	3.65***	1.40	1.63	1.71*	4.11***	1.93*
Student t	0.90	0.91	3.61***	-0.28	0.67	0.73	4.41***	0.52
Laplace	-0.22	-0.01	3.54***	0.22	-0.73	-0.46	3.65***	0.21
Skewed t	5.20***	4.71***	-1.49	3.62***	5.93***	5.51***	-0.01	3.32***
$\kappa = 0.25$								
GARCH	3.86***	3.30***	4.19***	2.92***	0.62	1.34	4.12***	1.06
TGARCH	2.98***	2.21**	3.39***	2.52**	-1.02	0.10	3.22***	0.18
HEAVY	-0.53	-0.72	2.02**	-3.62***	-2.24**	-0.92	1.25	-0.23
RGARCH	2.02**	2.15**	0.55	2.57**	2.35**	2.84***	2.61***	3.77***
GAS	3.20***	2.89***	5.52***	2.89***	1.39	0.83	5.55***	1.58
Normal	2.11**	2.35**	5.54***	1.64	2.13**	2.24**	5.63***	2.10**
Student t	1.91*	2.19**	5.36***	0.94	1.17	1.30	6.04***	1.08
Laplace	0.38	0.84	4.70***	1.40	-0.74	-0.13	5.02***	0.94
Skewed t	4.94***	4.50***	-2.25**	3.05***	5.80***	5.21***	0.15	3.18***

**Table 4: Censored likelihood scores**

This table reports censored likelihood scores corresponding with individual models and combined models, where the weights are based on optimizing the *csl* score function of (5), with  $B_t$  the left tail  $y_t < \hat{r}^\kappa$  with  $\hat{r}^\kappa$  the  $\kappa$ th quantile of the empirical CDF of the in-sample returns. We set  $\kappa$  equal to 0.15 and 0.25. The weights are repeatedly optimized based on a moving window of 750 evaluated density forecasts. In addition, *csl* scores are reported of combined models using equal weights (eqw). The bold numbers represent the scores exceeding the pooled *csl* score. All models are estimated with a moving window of 750 daily returns from the S&P500, DJIA, FTSE and Nikkei index through the period January, 2000 - June, 2013. The number of out-of-sample observations are equal to 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei)respectively.

	1-step ahead forecasts				5-step ahead forecasts			
	S&P500	DJIA	FTSE	Nikkei	S&P500	DJIA	FTSE	Nikkei
$\kappa = 0.15$								
GARCH N	-1052	-1014	-1019	-1002	-1081	-1031	-1050	-1083
GARCH T	-1030	-991	-1002	-937	-1038	-993	-1013	-956
GARCH Lap	-1019	-980	-1001	-934	-1029	-984	-1009	-952
GARCH ST	-1091	-1043	-964	-961	-1105	-1049	-973	-986
TGARCH N	-1038	-1002	-1016	-1013	-1066	-1024	-1049	-1067
TGARCH T	-1022	-984	-1001	-938	-1035	-994	-1016	-957
TGARCH Lap	-1012	-972	-999	-933	-1023	<b>-981</b>	-1008	-948
TGARCH ST	-1082	-1036	-959	-960	-1099	-1046	-967	-985
HEAVY N	-1026	-990	-988	-979	-1055	-1005	-1027	-1061
HEAVY T	-1013	-976	-983	-919	-1029	-986	-1006	-962
HEAVY Lap	<b>-1004</b>	<b>-966</b>	-986	-919	<b>-1019</b>	<b>-977</b>	-1000	<b>-944</b>
HEAVY ST	-1075	-1028	-946	-950	-1092	-1038	-959	-994
RGARCH N	-1022	-981	-990	-1011	-1080	-1022	-1050	-1124
RGARCH T	-1014	-974	-989	-931	-1049	-1007	-1022	-992
RGARCH Lap	-1008	-969	-991	-951	-1040	-996	-1016	-1000
RGARCH ST	-1075	-1028	<b>-940</b>	-986	-1109	-1059	-964	-1057
GAS T	-1031	-991	-1005	-941	-1043	-993	-1015	-965
GAS Lap	-1020	-981	-1001	-935	-1029	-984	-1010	-951
pooled csl	-1007	-967	-945	-917	-1027	-982	-958	-944
eqw	-1013	-974	-969	-923	-1031	-986	-983	-947
$\kappa = 0.25$								
GARCH N	-1480	-1443	-1364	-1338	-1507	-1458	-1398	-1421
GARCH T	-1454	-1419	-1343	-1274	-1461	-1417	-1358	-1295
GARCH Lap	-1440	-1404	-1336	-1273	-1447	-1405	-1348	-1291
GARCH ST	-1505	-1459	-1270	-1281	-1521	-1468	-1286	-1312
TGARCH N	-1468	-1433	-1363	-1349	-1495	-1454	-1399	-1405
TGARCH T	-1448	-1414	-1343	-1275	-1459	-1420	-1361	-1295
TGARCH Lap	-1433	-1397	-1334	-1272	<b>-1441</b>	-1401	-1347	-1287
TGARCH ST	-1496	-1449	-1265	-1276	-1513	-1463	-1280	-1309
HEAVY N	-1453	-1421	-1332	-1313	-1486	-1438	-1376	-1398
HEAVY T	-1436	-1405	-1324	-1253	-1454	-1415	-1352	-1300
HEAVY Lap	-1422	-1389	-1320	-1256	<b>-1437</b>	<b>-1398</b>	-1338	-1283
HEAVY ST	-1485	-1439	<b>-1244</b>	-1265	-1511	-1460	<b>-1266</b>	-1318
RGARCH N	-1454	-1414	-1336	-1349	-1517	-1458	-1400	-1464
RGARCH T	-1440	-1402	-1330	-1270	-1479	-1437	-1370	-1332
RGARCH Lap	-1427	-1391	-1325	-1288	-1460	-1418	-1354	-1338
RGARCH ST	-1494	-1450	<b>-1241</b>	-1342	-1547	-1499	-1280	-1428
GAS T	-1455	-1420	-1346	-1278	-1465	-1416	-1360	-1303
GAS Lap	-1440	-1405	-1336	-1274	-1446	-1405	-1349	-1291
pooled csl	-1420	-1381	-1250	-1243	-1446	-1402	-1267	-1276
eqw	-1430	-1394	-1295	-1254	-1450	-1407	-1316	-1282

**Table 5: Evaluation of 1-day Value-at-Risk estimates**

This table provides the accuracy of 1-day VaR estimates of the daily return of the S&P 500, DJIA, FTSE and Nikkei index, obtained by combining density forecasts with weights based on the *csl* scoring rule of (4), the log scoring rule of (1) or equal weights. In case of using equal weights, we report the approach by means of simulation (eqw(1)) and the approach that takes simply the average of all individual VaR estimates (eqw(2)). Further, we report results based on VaR estimates of the individual models. The columns represent for both 95% and 90% VaRs the number of violations, the percentage of violations with respect to the total number of VaR estimates in parentheses, the *p*-values of the Unconditional Coverage (UC) of Christoffersen (1998) and finally HAC-based *t*-statistics of the unconditional test on predictive ability of the combination method/individual model and the combined density forecasts with weights based on the *csl* score function, using the tick-loss function presented in (22). Bold numbers represent those models which have a *p*-value for the UC test above 5% for both the 90% and 95% VaR estimates. The number of estimated VaRs for each series is equal to 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei) respectively.

S&P500						
Model/Sc. rule	V(%)	$p_{uc}$	<i>t</i> -stat	V(%)	$p_{uc}$	<i>t</i> -stat
	95% VaR			90% VaR		
csl	110 (5.91)	0.078		198 (10.65)	0.358	
log	110 (5.91)	0.078	-1.18	200 (10.75)	0.284	-0.07
eqw(1)	111 (5.97)	0.063	-1.78*	199 (10.70)	0.320	-0.32
eqw(2)	114 (6.13)	0.031	-1.96**	207 (11.13)	0.110	-0.70
GARCH N	124 (6.67)	0.002	-3.33***	192 (10.32)	0.644	-2.18**
GARCH T	132 (7.10)	0.000	-3.40***	230 (12.37)	0.001	-2.76***
GARCH Lap	121 (6.51)	0.004	-3.17***	231 (12.42)	0.001	-2.70***
GARCH ST	122 (6.56)	0.003	-3.32***	211 (11.34)	0.058	-2.40**
TGARCH N	120 (6.45)	0.006	-2.63***	192 (10.32)	0.644	-0.90
TGARCH T	130 (6.99)	0.000	-2.73***	216 (11.61)	0.023	-1.43
TGARCH Lap	116 (6.24)	0.018	-2.51**	224 (12.04)	0.004	-1.45
TGARCH ST	108 (5.81)	<b>0.119</b>	-2.54**	191 (10.27)	<b>0.700</b>	-0.89
HEAVY N	125 (6.72)	0.001	-1.20	191 (10.27)	0.700	2.18**
HEAVY T	133 (7.15)	0.000	-1.25	215 (11.56)	0.028	0.32
HEAVY Lap	115 (6.18)	0.024	0.15	222 (11.94)	0.007	0.71
HEAVY ST	110 (5.91)	<b>0.078</b>	0.47	198 (10.65)	<b>0.358</b>	1.70*
RGARCH N	115 (6.18)	0.024	-1.35	182 (9.78)	0.756	0.07
RGARCH T	123 (6.61)	0.002	-1.96*	209 (11.24)	0.081	-1.51
RGARCH Lap	110 (5.91)	0.078	-1.51	218 (11.72)	0.016	-1.61
RGARCH ST	106 (5.70)	<b>0.176</b>	-0.77	190 (10.22)	<b>0.758</b>	-0.43
GAS T	129 (6.94)	0.000	-3.21***	224 (12.04)	0.004	-2.46**
GAS Lap	121 (6.51)	0.004	-2.94***	227 (12.20)	0.002	-2.50**
DJIA						
csl	110 (5.91)	0.080		201 (10.79)	0.258	
log	115 (6.18)	0.024	0.54	201 (10.79)	0.258	-1.09
eqw(1)	114 (6.12)	0.032	0.06	206 (11.06)	0.132	-0.73
eqw(2)	119 (6.39)	0.008	-0.41	207 (11.12)	0.114	-0.75
GARCH N	130 (6.98)	0.000	-2.85***	198 (10.63)	0.366	-1.87*
GARCH T	138 (7.41)	0.000	-2.74***	222 (11.92)	0.007	-2.37**
GARCH Lap	117 (6.28)	0.014	-2.26**	222 (11.92)	0.007	-2.30**
GARCH ST	125 (6.71)	0.001	-2.35**	213 (11.44)	0.042	-2.06**
TGARCH N	124 (6.66)	0.002	-1.83*	199 (10.69)	0.328	-1.14
TGARCH T	132 (7.09)	0.000	-1.84*	222 (11.92)	0.007	-1.69*
TGARCH Lap	108 (5.80)	0.122	-1.14	227 (12.19)	0.002	-1.50
TGARCH ST	108 (5.80)	<b>0.122</b>	-1.20	207 (11.12)	<b>0.114</b>	-1.08
HEAVY N	130 (6.98)	0.000	0.62	201 (10.79)	0.258	1.42
HEAVY T	138 (7.41)	0.000	0.07	214 (11.49)	0.035	-0.18
HEAVY Lap	108 (5.80)	0.122	1.53	217 (11.65)	0.020	0.47
HEAVY ST	111 (5.96)	0.064	1.57	203 (10.90)	0.200	0.87
RGARCH N	118 (6.34)	0.011	-0.31	189 (10.15)	0.829	0.78
RGARCH T	126 (6.77)	0.001	-1.53	203 (10.90)	0.200	-1.21
RGARCH Lap	106 (5.69)	<b>0.179</b>	-0.24	202 (10.85)	<b>0.228</b>	-0.89
RGARCH ST	104 (5.59)	<b>0.255</b>	-0.00	191 (10.26)	<b>0.712</b>	0.01
GAS T	136 (7.30)	0.000	-2.58***	213 (11.44)	0.042	-1.90*
GAS Lap	114 (6.12)	0.032	-2.32**	220 (11.82)	0.011	-2.02**

(continued from previous page)

FTSE						
Sc. rule/Model	V(%)	$p_{uc}$	$t$ -stat	V(%)	$p_{uc}$	$t$ -stat
		95% VaR			90% VaR	
csl	122 (6.50)	0.004		208 (11.08)	0.126	
log	112 (5.96)	0.063	0.92	209 (11.13)	0.109	1.25
eqw(1)	115 (6.12)	0.031	-0.77	206 (10.97)	0.167	-0.25
eqw(2)	116 (6.18)	0.024	-0.87	209 (11.13)	0.109	-0.42
GARCH N	122 (6.50)	0.004	-2.18**	203 (10.81)	0.248	-1.30
GARCH T	129 (6.87)	0.000	-2.29**	222 (11.82)	0.010	-1.73*
GARCH Lap	108 (5.75)	0.144	-1.90*	224 (11.93)	0.007	-1.87*
GARCH ST	112 (5.96)	0.063	-2.00**	212 (11.29)	0.068	-1.28
TGARCH N	120 (6.39)	0.008	-1.97**	204 (10.86)	0.218	-1.58
TGARCH T	130 (6.92)	0.000	-1.98**	223 (11.87)	0.008	-1.97**
TGARCH Lap	103 (5.48)	0.342	-1.45	229 (12.19)	0.002	-2.09**
TGARCH ST	103 (5.48)	<b>0.342</b>	-1.45	204 (10.86)	<b>0.218</b>	-1.46
HEAVY N	124 (6.60)	0.002	0.61	202 (10.76)	0.280	1.96**
HEAVY T	135 (7.19)	0.000	-0.10	219 (11.66)	0.019	1.14
HEAVY Lap	106 (5.64)	0.209	1.26	224 (11.93)	0.007	0.84
HEAVY ST	108 (5.75)	<b>0.144</b>	1.23	200 (10.65)	<b>0.353</b>	1.65*
RGARCH N	128 (6.82)	0.001	-1.58	203 (10.81)	0.248	0.42
RGARCH T	138 (7.35)	0.000	-2.32**	221 (11.77)	0.013	-1.74*
RGARCH Lap	115 (6.12)	0.031	-0.49	226 (12.03)	0.004	-1.82*
RGARCH ST	116 (6.18)	0.024	-0.58	203 (10.81)	0.248	-0.39
GAS T	126 (6.71)	0.001	-2.40**	217 (11.55)	0.028	-1.72*
GAS Lap	98 (5.22)	0.666	-2.14**	224 (11.93)	0.007	-1.71*
Nikkei						
csl	87 (4.94)	0.904		171 (9.70)	0.678	
log	87 (4.94)	0.904	-0.66	170 (9.65)	0.621	-1.49
eqw(1)	93 (5.28)	0.595	-1.58	166 (9.42)	0.414	-2.33**
eqw(2)	90 (5.11)	0.836	-2.05**	163 (9.25)	0.289	-2.75***
GARCH N	104 (5.90)	<b>0.091</b>	-1.34	166 (9.42)	<b>0.414</b>	-2.44**
GARCH T	108 (6.13)	0.035	-1.55	189 (10.73)	0.314	-2.24**
GARCH Lap	90 (5.11)	<b>0.836</b>	-1.19	182 (10.33)	<b>0.647</b>	-2.40**
GARCH ST	99 (5.62)	<b>0.242</b>	-1.13	178 (10.10)	<b>0.887</b>	-2.20**
TGARCH N	99 (5.62)	<b>0.242</b>	-1.40	161 (9.14)	<b>0.221</b>	-2.14**
TGARCH T	106 (6.02)	<b>0.058</b>	-1.91*	178 (10.10)	<b>0.887</b>	-2.13**
TGARCH Lap	86 (4.88)	<b>0.818</b>	-1.03	176 (9.99)	<b>0.987</b>	-1.93*
TGARCH ST	95 (5.39)	<b>0.456</b>	-1.53	166 (9.42)	<b>0.414</b>	-2.12**
HEAVY N	91 (5.16)	<b>0.752</b>	0.82	162 (9.19)	<b>0.254</b>	0.62
HEAVY T	99 (5.62)	<b>0.242</b>	0.45	176 (9.99)	<b>0.987</b>	1.43
HEAVY Lap	75 (4.26)	<b>0.142</b>	0.87	176 (9.99)	<b>0.987</b>	0.76
HEAVY ST	86 (4.88)	<b>0.818</b>	1.72*	170 (9.65)	<b>0.621</b>	1.42
RGARCH N	84 (4.77)	0.652	-1.90*	144 (8.17)	0.008	-2.82***
RGARCH T	87 (4.94)	<b>0.904</b>	-2.14**	157 (8.91)	<b>0.121</b>	-2.23**
RGARCH Lap	69 (3.92)	0.030	-2.76***	149 (8.46)	0.027	-3.17***
RGARCH ST	76 (4.31)	0.176	-2.71***	152 (8.63)	0.050	-3.20***
GAS T	106 (6.02)	<b>0.058</b>	-1.84*	191 (10.84)	<b>0.246</b>	-2.12**
GAS Lap	86 (4.88)	<b>0.818</b>	-2.07**	184 (10.44)	<b>0.538</b>	-2.80***



**Table 6: Evaluation of 5-day Value-at-Risk estimates**

This table provides the accuracy of 5-day VaR estimates of the daily return of the S&P 500, DJIA, FTSE and Nikkei index, obtained by combining density forecasts with weights based on the *csl* scoring rule of (4), the log scoring rule of (1) or equal weights. In case of using equal weights, we report the approach by means of simulation (eqw(1)) and the approach that takes simply the average of all individual VaR estimates (eqw(2)). Further, we report results based on VaR estimates of the individual models. For each combination method/model, we have 5 different sub-series of VaRs. The table reports for both 95% and 90% VaRs the sub-series corresponding with the lowest  $p$ -value of the test on unconditional coverage of the VaR estimates. The columns represents the corresponding number of violations, the percentage of violations with respect to the total number of VaR estimates in parentheses and the  $p$ -values of the Unconditional Coverage (UC) and Independence (Ind) test of Christoffersen (1998). The number of estimated VaRs for each series is equal to 372 (S&P 500), 372 (DJIA), 375 (FTSE) and 352 (Nikkei) respectively.

S&P500						
Model/sc. rule	V(%)	$p_{uc}$	$p_{m,ind}$	V(%)	$p_{uc}$	$p_{ind}$
	95% VaR			90% VaR		
csl	24 (6.45)	0.218	0.263	32 (8.63)	0.367	0.880
log	23 (6.18)	0.312	0.626	43 (11.56)	0.327	0.993
eqw(1)	26 (6.99)	0.096	0.511	46 (12.37)	0.141	0.396
eqw(2)	24 (6.45)	0.218	0.652	44 (11.83)	0.252	0.394
GARCH N	26 (6.99)	0.096	0.479	45 (12.10)	0.190	0.461
GARCH T	28 (7.53)	0.037	0.390	56 (15.05)	0.002	0.302
GARCH Lap	25 (6.72)	0.147	0.544	50 (13.44)	0.035	0.196
GARCH ST	25 (6.72)	0.147	0.544	48 (12.90)	0.073	0.567
TGARCH N	23 (6.18)	0.312	0.626	32 (8.63)	0.367	0.598
TGARCH T	27 (7.26)	0.060	0.979	46 (12.37)	0.141	0.164
TGARCH Lap	24 (6.45)	0.218	0.712	47 (12.63)	0.102	0.134
TGARCH ST	14 (3.76)	0.253	0.313	31 (8.36)	0.279	0.675
HEAVY N	27 (7.26)	0.060	0.455	43 (11.56)	0.327	0.993
HEAVY T	28 (7.53)	0.037	0.531	45 (12.10)	0.190	0.821
HEAVY Lap	26 (6.99)	0.096	0.889	47 (12.63)	0.102	0.648
HEAVY ST	23 (6.18)	0.312	0.690	43 (11.56)	0.327	0.993
RGARCH N	23 (6.18)	0.312	0.729	30 (8.06)	0.199	0.802
RGARCH T	24 (6.45)	0.218	0.712	43 (11.56)	0.327	0.993
RGARCH Lap	24 (6.45)	0.218	0.263	43 (11.56)	0.327	0.993
RGARCH ST	15 (4.03)	0.376	0.585	29 (7.80)	0.142	0.337
GAS T	28 (7.53)	0.037	0.363	52 (13.98)	0.015	0.570
GAS Lap	24 (6.45)	0.218	0.263	50 (13.44)	0.035	0.196
DJIA						
csl	26 (6.99)	0.096	0.511	46 (12.37)	0.141	0.733
log	25 (6.72)	0.147	0.579	46 (12.37)	0.141	0.733
eqw(1)	26 (6.99)	0.096	0.351	46 (12.37)	0.141	0.733
eqw(2)	27 (7.26)	0.060	0.418	48 (12.90)	0.073	0.567
GARCH N	32 (8.60)	0.004	0.830	48 (12.90)	0.073	0.567
GARCH T	29 (7.80)	0.022	0.570	53 (14.25)	0.010	0.254
GARCH Lap	28 (7.53)	0.037	0.492	47 (12.63)	0.102	0.134
GARCH ST	28 (7.53)	0.037	0.492	47 (12.63)	0.102	0.337
TGARCH N	26 (6.99)	0.096	0.351	43 (11.56)	0.327	0.609
TGARCH T	28 (7.53)	0.037	0.492	51 (13.71)	0.023	0.360
TGARCH Lap	24 (6.45)	0.218	0.668	45 (12.10)	0.190	0.461
TGARCH ST	25 (6.72)	0.147	0.755	44 (11.83)	0.252	0.532
HEAVY N	27 (7.26)	0.060	0.933	48 (12.90)	0.073	0.567
HEAVY T	28 (7.53)	0.037	0.977	53 (14.25)	0.010	0.494
HEAVY Lap	22 (5.91)	0.431	0.810	49 (13.17)	0.051	0.492
HEAVY ST	24 (6.45)	0.218	0.652	46 (12.37)	0.141	0.733
RGARCH N	23 (6.18)	0.312	0.729	30 (8.04)	0.194	0.763
RGARCH T	25 (6.72)	0.147	0.579	43 (11.56)	0.327	0.609
RGARCH Lap	13 (3.49)	0.160	0.331	30 (8.04)	0.194	0.763
RGARCH ST	13 (3.49)	0.160	0.331	29 (7.80)	0.142	0.570
GAS T	30 (8.06)	0.012	0.278	49 (13.17)	0.051	0.492
GAS Lap	27 (7.26)	0.060	0.418	45 (12.10)	0.190	0.821

(continued from previous page)

FTSE						
Sc. rule/Model	V(%)	$p_{uc}$	$p_{ind}$	V(%)	$p_{uc}$	$p_{ind}$
	95% VaR			90% VaR		
csl	25 (6.65)	0.162	0.312	33 (8.80)	0.430	0.504
log	23 (6.12)	0.336	0.184	32 (8.53)	0.333	0.171
eqw(1)	22 (5.85)	0.460	0.535	32 (8.53)	0.333	0.047
eqw(2)	23 (6.13)	0.330	0.736	32 (8.53)	0.333	0.047
GARCH N	24 (6.40)	0.232	0.232	34 (9.07)	0.541	0.098
GARCH T	25 (6.65)	0.162	0.080	41 (10.90)	0.564	0.029
GARCH Lap	22 (5.85)	0.460	0.030	33 (8.80)	0.430	0.074
GARCH ST	24 (6.40)	0.232	0.232	33 (8.80)	0.430	0.074
TGARCH N	22 (5.85)	0.460	0.535	30 (8.00)	0.183	0.090
TGARCH T	25 (6.65)	0.162	0.312	34 (9.07)	0.541	0.085
TGARCH Lap	15 (4.00)	0.358	0.623	31 (8.27)	0.250	0.118
TGARCH ST	15 (4.00)	0.358	0.623	28 (7.47)	0.088	0.176
HEAVY N	24 (6.38)	0.237	0.059	33 (8.80)	0.430	0.214
HEAVY T	25 (6.65)	0.162	0.788	41 (10.90)	0.564	0.073
HEAVY Lap	14 (3.73)	0.240	0.540	33 (8.80)	0.430	0.214
HEAVY ST	15 (4.00)	0.358	0.130	33 (8.80)	0.430	0.214
RGARCH N	26 (6.91)	0.106	0.375	32 (8.53)	0.333	0.430
RGARCH T	27 (7.18)	0.068	0.409	33 (8.80)	0.430	0.214
RGARCH Lap	24 (6.38)	0.237	0.230	31 (8.27)	0.250	0.774
RGARCH ST	26 (6.91)	0.106	0.375	32 (8.53)	0.333	0.430
GAS T	24 (6.40)	0.232	0.232	41 (10.90)	0.564	0.029
GAS Lap	16 (4.27)	0.504	0.169	33 (8.80)	0.430	0.192
Nikkei						
csl	11 (3.13)	0.084	0.399	20 (5.67)	0.003	0.890
log	13 (3.69)	0.239	0.317	20 (5.67)	0.003	0.890
eqw(1)	10 (2.83)	0.043	0.274	20 (5.67)	0.003	0.890
eqw(2)	11 (3.12)	0.082	0.341	20 (5.67)	0.003	0.890
GARCH N	14 (3.97)	0.356	0.575	24 (6.80)	0.034	0.767
GARCH T	13 (3.68)	0.234	0.492	27 (7.65)	0.126	0.957
GARCH Lap	12 (3.40)	0.144	0.414	23 (6.52)	0.021	0.678
GARCH ST	13 (3.68)	0.234	0.492	25 (7.08)	0.055	0.859
TGARCH N	12 (3.40)	0.144	0.414	24 (6.80)	0.034	0.569
TGARCH T	13 (3.68)	0.234	0.492	25 (7.08)	0.055	0.499
TGARCH Lap	10 (2.83)	0.043	0.444	19 (5.38)	0.002	0.979
TGARCH ST	12 (3.40)	0.144	0.414	25 (7.08)	0.055	0.499
HEAVY N	13 (3.69)	0.239	0.317	21 (5.95)	0.006	0.804
HEAVY T	14 (3.98)	0.362	0.576	25 (7.08)	0.055	0.859
HEAVY Lap	9 (2.56)	0.021	0.491	19 (5.38)	0.002	0.979
HEAVY ST	14 (3.98)	0.362	0.281	21 (5.95)	0.006	0.804
RGARCH N	13 (3.68)	0.234	0.080	18 (5.10)	0.001	0.009
RGARCH T	11 (3.12)	0.082	0.341	17 (4.82)	0.000	0.044
RGARCH Lap	9 (2.56)	0.021	0.215	14 (3.97)	0.000	0.575
RGARCH ST	12 (3.40)	0.144	0.005	16 (4.53)	0.000	0.003
GAS T	21 (5.97)	0.419	0.802	28 (7.93)	0.181	0.590
GAS Lap	11 (3.12)	0.082	0.341	26 (7.37)	0.085	0.431

regarding the frequency of violations and the test on equal accuracy.<sup>12</sup> Using equal weights leads to rejection of the nominal frequency of 5% for the DJIA (eqw(1) and eqw(2)) and S&P 500 (eqw(2)) indexes, while this is not the case for VaR estimates based on the *csl* score function. Furthermore, the violation frequencies for the 90% VaR estimates using the *csl* scoring rule are closer to the nominal value of 0.10 in case of the DJIA, S&P 500 and Nikkei indexes. According to the *t*-statistics of equal accuracy of the VaR estimates, using the *csl* score function produces significantly better VaR estimates than the benchmark in case of the S&P 500 index (95% VaR estimates) and the Nikkei index (95% and 90%).

Second, pooling based on the *csl* scoring rule improves pooling based on the log scoring rule, although the gain is less strong than the gain with respect to using equal weights. Considering the DJIA returns, using the whole density implies a violation frequency that is significantly different from 5% (using a significance level of 5%), while it is not significant using the *csl* scoring rule. Focusing on the S&P 500 and the Nikkei indexes, the number of violations does not differ, only the *t*-statistics are negative (although not significant) indicating that the VaR estimates of the *csl* score functions are slightly better. The opposite is true for the FTSE returns, where the violation frequency corresponding with the *csl* score function is too high (6.5%), and the positive *t*-statistics favor the log score function. This is in line with the results on the statistical tests presented in Table 2.

Third, only the HEAVY Skewed-*t* model consistently outperforms our method of combining density forecasts. Each remaining model fails at least once in the frequency of violations or in the test of equal accuracy. The best competitors are the HEAVY and RGARCH model classes. Consistent with the statistical results of 4.3.1, again the HEAVY model performs better according to the asymmetric tick-loss function and the number of violations than the RGARCH models although both model classes are quite similar in their specification. As noted earlier, this could be affected by the amount of parameter uncertainty when estimating the RGARCH class of models.

The differences between our various methods to estimate a VaR vanish if we put attention to the 5-day estimated VaRs, as indicated by Table 6. Using the Bonferroni bound corresponding with a 5% significance level, we conclude that using the *csl* scoring rule or

---

<sup>12</sup>For the sake of space, we do not show the *p*-values corresponding with the Independence test. Moreover, all methods/models pass this test. These results are available upon request.

the log scoring rule to obtain weights does not make a clear difference in the VaR estimates. In addition, the individual models perform also well. This could be explained partly by the decreasing power of the tests when the number of exceptions decreases.

To summarize, short-horizon VaR estimates improve compared to using equal weights and individual volatility models when using combined density forecasts based on the *csl* score function, either with respect to the nominal size and/or with respect to the statistical accuracy using the asymmetric tick-loss function of (22). The gain of using the *csl* score function instead of the log-score function is less strong, and holds in particular for the DJIA returns.

## 5 Conclusion

We investigate the benefits of combining density forecasts based on a specific region of interest. We develop a new density forecast method that combines density forecasts of different models based on the censored likelihood scoring rule (Diks *et al.*, 2011). Using daily returns from the S&P 500, DJIA, FTSE and Nikkei stock market indexes from 2000 until 2013, we apply our technique on recently developed univariate volatility models, including the HEAVY, GAS and Realized GARCH models.

Our results show that density forecasts in the tail are statistically more accurate if one pools density forecasts using the censored likelihood scoring rule than using density forecasts based on the log score rule, using the benchmark of equal weights or density forecasts of any individual volatility model. Second, we show that the 1-day 95% VaR estimates improve significantly compared to the combined density forecast method using equal weights and performs better than the method based on the log scoring rule in case of the DJIA stock market index. Moreover, the VaR estimates of 17 out of 18 individual models are beaten, either with respect to the nominal frequency of the VaR violations, or with respect to a statistical test on equal accuracy of the VaR estimates. Our results imply that risk managers and portfolio managers might benefit from combining density forecasts with the focus on the left tail using the *csl* scoring rule.

## References

- Aastveit, K.A., K.R. Gerdrup, A.S. Jore and L.A. Thorsrud (2011), Nowcasting GDP in real-time: A density combination approach, Working Paper.
- Amisano, G. and R. Giacomini (2007), Comparing density forecasts via weighted likelihood ratio tests, *Journal of Business and Economic Statistics* **25**, 177–190.
- Andersen, T., T. Bollerslev, F.X. Diebold and P. Labys (2003), Modeling and forecasting realized volatility, *Econometrica* **71**, 529–626.
- Bacharach, J. (1974), Bayesian dialogues, Working Paper.
- Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2008), Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica* **76**, 1481–1536.
- Bates, J.M. and C.W.J. Granger (1969), The combination of forecasts, *Operational Research* **20**, 451–468.
- Billio, M., R. Casarin, F. Ravazzolo and H.K. van Dijk (2013), Time-varying combinations of predictive densities using nonlinear filtering, *Journal of Econometrics* **177**, 211–232.
- Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307–327.
- Christoffersen, P. (1998), Evaluating interval forecasts, *International Economic Review* **39**, 841–862.
- Conflitti, C., C. De Mol and D. Giannone (2012), Optimal combination of survey forecasts, Working Paper.
- Creal, D., S.J. Koopman and A. Lucas (2013), Generalized autoregressive score models with applications, *Journal of Applied Econometrics* **28**, 777–795.
- Del Negro, M., R.B. Hasegawa and F. Schorfheide (2013), Generalised Density Forecast Combinations, Working Paper.

- Diebold, F.X. and R.S. Mariano (1995), Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**, 253–263.
- Diebold, F.X., T.A. Gunther and A.S. Tay (1998), Evaluating density forecasts with applications to financial risk management, *International Economic Review* **39**, 863–883.
- Diks, C., V. Panchenko and D. van Dijk (2011), Likelihood-based scoring rules for comparing density forecasts in tails, *Journal of Econometrics* **163**, 215–230.
- Fawcett, N., G. Kapetanios, J. Mitchell and S. Price (2014), Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance, Working Paper.
- Garratt, A., K. Lee, M.H. Peseran and Y. Shin (2003), Forecast uncertainties in macroeconomic modelling: an application to the UK economy, *Journal of the American Statistical Association* **98**, 829–838.
- Geweke, J. and G. Amisano (2011), Optimal prediction pools, *Journal of Econometrics* **164**, 130–141.
- Giacomini, R. and H. White (2006), Tests of conditional predictive ability, *Econometrica* **74**, 1545–1578.
- Giacomini, R. and I. Komunjer (2005), Evaluation and combination of conditional quantile forecasts, *Journal of Business and Economic Statistics* **23**, 416–431.
- Glosten, L., R. Jagannathan and D. Runkle (1993), On the relationship between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**, 1779–1801.
- Gneiting, T. and A.E. Raftery (2007), Strictly proper scoring rules, prediction and estimation, *Journal of the American Statistical Association* **102**, 359–378.
- Gneiting, T. and R. Ranjan (2013), Combining predictive distributions, *Electronic Journal of Statistics* **7**, 1747–1782.
- Granger, C.W.J. and M.H. Pesaran (2000), Economic and statistical measures of forecast accuracy, *Journal of Forecasting* **19**, 537–560.

- Hall, S.G. and J. Mitchell (2007), Combining density forecasts, *Journal of Forecasting* **23**, 1–13.
- Hansen, B.E. (1994), Autoregressive conditional density estimation, *International Economic Review* **35**, 705–730.
- Hansen, P.R., Z. Huang and H.H. Shek (2012), Realized GARCH: A joint model for returns and realized measures of volatility, *Journal of Applied Econometrics* **27**, 877–906.
- Jore, A.S., J. Mitchell and S. P. Vahey (2010), Combining forecast densities from VARs with uncertain instabilities, *Journal of Applied Econometrics* **25**, 621–634.
- Krüger, F. (2014), Combining Density Forecasts under various Scoring Rules: An Analysis of UK Inflation, Working Paper.
- Kupiec, P.H. (1995), Techniques for verifying the accuracy of risk measurement models, *Journal of Derivatives* **3**, 73–82.
- Mitchell, J. and S.G. Hall (2005), Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR fan charts of inflation, *Oxford Bulletin of Economics and Statistics* **67**, 995–1033.
- Nelson, D.B. (1991), Conditional heteroskedasticity in asset returns: a new approach, *Econometrica* **59**, 347–370.
- Rapach, D.E., J.K. Strauss and G. Zhou (2010), Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* **23**, 821–862.
- Shephard, N. and K. Sheppard (2010), Realising the future: forecasting with high-frequency-based volatility (heavy) models, *Journal of Applied Econometrics* **25**, 197–231.
- Stock, J. and M. Watson (2004), Combination forecasts of output growth in a seven-country data set, *Journal of Forecasting* **23**, 405–430.
- Timmermann, A. (2006), Forecast combinations, *Handbook of economic forecasting* **1**, 135–196.

Wallis, K.F. (2005), Combining density and interval forecasts: A modest approach, *Oxford Bulletin of Economics and Statistics* **67**, 983–994.



# Appendix

## A Optimizing weights

We follow Conflitti *et al.* (2012) to optimize the weights according to the log or *csl* score function of (3) and (5) respectively. We provide here only an outline of the algorithm.

Define  $\mathbf{p}(y_{t+1})$  as the vector of  $n$  density forecasts  $p_i(y_{t+1}) = p_{t+1}(y_{t+1}; Y_t, A_i)$  ( $i = 1, \dots, n$ ) of the variable  $y_{t+1}$  at time  $t$  over a one-day horizon. The combined density is then equal to:

$$p(y_{t+1}) = \mathbf{w}' \mathbf{p}(y_{t+1}) = \sum_{i=1}^n w_i p_i(y_{t+1}), \quad (\text{A.1})$$

with the assumption that the weights are positive and sum to one. For both scoring rules, we have to maximize the logarithm of the combined (censored) density over a given time period:

$$\Phi(\mathbf{w}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \log p(y_{t+1}). \quad (\text{A.2})$$

Note that we omitted the factor  $\frac{1}{T-1}$  in equation (4) and (1). This does not change the result as it is a constant. Define the  $(T-1) \times n$  matrix  $P$  with non-negative elements  $P_{ti} = p_i(y_{t+1})$ . Now, (A.2) can be rewritten as  $\frac{1}{T-1} \sum_{t=1}^{T-1} \log(P\mathbf{w}_t)$ . Denote  $\mathbf{w}_{opt}$  as the maximum of  $\Phi(\mathbf{w})$  subject to the weight constraints. Further, the Lagrange multiplier is introduced to take into account these constraints:

$$\Phi_\lambda(\mathbf{w}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \log(P\mathbf{w}_t) - \lambda \sum_{i=1}^n w_i. \quad (\text{A.3})$$

Instead of optimizing (A.3), Conflitti *et al.* (2012) consider the following ‘surrogate’ function, which depends on a vector  $\mathbf{a}$  of arbitrary weights:

$$\Psi_\lambda(\mathbf{w}; \mathbf{a}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{i=1}^n b_{ti} \log \left( \frac{w_i}{a_i} \sum_{l=1}^n \log P_{tl} a_l \right) - \lambda \sum_{i=1}^n w_i. \quad (\text{A.4})$$

with  $b_{ti} = \frac{P_{ti} a_i}{\sum_{l=1}^n P_{tl} a_l}$ . Further, the function has the properties  $\Psi_\lambda(\mathbf{a}; \mathbf{a}) = \Psi_\lambda(\mathbf{a})$  for any  $\mathbf{a}$

and  $\Psi_\lambda(\mathbf{w}; \mathbf{a}) \leq \Psi_\lambda(\mathbf{w})$  for any  $\mathbf{a}$  and  $\mathbf{w}$ .

The iterative algorithm is defined as

$$\mathbf{w}_\lambda^{(k+1)} = \arg \max_w \Psi_\lambda(\mathbf{w}; \mathbf{w}_\lambda^{(k)}) \quad (\text{A.5})$$

which yields a monotonic increase of  $\Psi_\lambda$ , according to the two aforementioned properties. Setting the derivatives of  $\Psi_\lambda(\mathbf{w}; \mathbf{w}_\lambda^{(k)})$  with respect to  $w_i$  equal to zero, leads to the maximum  $w_{\lambda,i} = (1/\lambda) \sum_{t=1}^{T-1} b_{ti}$ . Using the constraint that the weights should sum up to one, it holds that  $\lambda = T - 1$ . This changes (A.5) into

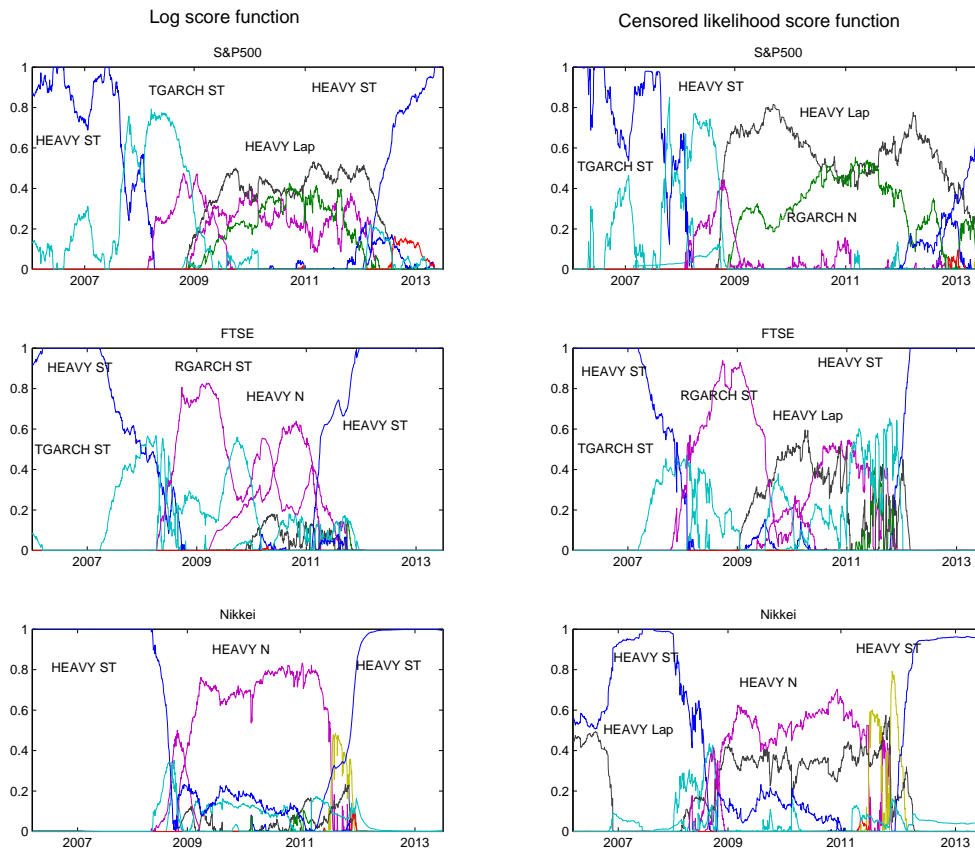
$$w_i^{(k+1)} = w_i^{(k)} \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{P_{ti}}{\sum_{l=1}^n P_{tl} w_l^{(k)}}, \quad (\text{A.6})$$

where we replace  $a_i$  by  $w_i^{(k)}$  in the expression of  $b_{ti}$ . We start the algorithm with equal weights, that is  $w_i^0 = 1/n$  and use as a stopping criterion a tolerance of  $10^{-6}$  of the sum of the absolute deviation of two successive iterations.

## B Pooling weights of remaining data sets and pooling results of the log score function

**Figure B.1: Pooling weights of the S&P 500, FTSE and Nikkei index**

This figure depicts the evolution of weights based on optimizing the logarithmic score function (left part) of (3) or the *csl* score function (right part) of (5) with a moving window of  $T = 750$  one-step ahead evaluated density forecasts using daily returns of the S&P500, FTSE and Nikkei indexes. In case of the *csl* score function,  $B_t$  the left tail  $y_t < \hat{r}^{0.25}$  with  $\hat{r}^{0.25}$  the 0.25th quantile of the empirical CDF of the in-sample returns. The labels refer to the models that have the highest weight at a given period. The abbreviations “ST”, “Lap” and “N” stand for Skewed-*t*, Laplace and Normal respectively.



**Table B.1: Evaluation of 1- and 5-day ahead censored density forecasts based on the log scoring rule**

This table reports results of testing equal predictive accuracy using the censored likelihood scoring rule of (4), with  $B_t$  the left tail  $y_t < \hat{r}^\kappa$  with  $\hat{r}^\kappa$  the  $\kappa$ th quantile of the empirical CDF of the in-sample returns. We set  $\kappa$  equal to 0.15 and 0.25 respectively. The weights are repeatedly optimized based on a the log score function of (3), using a moving window of 750 evaluated density forecasts. We focus on 1- and 5-step ahead density forecasts. The test statistic is given in (19) and compares censored density forecast with weights based on the log score function and density forecasts of each competing model, which are listed in Table 1. All models are estimated with a moving window of 750 daily returns from the S&P500, DJIA, FTSE and Nikkei index through the period January, 2000 - June, 2013. The test statistics are based on HAC-based standard errors and 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei) out-of-sample observations respectively.

Pooled (log score function) vs. individual								
	S&P500	DJIA	FTSE	Nikkei	S&P500	DJIA	FTSE	Nikkei
	1-step ahead forecasts				5-step ahead forecasts			
	$\kappa = 0.15$							
GARCH N	0.64	0.48	4.16***	1.86*	0.71	0.55	3.92***	1.81*
GARCH T	-0.85	-1.07	4.21***	0.35	-1.89*	-1.93*	4.47***	-0.34
GARCH Lap	-1.64	-1.89*	4.44***	0.11	-2.64***	-2.66***	4.32***	-0.75
GARCH ST	6.40***	5.01***	3.45***	3.35***	5.81***	4.69***	2.71***	3.57***
T GARCH N	-0.15	-0.17	4.36***	1.49	0.15	0.28	4.19***	1.68*
T GARCH T	-1.44	-1.56	4.15***	0.42	-2.03**	-1.77*	4.38***	-0.29
T GARCH Lap	-2.22**	-2.46**	4.23***	0.00	-3.06***	-2.94***	4.09***	-0.95
T GARCH ST	6.11***	4.77***	2.49**	3.27***	5.52***	4.77***	1.69*	3.41***
HEAVY N	-0.83	-0.77	3.08***	0.99	-0.33	-0.71	3.95***	1.38
HEAVY T	-2.12**	-2.18**	2.90***	-1.85*	-2.50**	-2.47**	4.16***	0.21
HEAVY Lap	-2.78***	-2.90***	3.15***	-1.52	-3.34***	-3.30***	3.54***	-1.46
HEAVY ST	5.75***	4.38***	0.02	2.84***	4.66***	3.70***	0.98	2.42**
RGARCH N	-1.16	-1.55	3.38***	1.88*	0.89	0.28	6.11***	3.30***
RGARCH T	-2.14**	-2.35**	3.46***	-0.20	-1.11	-0.90	5.67***	2.42**
RGARCH Lap	-2.47**	-2.61***	3.62***	1.50	-1.72*	-1.68*	4.93***	2.86***
RGARCH ST	4.70***	3.33***	-1.73*	4.79***	6.20***	6.40***	1.38	6.68***
GAS T	-0.71	-0.97	4.24***	0.65	-1.42	-1.84*	4.53***	0.47
GAS Lap	-1.59	-1.80*	4.43***	0.19	-2.59***	-2.63***	4.35***	-0.85
	$\kappa = 0.25$							
GARCH N	1.82*	1.58	6.01***	2.33**	1.43	1.17	5.28***	2.00**
GARCH T	0.68	0.49	6.01***	2.05**	-0.67	-0.84	6.06***	1.08
GARCH Lap	-0.29	-0.45	5.55***	1.97**	-1.61	-1.63	5.25***	0.60
GARCH ST	6.12***	5.09***	3.55***	3.09***	5.38***	4.44***	2.92***	3.11***
T GARCH N	1.17	1.09	6.42***	1.84*	1.08	1.07	5.68***	1.88*
T GARCH T	0.19	0.14	6.06***	2.13**	-0.73	-0.57	6.03***	1.05
T GARCH Lap	-0.80	-0.95	5.36***	1.75*	-2.00**	-1.86*	5.05***	0.20
T GARCH ST	5.34***	4.28***	2.69***	2.60***	4.64***	3.99***	1.89*	2.74***
HEAVY N	0.45	0.45	5.24***	1.35	0.73	0.42	5.69***	1.55
HEAVY T	-0.60	-0.45	4.92***	0.21	-1.07	-0.94	6.04***	1.01
HEAVY Lap	-1.46	-1.38	4.40***	0.38	-2.27**	-2.08**	4.61***	-0.08
HEAVY ST	5.05***	3.72***	-1.14	2.00**	4.63***	3.62***	0.31	2.28**
RGARCH N	0.55	0.15	5.70***	2.37**	2.36**	1.59	7.93***	3.59***
RGARCH T	-0.31	-0.60	5.50***	1.54	0.70	0.54	7.42***	3.28***
RGARCH Lap	-1.14	-1.28	4.80***	2.86***	-0.66	-0.74	5.77***	3.58***
RGARCH ST	5.40***	4.44***	-1.66*	7.17***	8.41***	8.03***	2.21**	9.02***
GAS T	0.71	0.50	6.03***	2.13**	-0.31	-0.83	6.12***	1.56
GAS Lap	-0.29	-0.41	5.52***	2.07**	-1.61	-1.65*	5.25***	0.64

**Table B.2: Log scores**

This table reports log scores corresponding with individual models and combined models, where the weights are based on optimizing the log score function of (3). The weights are repeatedly optimized based on a moving window of 750 evaluated density forecasts. The bold numbers represent the maximum of all models per data set. All models are estimated with a moving window of 750 daily returns from the S&P500, DJIA, FTSE and Nikkei index through the period January, 2000 - June, 2013. The number of out-of-sample observations are equal to 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei)respectively.

	S&P500	DJIA	FTSE	Nikkei	S&P500	DJIA	FTSE	Nikkei
	1-step ahead forecasts				5-step ahead forecasts			
GARCH N	-2687	-2610	-2338	-2500	-2735	-2647	-2400	-2642
GARCH T	-2651	-2574	-2314	-2434	-2672	-2588	-2348	-2491
GARCH Lap	-2642	-2568	-2338	-2454	-2662	-2583	-2364	-2499
GARCH ST	-2652	-2505	-1948	-2377	-2710	-2563	-2017	-2480
T GARCH N	-2642	-2563	-2320	-2509	-2701	-2620	-2384	-2618
T GARCH T	-2618	-2539	-2299	-2433	-2660	-2579	-2342	-2494
T GARCH Lap	-2616	-2539	-2327	-2453	-2648	-2570	-2357	-2498
T GARCH ST	-2582	-2419	-1905	-2364	-2666	<b>-2513</b>	-1977	-2483
HEAVY N	-2622	-2552	-2269	-2459	-2697	-2614	-2350	-2604
HEAVY T	-2603	-2532	-2263	-2399	-2660	-2584	-2321	-2490
HEAVY Lap	-2604	-2534	-2304	-2430	<b>-2646</b>	-2573	-2342	-2483
HEAVY ST	<b>-2542</b>	<b>-2402</b>	<b>-1812</b>	<b>-2301</b>	-2680	-2544	<b>-1914</b>	<b>-2464</b>
RGARCH N	-2635	-2560	-2279	-2518	-2780	-2682	-2416	-2766
RGARCH T	-2618	-2542	-2277	-2446	-2724	-2644	-2371	-2583
RGARCH Lap	-2616	-2544	-2314	-2480	-2703	-2624	-2380	-2597
RGARCH ST	-2612	-2489	-1835	-2670	-2872	-2748	-2041	-2988
GAS T	-2654	-2574	-2317	-2439	-2681	-2592	-2353	-2495
GAS Lap	-2644	-2568	-2339	-2457	-2662	-2583	-2366	-2494
pooled log	-2489	-2377	-1864	-2310	-2593	-2481	-1957	-2411