Received: 31 October 2016 | Revised: 2 March 2017 | Accepted: 3 March 2017
DOI: 10.1002/for.2472

RESEARCH ARTICLE

WILEY

The impact of parameter and model uncertainty on market risk predictions from GARCH-type models

David Ardia[1,2] | Jeremy Kolly[2,3] | Denis-Alexandre Trottier[2]

[1]Institute of Financial Analysis, University of Neuchâtel, Neuchâtel, Switzerland
[2]Finance, Insurance and Real Estate Department, Laval University, Québec, Canada
[3]Department of Management, University of Fribourg, Fribourg, Switzerland

Correspondence
Jeremy Kolly, Department of Management, University of Fribourg, Boulevard de Pérolles 90, CH-1700 Fribourg, Switzerland.
Email: jeremy.kolly@unifr.ch

Funding information
Swiss National Science Foundation, Grant Number: 158754; FQRSC, Grant Number: 2015-NP-179931

Abstract
We study the effect of parameter and model uncertainty on the left-tail of predictive densities and in particular on VaR forecasts. To this end, we evaluate the predictive performance of several GARCH-type models estimated via Bayesian and maximum likelihood techniques. In addition to individual models, several combination methods are considered, such as Bayesian model averaging and (censored) optimal pooling for linear, log or beta linear pools. Daily returns for a set of stock market indexes are predicted over about 13 years from the early 2000s. We find that Bayesian predictive densities improve the VaR backtest at the 1% risk level for single models and for linear and log pools. We also find that the robust VaR backtest exhibited by linear and log pools is better than the backtest of single models at the 5% risk level. Finally, the equally weighted linear pool of Bayesian predictives tends to be the best VaR forecaster in a set of 42 forecasting techniques.

KEYWORDS
GARCH models, Bayesian and frequentist estimation, predictive density combination, beta linear pool, censored optimal pooling, backtesting

1 | INTRODUCTION

Asset returns demonstrate volatility clustering and an abnormal amount of extreme values. The autoregressive conditional heteroskedastic (ARCH) model introduced by Engle (1982) is able to seize these empirical regularities. A more flexible specification, the generalized ARCH (GARCH) model, was later proposed by Bollerslev (1986). These models define the conditional volatility as a deterministic function of past innovations. However, they do not consider the leverage effect, that is, the asymmetric relation between news and volatility (Black, 1976). As a consequence, many asymmetric specifications for conditional volatility appeared around the 1990s (see, among others, Glosten, Jagannathan, & Runkle, 1993; Nelson, 1991; Zakoian, 1994). Furthermore, GARCH specifications were initially coupled with the normal conditional distribution. However, this appears insufficient to fully account for the asset return leptokurticity and skewness that can be empirically observed. Other distributions with fatter tails have been proposed, such as the standardized Student-*t* distribution (Bollerslev, 1987) or the generalized error distribution (Nelson, 1991), as well as methods to introduce skewness in these distributions (see, e.g., Fernández & Steel, 1998). Recently, GARCH-type models with complex updating mechanisms have appeared, such as those obtained from the generalized autoregressive score modeling framework (Creal, Koopman, & Lucas, 2013) with skewed and leptokurtic conditional distributions.

A predictive density fully depicts the uncertainty related to a prediction. GARCH-type models can typically be used to generate predictive densities for future returns of financial assets (e.g., indexes or stocks). In financial risk management, precise estimation of the left-tail of asset returns' predictive densities is crucial to reliably depict downside risk (Tay & Wallis, 2000). There are several kinds of predictive densities possessing different properties. Among them, Bayesian

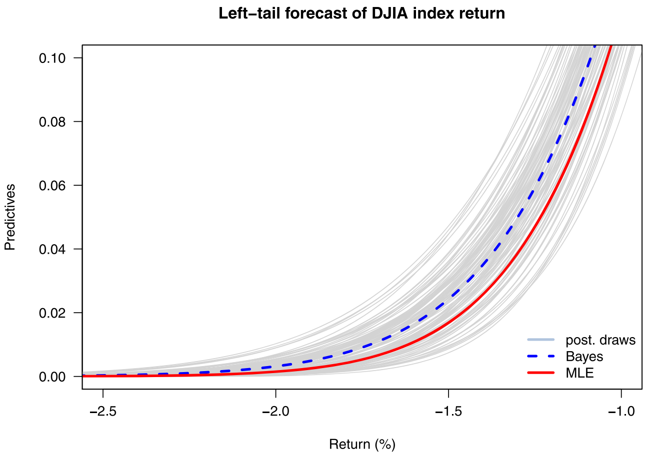

FIGURE 1 Typical left-tail of predictive densities generated by a GARCH-type model when we use particular posterior draws (thin solid lines), when parameter uncertainty is integrated using the Bayesian approach (bold dashed line), and when we plug ML estimates in the predictive (bold solid line). The Bayesian predictive is more conservative than the ML one and accounts for more likely scenarios [Colour figure can be viewed at wileyonlinelibrary.com]

predictive densities are of particular interest since they account for *parameter uncertainty* in a small-sample framework (Geweke & Amisano, 2010). Such predictive densities can improve out-of-sample left-tail predictive performance over those that do not integrate parameter uncertainty (Hoogerheide, Ardia, & Corré, 2012a), and the Bayesian approach is an appropriate way to account for parameter uncertainty when the purpose is to produce value-at-risk (VaR) estimates (Aussenegg & Miazhynskaia, 2006). However, it has never been shown in the literature that integrating parameter uncertainty can improve VaR forecasts; we aim at filling this gap. Figure 1 illustrates why integrating parameter uncertainty can be useful for left-tail prediction. The Bayesian predictive density (bold dashed line) is a particular averaging of the predictives that can be formed with individual posterior draws (thin solid lines). It is generally more conservative than the predictive density with plugged maximum likelihood (ML) estimates (bold solid line) and offers additional flexibility by accounting for all likely scenarios within the model structure. Nevertheless, it is also interesting to go beyond this structure by aggregating predictive densities originating from different models (see, among others, Genest & Zidek, 1986; Gneiting & Ranjan, 2013; Hall & Mitchell, 2007; Moral-Benito, 2015, section 5). This extra step allows us to account for *model uncertainty* and delivers further flexibility for downside risk prediction.[1]

In this research, we assess the impact of these two forms of uncertainty on the left-tail of predictive densities and in particular on VaR forecasts obtained from these densities. Our investigations are performed in the universe of GARCH-type models. Besides having been studied for decades in financial econometrics, these models are extensively used in the financial industry. The effect of parameter uncertainty is studied using Bayesian and ML estimation of GARCH-type models (Ardia, 2008). The effect of model uncertainty is investigated using the linear and log pools as well as the recent beta linear pool. Weights and parameters of the different pools are computed from past data. Methods for weight computation include Bayesian model averaging with predictive likelihoods (Eklund & Karlsson, 2007), as well as optimal pooling or OP (Geweke & Amisano, 2011, 2012). Broadly speaking, the former method averages measures of past predictive performance to form the weights, while the latter looks for the weights that maximize past predictive performance. We also use a censoring-based version of OP, referred to as COP, that allows us to focus on the left-tail (Opschoor et al. 2015). We contribute to the literature by applying this method to all of the previously mentioned pools, including the beta linear pool, and by comparing it to other combination methods, such as Bayesian model averaging. We investigate whether COP improves VaR forecasts for combinations of GARCH-type models.

Large forecasting experiments are carried out with several non-nested GARCH-type volatility specifications using skewed and heavy-tailed conditional distributions. We predict daily returns of a set of indexes over a window of about 13 years from the early 2000s. For each index, different predictive densities are produced and aggregated. Then, we evaluate VaR estimates obtained from individual and combined predictives. We also assess the quality of densities in the left-tail using probability integral transforms. We find that Bayesian predictive densities improve VaR estimates at the 1% risk level for individual models as well as for linear and log pools. We also find that the VaR backtest is more robust when linear or log pools are used and that VaR estimates from these methods are globally better than those of single models at the 5% risk level. Finally, the equally weighted linear pool of Bayesian predictives tends to be the best method for VaR prediction in a set of 42 forecasting techniques.

The outline of this paper is as follows. Section 2 presents GARCH-type models. Section 3 describes model estimation and the different types of predictive densities. Section 4 compares single model predictions in a first application to stock market indexes. Section 5 discusses the combination of predictive densities. Section 6 compares single and combined forecasts in a second application to stock market indexes. Section 7 concludes.

[1]Some studies already rely on model combination for VaR forecasting (see Massacci, 2015; Opschoor, van Dijk, & van der Wel, 2015; Pesaran, Schleicher, & Zaffaroni, 2009). However, they confine themselves to the linear pool and do not simultaneously account for parameter and model uncertainty.

2 | GARCH-TYPE MODELS

Let y_t be a return series with a negligible conditional mean such that we can write $y_t = \sigma_t \epsilon_t$, where the innovations ϵ_t are i.i.d. (independent and identically distributed) with zero mean and unit variance. In the generalized autoregressive conditionally heteroskedastic (GARCH) model, the conditional variance is given by[2]

$$\sigma_t^2 \equiv \alpha_0 + \alpha_1 y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (1)$$

where $\alpha_0 > 0$, $\alpha_1, \beta \geqslant 0$ to guarantee that Equation 1 is positive and where $\alpha_1 + \beta < 1$ to ensure stationarity. Although widely used in practice, the GARCH model does not account for the leverage effect first evidenced by Black (1976). This effect is the fact that negative returns tend to produce more volatility than positive ones. Around 1990, many specifications for conditional volatility appeared to capture this effect. We consider here three of them that are popular and non-nested. The first is the exponential GARCH (EGARCH) model (Nelson, 1991):

$$\ln \sigma_t^2 \equiv \alpha_0 + \alpha_1(|\epsilon_{t-1}| - E|\epsilon_{t-1}|) + \gamma \epsilon_{t-1} + \beta \ln \sigma_{t-1}^2, \quad (2)$$

where $|\beta| < 1$ is required for stationarity. The second is the GJR model (Glosten et al., 1993):

$$\sigma_t^2 \equiv \alpha_0 + \alpha_1 y_{t-1}^2 + \gamma y_{t-1}^2 I\{y_{t-1} < 0\} + \beta \sigma_{t-1}^2, \quad (3)$$

where $\alpha_0 > 0, \alpha_1, \beta \geqslant 0$ and $\alpha_1 + \gamma \geqslant 0$ and where $I\{\cdot\}$ is an indicator function equal to one when the condition in brackets holds and zero otherwise. Stationarity is ensured when $\alpha_1 + \gamma E[\epsilon_t^2 I\{\epsilon_t < 0\}] + \beta < 1$. The third is the threshold GARCH (TGARCH) model (Zakoian, 1994):

$$\sigma_t \equiv \alpha_0 + \alpha_1^+ y_{t-1}^+ - \alpha_1^- y_{t-1}^- + \beta \sigma_{t-1}, \quad (4)$$

where $\alpha_0 > 0, \alpha_1^+, \alpha_1^-, \beta \geqslant 0$ and where $y_t^+ \equiv \max\{y_t, 0\}$ and $y_t^- \equiv \min\{y_t, 0\}$. The stationarity condition for this model can be found in Francq and Zakoian (2010). In Equations 2–3, the leverage effect is captured by γ. In Equation 4, it stems from the difference between α_1^+ and α_1^-. Note that these three asymmetric volatility models exhibit different news impact curves. It is also noteworthy that the GJR and TGARCH models are nested in the asymmetric power ARCH model of Ding, Granger, and Engle (1993). However, we prefer the GJR and TGARCH models as they are more parsimonious and provide a more stable estimation.

Recently, Creal et al. (2013) suggested the generalized autoregressive score (GAS) modeling framework for latent variables, which uses the derivative of the log predictive likelihood as an updating mechanism. When applied to the conditional variance, it gives rise to a class of GARCH-type models. The GAS model is given by

$$\sigma_t^2 \equiv \alpha_0 + \alpha_1 s_{t-1} + \beta \sigma_{t-1}^2, \quad s_t \equiv S_t \nabla_t, \quad \nabla_t \equiv \frac{\partial}{\partial \sigma_t^2} \ln p(y_t | \sigma_t^2),$$

where $\alpha_0 > 0$, $\alpha_1, \beta \geqslant 0$ and where $p(\cdot)$ denotes a density function. The scaling factor S_t is defined as the inverse of the information matrix. We require $\beta < 1$ for stationarity and we ensure positive volatility. When the innovations are i.i.d. standard normal, this model is equivalent to Equation 1. However, more advanced distributions lead to complex updating mechanisms.

To properly account for the leptokurticity exhibited by the empirical distribution of asset returns, we model innovations with the standardized Student-t distribution with $\nu > 2$ degrees of freedom as proposed by Bollerslev (1987):

$$p(\epsilon_t|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)}} \left(1 + \frac{\epsilon_t^2}{\nu - 2}\right)^{-\frac{\nu+1}{2}},$$

where $\Gamma(\cdot)$ is the gamma function. As a non-nested alternative, we also consider the generalized error distribution (GED) with zero mean and unit variance (see Nelson, 1991):

$$p(\epsilon_t|\lambda) = \frac{\lambda \exp\left(-\frac{1}{2}|\epsilon_t/\varphi_\lambda|^\lambda\right)}{\varphi_\lambda 2^{\frac{\lambda+1}{\lambda}} \Gamma\left(\frac{1}{\lambda}\right)}, \quad \varphi_\lambda \equiv \left[\frac{\Gamma\left(\frac{1}{\lambda}\right)}{2^{\frac{2}{\lambda}} \Gamma\left(\frac{3}{\lambda}\right)}\right]^{\frac{1}{2}}.$$

The parameter $\lambda > 0$ controls tail thickness. The GED reduces to the standard normal distribution when $\lambda = 2$, while the Laplace distribution appears when $\lambda = 1$. Empirical distributions of asset returns are also typically skewed. We introduce skewness in the above distributions using the approach of Fernández and Steel (1998). A standardized skew distribution is given by

$$p^*(\epsilon_t|\xi, \vartheta) = \frac{2\sigma_\xi}{\xi + \frac{1}{\xi}} \, p \left[\frac{\sigma_\xi \epsilon_t + \mu_\xi}{\xi} I\{\epsilon_t \geqslant -\mu_\xi/\sigma_\xi\} + \xi(\sigma_\xi \epsilon_t + \mu_\xi) I\{\epsilon_t < -\mu_\xi/\sigma_\xi\} | \vartheta\right],$$

where $\xi > 0$ is the skewness parameter, ϑ denotes the parameters of the initial distribution and

$$\mu_\xi \equiv m_1 \left(\xi - \frac{1}{\xi}\right), \quad \sigma_\xi^2 \equiv (1 - m_1^2)\left(\xi^2 + \frac{1}{\xi^2}\right) + 2m_1^2 - 1,$$
$$m_1 \equiv 2 \int_0^\infty u \, p(u|\vartheta) \mathrm{d}u.$$

The previous expressions can be used to obtain a skew version of any symmetric unimodal density function which has zero mean and unit variance. We use formulas in Trottier and Ardia (2016) for computing moments of skew distributions.

[2]We consider GARCH-type models in their (1, 1) form. In this model class, best performances are often obtained from the most parsimonious specifications.

3 | MODEL ESTIMATION AND PREDICTIVE DENSITIES

Both Bayesian and maximum likelihood (ML) estimation methodologies are considered. To perform Bayesian estimation of a GARCH-type model M_k, we use an independence chain Metropolis–Hastings algorithm (Tierney, 1994) that simulates the posterior density $p(\theta_k|y, M_k)$ of the parameter vector $\theta_k \in \Theta_k$ given the sample of data y. The proposal distribution is constructed with the MitISEM method proposed in Hoogerheide, Opschoor and van Dijk (2012b). Furthermore, we use diffuse proper priors as in Ardia (2008) and confirmed with sensitivity analyses that they have a negligible influence on posterior results. To carry out ML estimation of M_k, we look for the vector of parameter estimates $\hat{\theta}_k \in \Theta_k$ that maximizes the log-likelihood function of the sample y using numerical methods.

GARCH-type models can be used to generate predictive densities for future asset returns. Let $\mathcal{M} \equiv \{M_1, \dots, M_K\}$ be a set of such models. The Bayesian predictive density provided by M_k for y_t given the sample $Y^r_{t-1} \equiv (y_{t-r}, \dots, y_{t-1})'$ of r past observations can be written as[3]

$$p(y_t|Y^r_{t-1}, M_k) = \int_{\Theta_k} p(y_t|Y^r_{t-1}, \theta_k, M_k) p(\theta_k|Y^r_{t-1}, M_k) \mathrm{d}\theta_k, \quad (5)$$

where $p(y_t|Y^r_{t-1}, \theta_k, M_k)$ is known analytically in GARCH-type models. The density in Equation 5 can easily be evaluated from a posterior sample. Furthermore, we see that it accounts for parameter uncertainty. This feature has already proven useful for improving GARCH predictive performance in terms of log score (Geweke & Amisano, 2010) and censored log score (Hoogerheide et al. 2012a). Another predictive density that also accounts for parameter uncertainty could be built, for instance, with the asymptotic sampling distribution of ML parameter estimates. However, as explained in Geweke and Amisano (2010), it is difficult to interpret besides being a large sample approximation. Furthermore, Aussenegg and Miazhynskaia (2006) show that the Bayesian approach is advantageous over other methods that account for parameter uncertainty when the purpose is to produce VaR estimates.

Of course, we can also produce predictive densities that condition on particular parameter values instead of integrating parameter uncertainty. Posterior means $\bar{\theta}_k$ computed from Y^r_{t-1} can be used to form the density $p(y_t|Y^r_{t-1}, \bar{\theta}_k, M_k)$, which we call the PM predictive density. It is also common to consider the ML predictive density $p(y_t|Y^r_{t-1}, \hat{\theta}_k, M_k)$, where ML estimates $\hat{\theta}_k$ obtained over Y^r_{t-1} are plugged into the density. In large samples, Bayesian, PM and ML predictives are similar because the posterior density is very concentrated around the mode of the likelihood function. However, in practice, they exhibit important differences as highlighted in Sections 4 and 6 for the left-tail.

Finally, the predictive densities introduced above can be used to compute a fundamental risk measure that is extensively used in the industry, known as the value-at-risk or VaR (Duffie & Pan, 1997; Jorion, 2007). It can be calculated as the $100\delta\%$ quantile of these predictive densities. In general, the risk level $\delta \in (0, 1)$ is fixed to 0.01 or 0.05 to consider left-tail risk. The above predictives can also be used to obtain the probability integral transform (PIT), that is, the predicted probability of having an outcome smaller than or equal to the actual realization. For a given sample, PITs from one-step-ahead predictives will be i.i.d. uniform if the predictives correspond to the data-generating process (DGP) densities (Diebold, Gunther, & Tay, 1998). Therefore, studying PIT independence and distributions allows us to evaluate one-step-ahead predictive densities. In what follows, we will assess the quality of our left-tail forecasts by considering only the PITs associated with left-tail outcomes.

4 | APPLICATION WITH STOCK MARKET INDEXES I

We consider daily returns (in percentage points) of eight major stock market indexes (CAC 40, DAX, DJIA, FTSE, Nikkei, NASDAQ, SMI and S&P 500) provided by the Oxford-Man Institute. These samples start from January 3, 2000 and are made up of 3265 observations (about 13 years). Note that the series are demeaned and that a first-order autoregressive filter is applied to each of them to focus on volatility and higher conditional moments.

We work here with the EGARCH, GJR, TGARCH, and GAS models with skew Student-t and skew GED innovations. These non-nested GARCH-type models are used to produce 1-day-ahead Bayesian, PM and ML predictive densities for each index. Rolling windows of 750 past observations are used for posterior and ML estimations. Such windows allow us to take into account potential parameter instability over time. Note that we set aside the first 500 predictives to be consistent with the application of Section 6, where they are used to form combination weights. For each series, we use predictives to compute 1% and 5% VaR estimates as well as PITs below 5%.

4.1 | Backtest methodologies

To backtest our VaR estimates, we first consider the standard unconditional coverage (UC) test (Kupiec, 1995), which is simply a likelihood ratio (LR) test for the correct proportion of VaR exceedances, or hits, and the conditional coverage (CC)

[3] Here we consider one-step-ahead predictions, but extension to larger forecast horizons is straightforward.

test of Christoffersen (1998), where the alternative hypothesis is that the hits follow a first-order Markov chain. We do not perform the CC test when the 1% risk level is involved as it is often invalid due to a lack of consecutive hits at this risk level. To gain power, we complement these tests with the Monte Carlo test of unconditional coverage (MCS) proposed in Ziggel, Berens, Weiss, and Wied (2014). It is based on the statistic:

$$\text{MCS} \equiv \sum_{t=1}^{T} I_t + \epsilon,$$

where $I_t \equiv I\{y_t < \text{VaR}_t^{\delta}\}$ and $\epsilon \sim$ i.i.d. $N(0, 0.001^2)$. Critical values for this test are obtained via Monte Carlo simulation. As our VaR estimates are generally not sufficiently conservative, we perform upper-tailed MCS tests. We also consider the CAViaR test (Engle & Manganelli, 2004), as implemented in Berkowitz, Christoffersen, and Pelletier (2011). We estimate the following logit model:

$$\Pr(I_t) \equiv \frac{\exp(\phi_0 + \phi_1 I_{t-1} + \phi_2 \text{VaR}_t^{\delta})}{1 + \exp(\phi_0 + \phi_1 I_{t-1} + \phi_2 \text{VaR}_t^{\delta})},$$

and perform an LR test of the joint hypothesis ($\phi_0 = \ln[\delta/(1-\delta)], \phi_1 = 0, \phi_2 = 0$). Finally, we calculate the (tick) asymmetric linear losses induced by our VaR forecasts:

$$L(y_t, \text{VaR}_t^{\delta}) \equiv (\delta - I_t)(y_t - \text{VaR}_t^{\delta}),$$

and test the significance of differences between mean tick losses with the Diebold–Mariano (DM) test (Diebold & Mariano, 1995) using a heteroskedasticity and autocorrelation consistent variance estimate.

To analyze PITs below 5%, we begin by rescaling and normalizing them (see Christoffersen & Pelletier, 2004). Then, we carry out the ARCH, JB and LR tests suggested in Deschamps (2012). The ARCH test is an F-test of the nullity of autoregression coefficients (intercept excluded) in a six-order autoregressive process of the squared PITs. We use a heteroskedasticity consistent covariance matrix estimate for this test. The JB test is the Jarque–Bera test for normality and the LR test is simply a test of $N(0, 1)$ against an unconstrained normal alternative.

4.2 | Backtest results

Table 1 reports the numbers of rejections at the 5% significance level over the set of eight indexes for the backtest of 2015 VaR estimates at the 1% and 5% risk levels. These quantities are obtained from Bayesian, PM, and ML predictives generated by our GARCH-type models. We see that, at the 1% risk level, Bayesian predictives show globally fewer rejections than PM or ML predictives. At the 5% risk level, the evidence is less clear. Performance of Bayesian and PM predictives is similar. Moreover, they both tend to provide better results in

TABLE 1 Numbers of rejections at the 5% significance level over eight indexes (CAC 40, DAX, DJIA, FTSE, Nikkei, NASDAQ, SMI, and S&P 500) for the backtest of our GARCH-type models using 2015 VaR estimates obtained from 1-day-ahead Bayesian, PM, and ML predictives

	Bayesian predictives			PM predictives			ML predictives		
	UC	MCS	CAViaR	UC	MCS	CAViaR	UC	MCS	CAViaR
1% risk level									
egarch.st	3	3	2	3	4	3	3	3	2
egarch.sged	3	3	1	3	3	1	3	4	2
gjr.st	0	1	0	0	1	0	0	1	1
gjr.sged	0	0	0	0	1	0	1	1	1
tgarch.st	2	1	0	2	1	0	1	1	0
tgarch.sged	1	1	0	1	1	0	1	1	0
gas.st	1	2	1	2	4	1	2	4	1
gas.sged	0	0	0	0	1	1	2	4	1
5% risk level									
egarch.st	1	2	2	1	3	2	2	2	1
egarch.sged	0	0	2	0	1	2	1	2	2
gjr.st	0	0	1	0	0	1	0	2	0
gjr.sged	0	0	0	0	0	0	0	0	0
tgarch.st	0	1	2	0	1	2	1	2	1
tgarch.sged	0	0	2	0	0	2	1	1	1
gas.st	1	3	0	1	2	0	1	2	0
gas.sged	0	0	0	0	0	0	0	0	0

Note. UC, MCS, and CAViaR tests are presented in Subsection 4.1. Definitions of model acronyms can be obtained from Table A1.

terms of unconditional coverage than ML predictives, while their CAViaR tests show more rejections than for ML predictives. Regarding models' performance, the skew GED tends to improve performance, especially at the 5% risk level. Note the perfect backtest of the GJR and GAS models with skew GED innovations at the 5% risk level. Globally, the EGARCH model provides the worst outcomes.

Table 2 presents the numbers of rejections over the eight indexes at the 5% significance level for the tests used to analyze rescaled PITs below 5%. PITs are computed from 2015 Bayesian, PM, and ML predictives generated by our GARCH-type models. We observe few rejections for the ARCH test whatever the predictive used. Regarding normality, the JB and LR tests favor Bayesian predictives over PM or ML predictives. Note, however, that the JB test for the skew Student-t TGARCH model is rejected five times over eight with Bayesian predictives.

In summary, the backtest in this application tends to favor 1% VaR estimates and left-tail forecasts generated from Bayesian predictives. For single models, it is therefore better to forecast the left-tail and compute 1% VaR estimates from predictives that integrate parameter uncertainty than from predictives that do not, such as the PM or ML ones. Now it remains to investigate if accounting for model uncertainty also improves our forecasts and under what circumstances. For this purpose, we consider model combination.

5 | COMBINATION OF PREDICTIVE DENSITIES

Predictive densities generated from models in $\mathcal{M}$ can be aggregated together. The older and probably the more intuitive formula used for this purpose is the linear pool (Stone, 1961):[4]

$$p_{\text{lin}}(y_t|Y^r_{t-1}, w_{t-1}) \equiv \sum_{k=1}^{K} w_{t-1,k}\, p(y_t|Y^r_{t-1}, M_k), \quad (6)$$

where $w_t \equiv (w_{t,1}, \dots, w_{t,K})'$ is a weight vector depending on data up to time t and satisfying the conditions $\sum_{k=1}^{K} w_{t,k} = 1$ and $w_{t,1}, \dots, w_{t,K} \geqslant 0$ in order for Equation 6 to be a valid density. The linear pool may be multimodal and tends to be overdispersed. Nevertheless, it performs well in many applications. According to Gneiting and Ranjan (2013), its empirical successes may be due to individual densities' underdispersion relative to the true density. On the other hand, Krüger (2014) shows that important scoring rules for the linear pool including the log scoring rule satisfy a lower bound. For those scoring rules, the linear pool score cannot be lower than a weighted average of individual densities' scores. He explains that the linear pool should thus outperform its components on average over time because, unlike its components' score, the linear pool score will exceed the lower bound at each time period.

Alternative nonlinear pools have appeared in the literature. A popular one is the log pool:

$$p_{\log}(y_t|Y^r_{t-1}, w_{t-1}) \equiv \frac{\prod_{k=1}^{K} p(y_t|Y^r_{t-1}, M_k)^{w_{t-1,k}}}{\int_{-\infty}^{\infty} \prod_{k=1}^{K} p(u|Y^r_{t-1}, M_k)^{w_{t-1,k}} \mathrm{d}u},$$

where the weights meet the same constraints as those of the linear pool for convenience. Its log kernel corresponds to a weighted average of log densities of individual models. The log pool is generally unimodal and less dispersed than the linear pool (Genest & Zidek, 1986). Moreover, its log score has a lower bound (Kascha & Ravazzolo, 2010; Krüger, 2014). In the area of inflation forecasting, Krüger (2014) finds some empirical evidence in favor of the log pool against the linear one, while discrimination is more difficult in Kascha and Ravazzolo (2010).

As discussed, for example, in Gneiting and Ranjan (2011), it is important for a predictive density to be well calibrated, that is, to be statistically consistent with the DGP. Inadequate calibration can be diagnosed by the PIT distribution. The beta linear pool devised by Ranjan and Gneiting (2010) and Gneiting and Ranjan (2013) comes along with an improved calibration. Its density is obtained by applying a beta distribution function to the distribution function of the linear pool and taking the derivative. We write it as follows:

$$p_{\text{blin}}(y_t|Y^r_{t-1}, w_{t-1}) \equiv \beta_{a,b}\left[\sum_{k=1}^{K} w_{t-1,k} \int_{-\infty}^{y_t} p(u|Y^r_{t-1}, M_k)\mathrm{d}u\right] \times \sum_{k=1}^{K} w_{t-1,k}\, p(y_t|Y^r_{t-1}, M_k),$$

where $\beta_{a,b}(\cdot)$ is the beta density with shape parameters $a, b > 0$ and where the non-negative weights add to one. The linear pool results when $a = b = 1$. Note that the beta linear pool is a special case of the generalized linear pool introduced by Kapetanios, Mitchell, Price, and Fawcett (2015) that lets the weights depend on y_t. Recent studies (Bassetti, Casarin, & Ravazzolo, 2015; Casarin, Mantoan, & Ravazzolo, 2016) also consider a mixture of beta calibration functions for pooling schemes; however, this will not be used in this research.

Besides the aggregation formula, the determination of the weights is also crucial. A basic solution to this problem is to use equal weights. This approach is successful for combining point forecasts (Clemen, 1989; Smith & Wallis, 2009; Stock & Watson, 2004). When the weights of a linear pool of Bayesian predictive densities are defined as posterior model

[4]For notational simplicity, we present aggregation formulas with Bayesian predictive densities. However, they also apply to other types of predictives.

TABLE 2 Numbers of rejections at the 5% significance level over eight indexes (CAC 40, DAX, DJIA, FTSE, Nikkei, NASDAQ, SMI, and S&P 500) for the analysis of rescaled PITs below 5%

	Bayesian predictives			PM predictives			ML predictives		
	ARCH	JB	LR	ARCH	JB	LR	ARCH	JB	LR
egarch.st	1	1	2	0	1	4	2	1	2
egarch.sged	2	3	2	0	3	3	0	4	5
gjr.st	0	2	1	1	3	3	0	2	1
gjr.sged	0	4	2	1	4	3	0	4	3
tgarch.st	1	5	1	0	1	3	0	1	1
tgarch.sged	0	4	2	0	4	1	0	4	2
gas.st	0	2	1	1	3	3	1	2	1
gas.sged	0	2	3	0	3	5	1	3	4

Note. PITs are obtained from 2015 1-day-ahead Bayesian, PM and ML predictives produced by our GARCH-type models. The ARCH, JB and LR tests are presented in Subsection 4.1. Definitions of model acronyms can be obtained from Table A1.

probabilities, we obtain the Bayesian model averaging (BMA) method. This approach stems naturally from probability rules and formally accounts for model uncertainty (Leamer, 1978, chapter 4). Of course, we can also heuristically account for model uncertainty by using posterior model probabilities with other aggregation formulas and other kinds of predictive densities. In the BMA method, the posterior model probability of a single model in $\mathcal{M}$ will be equal to one asymptotically even if this model is false. This situation is illustrated in Ardia and Kolly (2016). As this is questionable, we prefer to consider an implementation of BMA where marginal likelihoods are replaced by predictive likelihoods (Eklund & Karlsson, 2007). In this framework, BMA weights can be written as

$$w_{t-1,k}^{\mathrm{BMA}} \equiv \frac{p(Y_{t-1}^{s} | Y_{t-s-1}^{r}, M_k)}{\sum_{l=1}^{K} p(Y_{t-1}^{s} | Y_{t-s-1}^{r}, M_l)},$$

where we assume equal prior model probabilities and where s corresponds to the size of the weight estimation window. This implementation of BMA implies slower convergence to a single model and allows us to work with diffuse priors. Note that the BMA weights do not depend on the aggregation formula and that BMA does not provide a way to estimate a and b in the beta linear pool.

An alternative to BMA is the optimal pooling (OP) approach introduced in Hall and Mitchell (2007) and subsequently deepened in Geweke and Amisano (2011, 2012). For a given pool $p_c(y_t | Y_{t-1}^{r}, w_{t-1})$, the OP weights are given by

$$w_{t-1}^{\mathrm{OP}-c} \equiv \underset{w_{t-1}}{\arg\max} \sum_{\tau=t-s}^{t-1} \ln p_c(y_\tau | Y_{\tau-1}^{r}, w_{t-1}), \tag{7}$$

where optimal weights must be found in the unit simplex. The objective function in Equation 7 adds pool log scores over the sample Y_{t-1}^{s} and its maximization corresponds to the minimization of the Kullback–Leibler (KLIC) distance from the DGP to the pool (Hall & Mitchell, 2007). Furthermore, when the DGP is a model or a linear or log pool of models in $\mathcal{M}$, the OP method asymptotically recovers true weights (Krüger, 2014). However, we observed in our own experiments that uncertainty around the true weights can remain substantial even with large estimation windows. On the other hand, when the DGP cannot be obtained from models in $\mathcal{M}$, several OP limiting weights are typically positive.

In risk management, pool predictive densities can be used to obtain VaR estimates (see, e.g., Ardia & Kolly, 2016). Therefore, it is possible to improve these estimates by using weights that give more importance to left-tail outcomes. Opschoor et al. (2015) propose to replace the log score in Equation 7 by the censored scoring rule introduced in Diks, Panchenko, and van Dijk (2011).[5] This scoring rule is equal to the log predictive likelihood when the outcome falls below a threshold and to the log probability mass above the threshold otherwise. It thus neglects the shape of the predictive distribution above the threshold. The censored OP (COP) weights can be obtained as follows:

$$w_{t-1}^{\mathrm{COP}-c} \equiv \underset{w_{t-1}}{\arg\max} \sum_{\tau=t-s}^{t-1} I\{y_\tau < \hat{q}_\psi\} \ln p_c(y_\tau | Y_{\tau-1}^{r}, w_{t-1}) + I\{y_\tau \geqslant \hat{q}_\psi\} \ln \int_{\hat{q}_\psi}^{\infty} p_c(u | Y_{\tau-1}^{r}, w_{t-1}) \mathrm{d}u, \tag{8}$$

subject to the usual weight restrictions. We define the censoring bound $\hat{q}_\psi$ as the empirical $100\psi\%$ quantile obtained from the sample Y_{t-1}^{s}. The choice of ψ thus determines the percentage of uncensored observations used to compute the weights.

Specific expressions are very useful for implementing the COP method. One is provided in Opschoor et al. (2015) in the linear pool case, while another is given in Ardia and Kolly

[5]Note that another approach using censoring in a BMA framework to generate weights that give more importance to tail events is proposed in Gatarek, Hoogerheide, Hooning, and van Dijk (2014).

(2016) in the log pool case. For the beta linear pool, it can be shown that Equation 8 reduces to

$$w_{t-1}^{\text{COP-blin}} \equiv \underset{w_{t-1},a,b}{\arg\max} \sum_{\tau=t-s}^{t-1} I\{y_\tau < \hat{q}_\psi\} \ln p_{\text{blin}}(y_\tau | Y_{\tau-1}^r, w_{t-1})$$
$$+ I\{y_\tau \geqslant \hat{q}_\psi\} \ln \left(1 - B_{a,b} \left[\sum_{k=1}^{K} w_{t-1,k} \int_{-\infty}^{\hat{q}_\psi} p(u|Y_{\tau-1}^r, M_k) \mathrm{d}u \right] \right),$$

where $B_{a,b}(\cdot)$ is the beta distribution function. Let us now illustrate the relevance of using COP for the beta linear pool when we are interested in the calibration of the left-tail. We simulate 1000 observations from the mixture DGP $0.6N(-2, 1) + 0.4N(2, 1)$ to form an estimation period and 1000 other observations from the same DGP to constitute a forecasting period. We consider the misspecified models $N(-1, 1)$ and $N(1, 1)$ and combine them with linear and beta linear pools whose weights and parameters are estimated with the OP and COP methods on the estimation period. In the COP method, the censoring bound is fixed to -2 such that about 30% of the lowest observations are uncensored. Then, we compute PITs below 30% on the forecasting period and rescale them so that they lie between 0 and 1. Figure 2 presents the empirical distribution functions of these tail PITs with straight lines passing through (0, 0) and (1, 1) representing the ideal behavior. We see that the tail PITs provided by the beta linear pool are closer to the straight line. This was already highlighted in Bassetti et al. (2015) and Casarin et al. (2016). Furthermore, we observe that the calibration is even better when COP is used to estimate the beta linear pool. We have this because COP focuses on the left-tail.

6 | APPLICATION WITH STOCK MARKET INDEXES II

Owing to the computational cost implied by model combination, and by analogy with Opschoor et al. (2015), we consider here a subset of four indexes (DJIA, FTSE, Nikkei and S&P 500) among those described in Section 4. We work again with the EGARCH, GJR, TGARCH, and GAS models with skew Student-t and skew GED innovations.

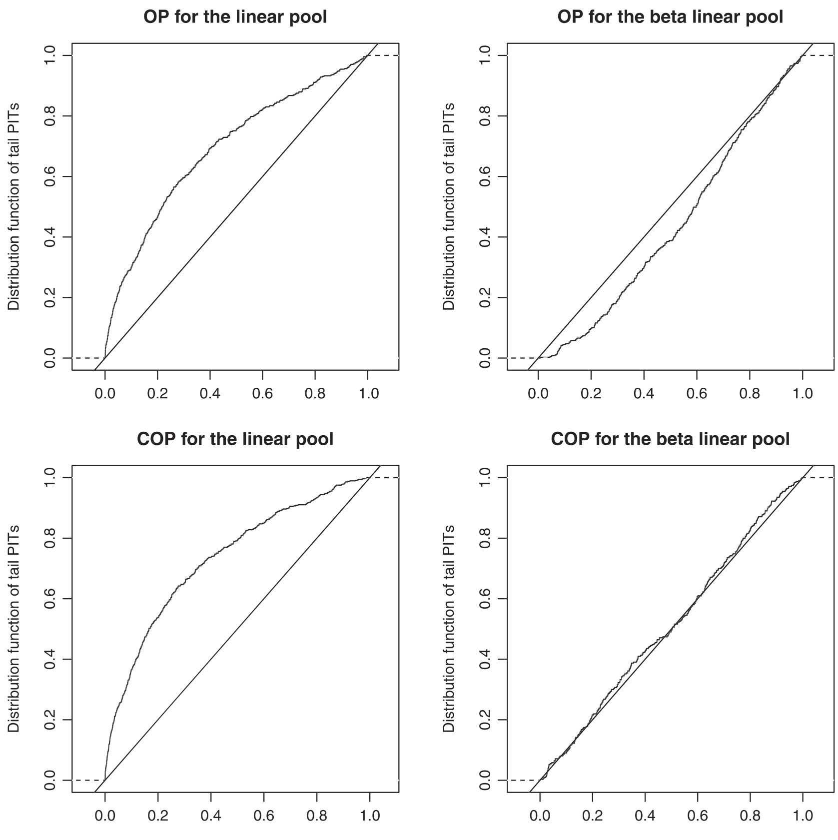

FIGURE 2 Empirical distribution functions of tail PITs rescaled to lie between 0 and 1, with ideal straight lines passing through (0, 0) and (1, 1). Tail PITs are produced by linear and beta linear pools of misspecified models over a forecasting period. Weights are estimated with the OP and COP methods from observations of an estimation period

TABLE 3 Statistical significance results for the backtest of 1-day-ahead VaR forecasts at the 1% risk level obtained from Bayesian and ML predictive densities produced by our GARCH-type models and from combinations of these densities. VaR forecasts are generated for 2015 daily returns of the DJIA, FTSE, Nikkei and S&P 500 indexes

	Bayesian predictives			ML predictives		
	UC	MCS	CAViaR	UC	MCS	CAViaR
DJIA						
egarch.st	**	***	***	**	***	***
egarch.sged	**	**	***	***	***	***
gjr.st					*	**
gjr.sged						*
tgarch.st						
tgarch.sged					*	
gas.st		*			*	**
gas.sged			*	*	*	***
lin.ew						
lin.bma						
lin.op						*
lin.cop-0.15			*		*	
lin.cop-0.25			*			
log.ew					*	
log.bma		*	*			
log.op			*			
log.cop-0.15			**			*
log.cop-0.25			**		*	
beta.op			***			**
beta.cop-0.15			***			***
beta.cop-0.25			***			***
FTSE						
egarch.st						
egarch.sged					*	
gjr.st						
gjr.sged		*				
tgarch.st						
tgarch.sged		*	*		*	
gas.st	*	*		**	**	*
gas.sged		*		**	**	*
lin.ew				*	**	
lin.bma		*		*	**	
lin.op		*		**	**	
lin.cop-0.15				*	**	
lin.cop-0.25				*	**	
log.ew				**	**	
log.bma		*		*	**	
log.op				*	**	
log.cop-0.15				*	*	
log.cop-0.25				*	**	
beta.op	*	**	***	**	**	***
beta.cop-0.15			**			*
beta.cop-0.25			**			**
Nikkei						
egarch.st					*	
egarch.sged				*	*	
gjr.st						
gjr.sged						
tgarch.st						
tgarch.sged						
gas.st						
gas.sged						
lin.ew						
lin.bma						
lin.op						
lin.cop-0.15						
lin.cop-0.25						
log.ew						
log.bma						
log.op						
log.cop-0.15						
log.cop-0.25						
beta.op	**	***	***	*	**	*
beta.cop-0.15		*				*
beta.cop-0.25	**	**	**	**	**	*
S&P 500						
egarch.st	**	***	*	**	***	*
egarch.sged	**	**		**	**	
gjr.st					*	
gjr.sged						
tgarch.st						
tgarch.sged						
gas.st		*		*	**	*
gas.sged					*	
lin.ew						
lin.bma				*	*	
lin.op		*			*	
lin.cop-0.15		*			*	
lin.cop-0.25		*				
log.ew					*	
log.bma				*	**	*
log.op	*	**	*	*	**	*
log.cop-0.15		*			*	
log.cop-0.25	*	**				
beta.op	**	**	***		*	**
beta.cop-0.15			**			*
beta.cop-0.25			*			*

Note. Asterisks indicate significance at 10%(*), 5%(**), and 1%(***). The UC, MCS and CAViaR tests are presented in Subsection 4.1. Definitions of model acronyms can be obtained from Table A1.

TABLE 4 Statistical significance results for the backtest of 1-day-ahead VaR forecasts at the 5% risk level obtained from Bayesian and ML predictive densities produced by our GARCH-type models and from combinations of these densities. VaR forecasts are generated for 2015 daily returns of the DJIA, FTSE, Nikkei, and S&P 500 indexes

	Bayesian predictives				ML predictives			
	UC	CC	MCS	CAViaR	UC	CC	MCS	CAViaR
DJIA								
egarch.st			*				*	
egarch.sged								
gjr.st			*				*	
gjr.sged								
tgarch.st								
tgarch.sged								
gas.st	**	*	**		*		**	
gas.sged			*				*	
lin.ew								
lin.bma								
lin.op								
lin.cop-0.15								
lin.cop-0.25								
log.ew								
log.bma								
log.op								
log.cop-0.15							*	
log.cop-0.25								
beta.op				*				**
beta.cop-0.15								
beta.cop-0.25								
FTSE								
egarch.st	*	**	**	**				*
egarch.sged		**		**		*		**
gjr.st			*				*	
gjr.sged			*					
tgarch.st		**	*	**	*	*	**	*
tgarch.sged		***	*	**		*	*	*
gas.st	*		**				*	
gas.sged							*	
lin.ew								
lin.bma		*	*				*	
lin.op		*	**				*	
lin.cop-0.15		*	*				*	
lin.cop-0.25			*					
log.ew			*					
log.bma		*	*				*	
log.op		*	*				*	
log.cop-0.15		*	*				*	
log.cop-0.25		*	**				*	
beta.op		**	*	***				***
beta.cop-0.15				***				***
beta.cop-0.25				***				***
Nikkei								
egarch.st								
egarch.sged								
gjr.st								
gjr.sged								
tgarch.st								
tgarch.sged								
gas.st								
gas.sged								
lin.ew								
lin.bma								
lin.op								
lin.cop-0.15								
lin.cop-0.25								
log.ew								
log.bma								
log.op								
log.cop-0.15								
log.cop-0.25								
beta.op								
beta.cop-0.15								
beta.cop-0.25								
S&P 500								
egarch.st	**	*	**		**	*	***	
egarch.sged			*		**		**	
gjr.st			*		*		**	
gjr.sged								
tgarch.st								
tgarch.sged								
gas.st	*		**		**	*	**	
gas.sged							*	
lin.ew								
lin.bma								
lin.op							*	
lin.cop-0.15								
lin.cop-0.25								
log.ew								
log.bma								
log.op							*	
log.cop-0.15								
log.cop-0.25							*	
beta.op								
beta.cop-0.15								
beta.cop-0.25								

Note. Asterisks indicate significance at 10%(*), 5%(**), and 1%(***). The UC, CC, MCS, and CAViaR tests are presented in Subsection 4.1. Definitions of model acronyms can be obtained in Table A1.

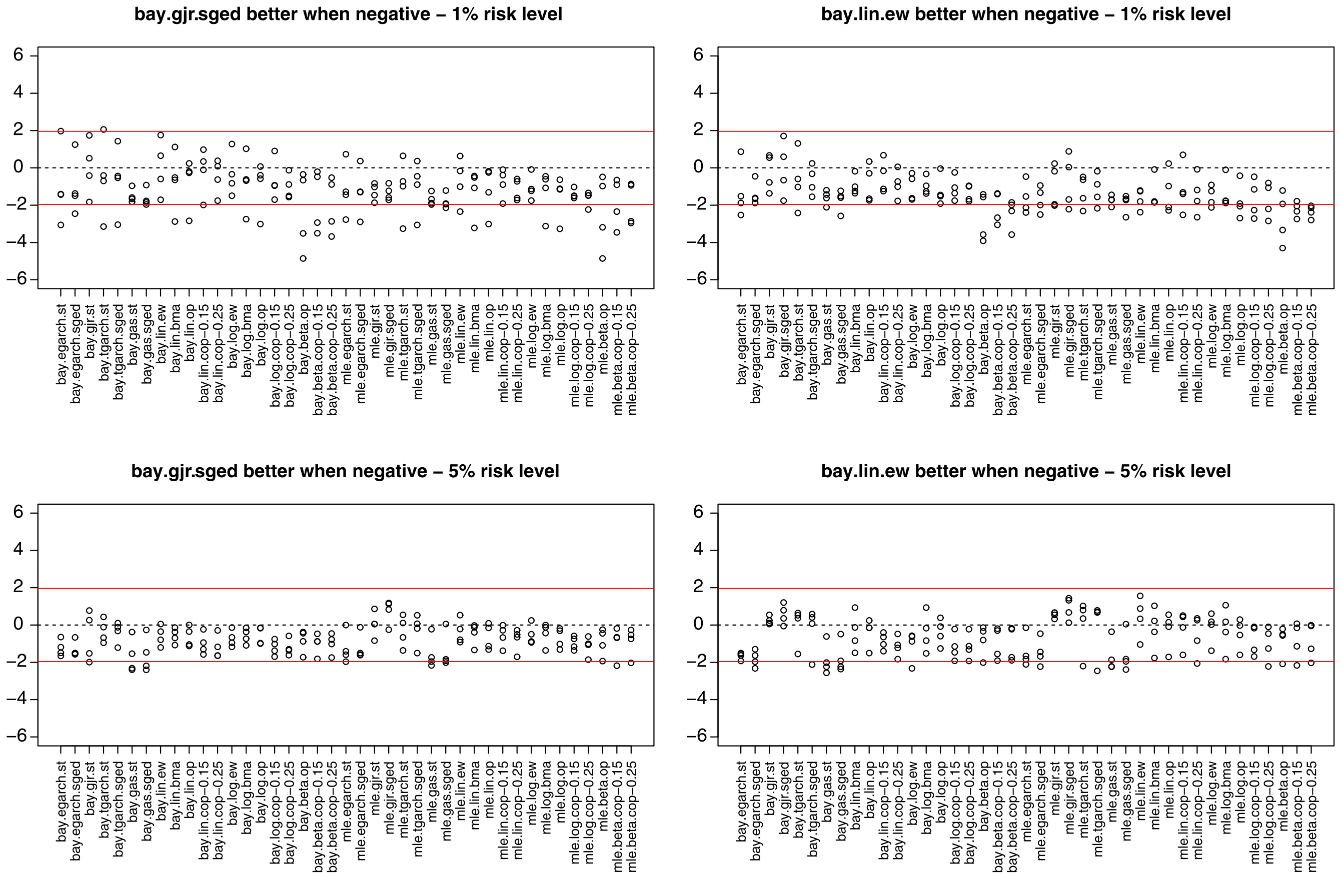

FIGURE 3 DM test statistics for significance of differences between mean (tick) asymmetric linear losses computed for 2015 daily returns of the DJIA, FTSE, Nikkei, and S&P 500 indexes. Left panels compare the skew GED GJR model with Bayesian predictives to all other models. Right panels compare the EW linear pool of Bayesian predictives to all other models. Top panels use 1% VaR forecasts, and bottom panels consider 5% VaR forecasts. For each model pair, each circle corresponds to a different index. Red lines indicate asymptotic critical values at the 5% level. More details on the testing methodology can be found in Subsection 4.1. Definitions of model acronyms can be obtained from Table A1 [Colour figure can be viewed at wileyonlinelibrary.com]

We produce 1-day-ahead Bayesian and ML predictive densities for each index using rolling estimation windows of 750 past observations. Predictives are then aggregated using linear and log pools with equal, BMA, OP, and COP weights and the beta linear pool with OP and COP weights. Rolling windows of 500 past observations are used for weight estimation.[6] They allow us to account for potential variation in model prevalence over time. In the COP method, we define censoring bounds such that 15% and 25% of the lowest observations of these windows are uncensored. This way, we keep enough uncensored observations to form model weights based on their left-tail performance. For each index, 1% and 5% VaR estimates as well as PITs below 5% are computed.

6.1 | Backtest results

The methodologies used to backtest our forecasts are described in Subsection 4.1. Tables 3 and 4 present the backtest of 2015 VaR estimates at the 1% and 5% risk levels, respectively, obtained from Bayesian and ML predictives of our models and model combinations. We only report statistical significance results at 1%, 5%, and 10% to ease analysis. Tables with test statistics are available upon request from the authors. We start by studying results at the 1% risk level in Table 3. We observe that Bayesian predictives globally provide better unconditional coverage for single models and under linear and log pooling. A notable exception is the log pool with COP weights and 25% censoring for the S&P 500 index, where ML predictives are preferable. Regarding individual models, the skew Student TGARCH model delivers the best outcomes, whereas the performance of the EGARCH and GAS models is poor. Interestingly, the results among linear or log pools are more similar than among single models given a particular index. Note the very good VaR backtest of the equally weighted (EW) linear pool. On the other hand, the beta linear pool shows puzzling outcomes. It can give correct

[6]Parameters a and b in the beta linear pool are also estimated on these windows.

unconditional coverage, especially when COP is used for estimation. However, its CAViaR test is systematically rejected, indicating that the calibration provided by the beta linear pool is detrimental to the independence of VaR violations. Moreover, Bayesian predictives seem to be harmful to the beta linear pool.

We now turn to the 5% risk level in Table 4. In this case, discrimination among Bayesian and ML predictives is more difficult. Bayesian predictives are favored for the S&P 500 index, whereas they are not for the FTSE index. For single models, the EGARCH model and the skew Student GAS models give bad results again. Note also the poor VaR backtest exhibited by the TGARCH model for the FTSE index. Given a particular index, results of combination methods are again more homogeneous than among single models. They are also globally better than those of single models. It is noteworthy that the EW linear pool is the only method providing a perfect backtest for the FTSE index and that the puzzling behavior of the beta linear pool can only be observed for the FTSE index.

Figures 3 and 4 present DM test statistics for significance of differences between mean (tick) asymmetric linear losses computed for 2015 VaR estimates at the 1% and 5% risk levels. Figure 3 compares the skew GED GJR model and the EW linear pool with Bayesian predictives to all other forecasting techniques, while Figure 4 compares these two methods with ML predictives to all other techniques. In both Figures, we observe at the 1% risk level that—relative to our two benchmarks—forecasting techniques using ML predictives are less accurate than those using Bayesian predictives. This phenomenon is not observable at the 5% risk level. Overall, the skew GED GJR model and the EW linear pool exhibit similar performance. However, a larger number of significant differences favor the EW linear pool at the 5% risk level. Finally, note that the beta linear pool performs very badly against both benchmarks at the 1% risk level, whereas there are few significant differences at the 5% risk level.

Table 5 presents the analysis of rescaled PITs below 5% derived from 2015 Bayesian and ML predictives produced by our models and model combinations. Again, we only report statistical significance results at 1%, 5%, and 10% to facilitate analysis and can provide tables with test statistics upon request. Regarding the ARCH test, it shows few rejections.

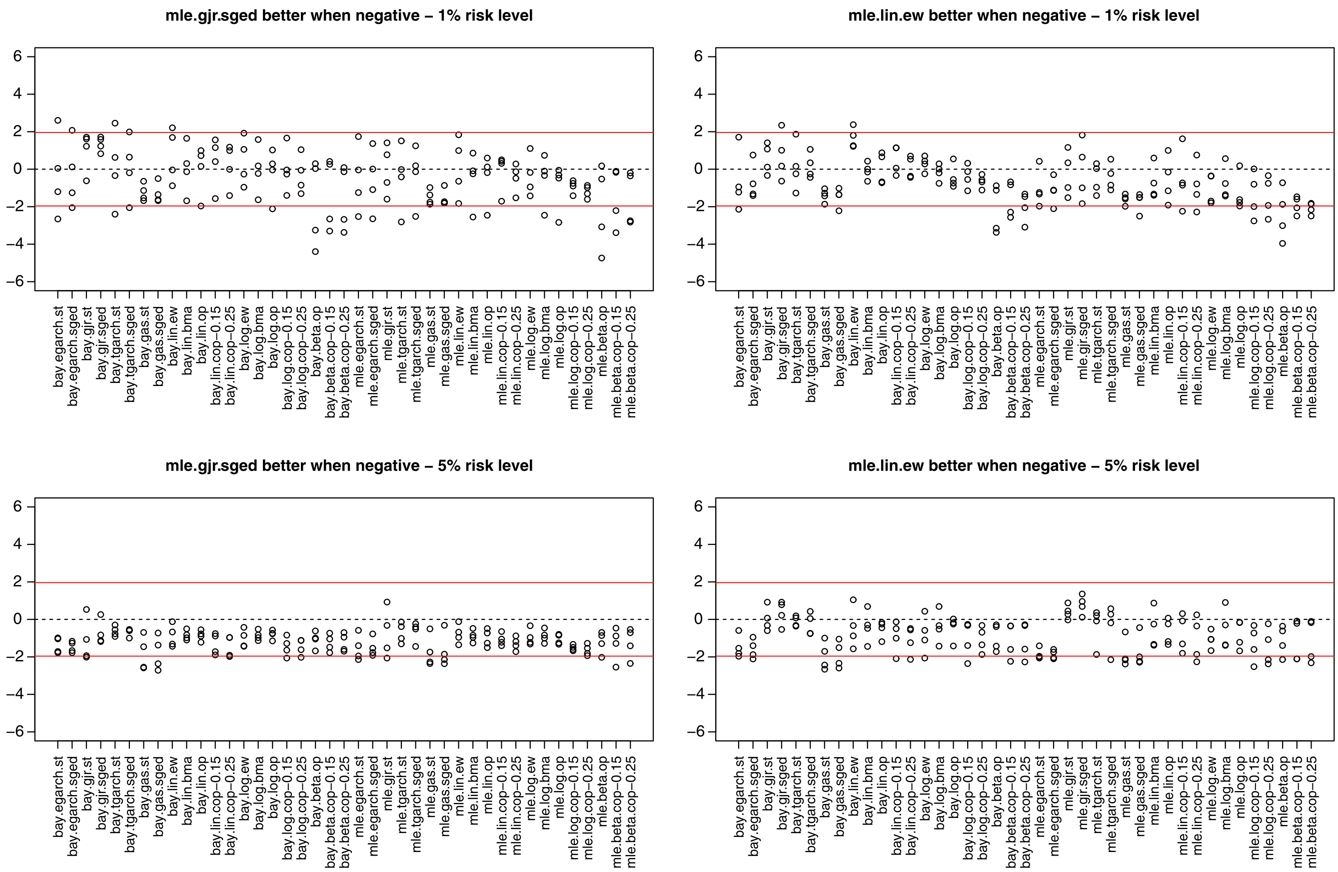

FIGURE 4 DM test statistics for significance of differences between mean (tick) asymmetric linear losses computed for 2015 daily returns of the DJIA, FTSE, Nikkei, and S&P 500 indexes. Left panels compare the skew GED GJR model with ML predictives to all other models. Right panels compare the EW linear pool of ML predictives to all other models. Top panels use 1% VaR forecasts, and bottom panels consider 5% VaR forecasts. For each model pair, each circle corresponds to a different index. Red lines indicate asymptotic critical values at the 5% level. More details on the testing methodology can be found in Subsection 4.1. Definitions of model acronyms can be obtained in Table A1 [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Statistical significance results for the analysis of rescaled PITs for observations below 1-day-ahead VaR forecasts at the 5% risk level obtained from Bayesian and ML predictive densities and combinations of such densities. VaR forecasts are generated for 2015 daily returns of the DJIA, FTSE, Nikkei and S&P 500 indexes

TABLE 5 Continued

	Bayesian predictives			ML predictives		
	ARCH	JB	LR	ARCH	JB	LR
DJIA						
egarch.st						
egarch.sged			*			**
gjr.st						
gjr.sged		***			***	
tgarch.st		**				
tgarch.sged		*			***	
gas.st						
gas.sged					*	
lin.ew						
lin.bma					*	
lin.op						
lin.cop-0.15		*				
lin.cop-0.25		***				
log.ew		*				
log.bma					***	
log.op					**	
log.cop-0.15		**				
log.cop-0.25		**			*	
beta.op					***	
beta.cop-0.15		***			***	
beta.cop-0.25		***			***	
FTSE						
egarch.st				**		
egarch.sged						**
gjr.st				*		
gjr.sged			**			**
tgarch.st	*					
tgarch.sged	**					*
gas.st	*					*
gas.sged	*		*			**
lin.ew	**					
lin.bma						
lin.op						
lin.cop-0.15						
lin.cop-0.25						
log.ew						*
log.bma						
log.op						
log.cop-0.15						
log.cop-0.25						
beta.op						
beta.cop-0.15						
beta.cop-0.25						
Nikkei						
egarch.st			***			**
egarch.sged		***	***		***	***
gjr.st		**	*		***	**
gjr.sged		***	**		***	***
tgarch.st			**			**
tgarch.sged		***	***		***	***
gas.st		***	***		***	***
gas.sged		***	***		***	***
lin.ew		***	***		**	**
lin.bma		***	**		***	***
lin.op		**	**		***	**
lin.cop-0.15		***	**		***	**
lin.cop-0.25		***	**		***	***
log.ew		***	***		***	***
log.bma		***	***		***	***
log.op		***	**		***	***
log.cop-0.15	**	***	***		***	***
log.cop-0.25		***	***		***	***
beta.op			**			**
beta.cop-0.15		**			***	
beta.cop-0.25		*	**		***	**
S&P 500						
egarch.st						
egarch.sged					***	**
gjr.st						
gjr.sged		**			***	
tgarch.st						
tgarch.sged		***	*		***	
gas.st						
gas.sged		**			**	
lin.ew			*		**	
lin.bma					**	*
lin.op						
lin.cop-0.15					***	
lin.cop-0.25					***	
log.ew					***	
log.bma					***	
log.op					***	
log.cop-0.15					***	
log.cop-0.25					***	
beta.op		*	*		***	**
beta.cop-0.15		***	*		***	
beta.cop-0.25		***			***	

Note. Asterisks indicate significance at 10%(*), 5%(**), and 1%(***). The ARCH, JB and LR tests are presented in Subsection 4.1. Definitions of model acronyms can be obtained in Table A1.

There is, however, an issue with normality for the Nikkei index and also for the S&P 500 index to a lesser extent. We found that this is due to very large negative returns that occur for these indexes and that are not captured by predictive densities. Note also that it is harder for ML predictives to achieve normality for the S&P 500 index and that tail PITs of the beta linear pool are often not normally distributed.

We reported in Subsection 4.2 that Bayesian predictives integrating parameter uncertainty improve 1% VaR forecasts for single models. We found here that 1% VaR forecasts from Bayesian predictives are also better under linear and log pooling. Besides this, we observed that linear and log pools are preferred to single models at the 5% risk level. We also highlighted that they provide a more homogeneous VaR backtest than single models. This suggests that these methods are robust against their worst-performing components. Among all forecasting techniques, the better VaR forecaster tends to be the EW linear pool of Bayesian predictives. This result can be explained by the superiority of VaR forecasts derived from Bayesian predictives and by the success of the EW combination in the literature. We also remarked in our own experiments that the optimal weight estimates exhibit considerable uncertainty. Finally, focusing on the left-tail with COP does not provide the expected results.

7 | CONCLUSION

In this study, we use a set of GARCH-type models to assess the influence of parameter and model uncertainty on the left-tail of predictive densities and, especially, on VaR estimates. Our main findings can be summarized as follows. First, accounting for parameter uncertainty within the Bayesian framework improves the VaR backtest at the 1% risk level for single models as well as linear and log pools. It tends also to improve left-tail forecasts as indicated by tail PIT analyses. Second, accounting for model uncertainty via linear or log pooling produces robust VaR backtests. Moreover, these pooling methods present better VaR estimates than single models at the 5% risk level. Third, the EW linear pool of Bayesian predictive densities tends to be the best VaR forecaster among 42 forecasting techniques.

Regarding the relative performance of single models, the worst VaR estimates are undoubtedly provided by the EGARCH model, whereas the best ones tend to be produced by the GJR model with a preference for skew GED innovations. We also demonstrate that GAS models with skew Student and GED innovations are competitive GARCH-type models. Lastly, we propose a novel combination scheme—the beta linear pool with COP weights—that simultaneously calibrates and focuses on the left-tail. This method gives puzzling results in our applications. Its VaR forecasts can show correct unconditional coverage while generating dependent VaR violations. In further research, it would be interesting to refine this method and to use it in other applications. Another interesting extension would be to see how priors that amplify or moderate time-varying volatility affect left-tail forecasts incorporating parameter and/or model uncertainty.

ACKNOWLEDGMENTS

This paper was previously titled "Predicting market risk with combinations of GARCH-type models." The authors thank the Editor (Terence Mills), two referees, Kris Boudt, William Doehler, and Lennart Hoogerheide, as well as conference and seminar participants at the European Meeting of the Econometric Society 2016 (Geneva, Switzerland), at the International Symposium on Forecasting 2015 (Riverside, CA, USA) and at Laval University (Québec, Canada) for their valuable comments. The authors also thank Keven Bluteau and Yoann Racine for research assistance. David Ardia acknowledges FRQSC, IFM2 Montreal and IIF for financial support. Jeremy Kolly is grateful to the Swiss National Science Foundation for financial support (grant #158754) and to the Finance, Insurance and Real Estate Department at Laval University for its hospitality. Denis-Alexandre Trottier is thankful to the FRQSC for financial support. The authors are solely responsible for the views expressed in this paper and for any remaining errors or shortcomings.

REFERENCES

Ardia, D. (2008). *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications*. Berlin, Germany: Springer.

Ardia, D., & Kolly, J. (2016). Predicting market risk with density combination: An introduction. *Wilmott Journal*, *81*, 52–57.

Aussenegg, W., & Miazhynskaia, T. (2006). Uncertainty in Value-at-Risk estimates under parametric and nonparametric modeling. *Financial Markets and Portfolio Management*, *20*(3), 243–264.

Bassetti, F., Casarin, R., & Ravazzolo, F. (2015). Bayesian nonparametric calibration and combination of predictive distributions. (Working Paper), Venice, Italy: Ca′ Foscari University of Venice.

Berkowitz, J., Christoffersen, P. F., & Pelletier, D. (2011). Evaluating Value-at-Risk models with desk-level data. *Management Science*, *57*(12), 2213–2227.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section*, American Statistical Association, pp. 177–181.

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

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, *69*(3), 542–547.

Casarin, R., Mantoan, G., & Ravazzolo, F. (2016). Bayesian calibration of generalized pools of predictive distributions. *Econometrics*, *4*(1), 17.

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

Christoffersen, P. F., & Pelletier, D. (2004). Backtesting Value-at-Risk: A duration-based approach. *Journal of Financial Econometrics*, *2*(1), 84–108.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

Deschamps, P. J. (2012). Bayesian estimation of generalized hyperbolic skewed student GARCH models. *Computational Statistics and Data Analysis*, *56*(11), 3035–3054.

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253–263.

Diks, C., Panchenko, V., & van Dijk, D. (2011). Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, *163*(2), 215–230.

Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106.

Duffie, D., & Pan, J. (1997). An overview of Value at Risk. *Journal of Derivatives*, *4*(3), 7–49.

Eklund, J., & Karlsson, S. (2007). Forecast combination and model averaging using predictive measures. *Econometric Reviews*, *26*(2–4), 329–363.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive Value at Risk by regression quantiles. *Journal of Business and Economic Statistics*, *22*(4), 367–381.

Fernández, C., & Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, *93*(441), 359–371.

Francq, C., & Zakoian, J. M. (2010). *GARCH models: Structure, Statistical Inference and Financial Applications*. Chichester, UK: Wiley.

Gatarek, L. T., Hoogerheide, L. F., Hooning, K., & van Dijk, H. K. (2014). Censored posterior and predictive likelihood in Bayesian left-tail prediction for accurate Value at Risk estimation. (Working Paper), Amsterdam, Netherlands: Tinbergen Institute.

Genest, C., & Zidek, J. V. (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, *1*(1), 114–135.

Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230.

Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, *164*(1), 130–141.

Geweke, J., & Amisano, G. (2012). Prediction with misspecified models. *American Economic Review*, *102*(3), 482–486.

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

Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business and Economic Statistics*, *29*(3), 411–422.

Gneiting, T., & Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, *7*, 1747–1782.

Hall, S. G., & Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, *23*(1), 1–13.

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

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

Jorion, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk* (3rd ed.). New York, NY: McGraw-Hill.

Kapetanios, G., Mitchell, J., Price, S., & Fawcett, N. (2015). Generalised density forecast combinations. *Journal of Econometrics*, *188*(1), 150–165.

Kascha, C., & Ravazzolo, F. (2010). Combining inflation density forecasts. *Journal of Forecasting*, *29*(1–2), 231–250.

Krüger, F. (2014). Combining Density Forecasts under Various Scoring Rules: An Analysis of UK Inflation. (Working Paper), Heidelberg, Germany: Heidelberg Institute for Theoretical Studies.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, *3*(2), 73–84.

Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York, NY: Wiley.

Massacci, D. (2015). Predicting the distribution of stock returns: Model formulation, statistical evaluation, VaR analysis and economic significance. *Journal of Forecasting*, *34*(3), 191–208.

Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys*, *29*(1), 46–75.

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

Opschoor, A., van Dijk, D., & van der Wel, M. (2015). Combining density forecasts using censored likelihood scoring rules. (Working Paper), Amsterdam, Netherlands: Tinbergen Institute.

Pesaran, M. H., Schleicher, C., & Zaffaroni, P. (2009). Model averaging in risk management with an application to futures markets. *Journal of Empirical Finance*, *16*(2), 280–305.

Ranjan, R., & Gneiting, T. (2010). Combining probability forecasts. *Journal of the Royal Statistical Society, Series B*, *72*(1), 71–91.

Smith, J., & Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, *71*(3), 331–355.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.

Stone, M. (1961). The opinion pool. *Annals of Mathematical Statistics*, *32*(4), 1339–1342.

Tay, A. S., & Wallis, K. F. (2000). Density forecasting: A survey. *Journal of Forecasting*, *19*(4), 235–254.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, *22*(4), 1701–1728.

Trottier, D.-A., & Ardia, D. (2016). Moments of standardized Fernández–Steel skewed distributions: Applications to the estimation of GARCH-type models. *Finance Research Letters*, *18*, 311–316.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, *18*(5), 931–955.

Ziggel, D., Berens, T., Weiss, G. N. F., & Wied, D. (2014). A new set of improved Value-at-Risk backtests. *Journal of Banking and Finance*, *48*, 29–41.

How to cite this article: Ardia D, Kolly J, Trottier D-A. The impact of parameter and model uncertainty on market risk predictions from GARCH-type models. *Journal of Forecasting*. 2017;36:808–823. https://doi.org/10.1002/for.2472

APPENDIX

Some acronyms

TABLE A1 Definitions of model acronyms used in tables and figures for Bayesian predictives. Model acronyms for ML predictives are similar but start with "mle". Sometimes the type of predictives used is not specified in the acronym when it is clear from the context

bay.egarch.st	Skew Student EGARCH model (Bayesian predictives)
bay.egarch.sged	Skew GED EGARCH model (Bayesian predictives)
bay.gjr.st	Skew Student GJR model (Bayesian predictives)
bay.gjr.sged	Skew GED GJR model (Bayesian predictives)
bay.tgarch.st	Skew Student TGARCH model (Bayesian predictives)
bay.tgarch.sged	Skew GED TGARCH model (Bayesian predictives)
bay.gas.st	Skew Student GAS model (Bayesian predictives)
bay.gas.sged	Skew GED GAS model (Bayesian predictives)
bay.lin.ew	EW linear pool of Bayesian predictives
bay.lin.bma	Linear pool of Bayesian predictives with BMA weights
bay.lin.op	Linear pool of Bayesian predictives with OP weights
bay.lin.cop-0.15	Linear pool of Bayesian predictives with COP weights (15% censoring)
bay.lin.cop-0.25	Linear pool of Bayesian predictives with COP weights (25% censoring)
bay.log.ew	EW log pool of Bayesian predictives
bay.log.bma	Log pool of Bayesian predictives with BMA weights
bay.log.op	Log pool of Bayesian predictives with OP weights
bay.log.cop-0.15	Log pool of Bayesian predictives with COP weights (15% censoring)
bay.log.cop-0.25	Log pool of Bayesian predictives with COP weights (25% censoring)
bay.beta.op	Beta linear pool of Bayesian predictives with OP weights
bay.beta.cop-0.15	Beta linear pool of Bayesian predictives with COP weights (15% censoring)
bay.beta.cop-0.25	Beta linear pool of Bayesian predictives with COP weights (25% censoring)