Policy Research Working Paper 8165

Variance and Skewness in Density Predictions

A World GDP Growth Forecast Assessment

Fabian Mendez-Ramos

WORLD BANK GROUP
Development Research Group
Macroeconomics and Growth Team
August 2017

Policy Research Working Paper 8165

Abstract

The paper introduces a Bayesian cross-entropy forecast (BCEF) procedure to assess the variance and skewness in density forecasting. The methodology decomposes the variance and skewness of the predictive distribution accounting for the shares of selected risk factors. The method assigns probability distributions to baseline-projections of an economic or social random variable—for example, gross domestic product growth, inflation, population growth, poverty headcount, among others—combining ex-post and ex-ante market information. The generated asymmetric density forecasts use information derived from surveys on expectations and implied statistics of predictive models. The BCEF procedure is applied to produce world GDP growth forecasts for three-year horizons using information spanning the period of October 2005–August 2015. The scores indicate that the BCEF density forecasts are more accurate and reliable than some naïve—symmetric and normal distributed confidence interval—predictions, illustrating the value-added of the introduced methodology.

This paper is a product of the Macroeconomics and Growth Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at fmendezramos@worldbank.org.

Variance and skewness in density predictions: A world GDP growth forecast assessment

Fabian Mendez-Ramos[1]

(Version: November 1, 2019)

JEL Classification: *C11, C53, C54*
Keywords: *Non-symmetric density forecasts, forecast uncertainty, Bayesian cross-entropy, scoring rules, the continuous ranked probability score, world GDP growth, implied volatility, consensus forecasts, the balance of risks, variance decomposition, skewness decomposition.*

I am thankful to Israel Osorio Rodarte, Raul Ibarra Ramírez, Aqib Aslam, Norman Loayza, colleagues from the Development Prospects Group: Delfin S. Go, Ayhan Kose, Franziska L. Ohnsorge and participants of the DECRG Seminar Series in Kuala Lumpur for helpful comments, inputs, and suggestions at various stages of the study. I am also thankful to Dr Michael McCracken and to several anonymous reviewers for helpful and constructive comments in previous versions of the paper. Contact address: Fabian Mendez Ramos, Development Research Group, The World Bank. Email: fmendezramos@worldbank.org.

"... if one model outforecasts another, the poorer of the two cannot be sensibly used for policy purposes".

Granger (1986).

1. Introduction

The relationship between informed decisions and knowledge about foreseen risks remains a central theme in the areas of policymaking and investment. The lack of suitable predictive models and basic rules of thumb leaves policy-makers and investors make decisions without proper knowledge of upside and downside risk dimensions associated with a random phenomenon. Especially in economics, the construction of appropriate predictive models should be a binding task before the generation of a policy or investment assessment. If, in terms of central measures and confidence intervals, a model is not able to predict trustworthy outcomes, the model itself is not useful for policy benchmarking as its predictions are misleading.

In this paper, we introduce a Bayesian cross-entropy forecasting (BCEF) procedure to generate density forecasts [2] solving an information-theoretical problem under noisy and scattered information. Our method combines ex-post and *ex-ante* knowledge of the variable of interest and risk factors that affect it. The empirical application generates density forecasts of world GDP growth associating selected risk factors: terms spreads, absolute deviations of inflation targets, energy prices, and the S&P 500 index prices. The BCEF method is Bayesian in the sense that allows the use of prior probability distributions to initially assume the structure of parameters that affect the distribution of our forecasted variable.

There are four key advantages of using the BCEF procedure to generate density forecasts. First, prior information of underlying unknown and unobserved parameters—or latent variables such as error terms—can be incorporated into a defined system to find less noisy levels of information in the noisy parameters. Second, the Bayesian entropy procedure recovers and processes information when the underlying sampling model is partly or incorrectly known, and

[2] From here onwards, density forecast, probabilistic forecast, or fan chart are used interchangeably.

the data is limited, partial, or incomplete (Golan, Judge, and Miller 1996). Third, the BCEF procedure complements already established forecasting econometric methods and assesses the uncertainty of a point forecast—most likely outcome—quantifying the magnitude of the variance and skewness of the confidence interval.[3] Fourth, the BCEF method also quantifies the risk-factor shares affecting the assessed variance and skewness of the density forecast. In this regard, the introduced procedure includes decompositions of *ex-ante* variance and skewness [4] of the predicted random variable. Both the variance and the skewness decompositions highlight the importance of *ex-ante* risk factor information in shaping the density forecasts.

The BCEF method is an information-theoretical procedure that constructs less information-noisy density forecasts given three significant restrictions: the constraint of a parametric functional form (of the density forecast), the consistency and behavioral constraints, and the selected prior information. Our BCEF procedure combines the minimum cross-entropy method (Golan, Judge, and Robinson 1994, Golan 2017) with *ex-ante* available data to predict future outcomes in the form of density forecasts. Thus, the introduced methodology adopts the Kullback-Leibler—or relative- or cross-entropy—function. By construction, the method relies on a principle that optimizes the error distribution of the variables of interest: minimization of the noise between the prior and the posterior distribution of the error terms. The solution of the BCEF method provides a posterior distribution of error terms assigned to parameters that describe the distribution of a study variable: world GDP growth in our empirical exercise. These posterior distributions are the closest to the prior distribution of the error components constrained in satisfying all the consistency and behavioral conditions of the model. As a minimum cross-entropy procedure, the BCEF procedure is

[3] Although in the BCEF procedure the point forecasts are taken as given, the accuracy of the density forecasts are still meaningful as they are evaluated in ex-post manner using proper scoring rules. The evaluation of the density forecasts via scoring rules gives the incentives to elicit accurate predictive distributions. The variance and skewness of the forecasted confidence intervals play an important role in the final ex-post score. See Appendix 1 for more details about the incentive mechanisms to release honest density forecasts behind proper scoring rules.

[4] We use the terms *ex-ante variance* and *ex-ante skewness* to denote a conditional variance and skewness given information that considers both *ex-ante* and ex-post observations.

optimal in the sense that it minimizes information losses in moving from a prior distribution assumption to a posterior distribution estimate.[5]

Whereas forecasters disagree in assigning prior distributions based on their judgment and personal information, these disagreements are in part a reflection of underlying uncertainty (Al-Najjar and Shmaya 2015). The BCEF procedure is flexible to exploit heterogeneity in beliefs embedded in assumed prior probability distributions of the parameters defining the study variable. We can get BCEF estimates under an almost infinite number of prior distributions, which, in most cases, the results will converge to similar posterior estimates.[6]

A critical characteristic of our BCEF method is that it uses *ex-ante* information to generate density forecasts. The used *ex-ante* information is extracted from surveys on expectations and implied statistics of forecasting models. While the use of *ex-ante* information to predict uncertainty in future outcomes is becoming more common, there is still some disagreement regarding its predictive power. Rich and Tracy (2010) build up uncertainty statistics to analyze inflation forecasts from the Survey of Professional Forecasters (SPF). Their results found little evidence that disagreement among forecasters is a good proxy for inflation uncertainty. Using the same SPF data, in contrast, Giordani and Söderlind (2003) find that difference among forecasters is a good measure for inflation and output growth uncertainties. More recent literature uses *ex-ante* forecasts errors based on SPF sources to improve the accuracy of uncertainty estimates around macroeconomic forecasts (Clark, McCracken, and Mertens 2018).

The empirical application of the introduced BCEF method generates monthly density forecasts of world GDP growth based on ex-post information spanning the period of October 2005–August 2015. The BCFE density forecasts for world GDP growth assess three annual horizons. The accuracy of the BCEF density forecasts is evaluated ex-post using the continuous

[5] In the language of Bayesian inference, the BCEF procedure capture a measure of the information gained when one revises one's beliefs from the prior probability distribution to the posterior probability distribution.

[6] Using incorrect priors will result in solutions that may be far from the underlying truth. In statistical terms, the solution will be biased (Golan 2017, p. 314).

ranked probability score (CRPS). Besides, the BCEF forecasts are compared with "naïve" density forecasts. This naïve forecasts are time-inexpensive and only require the standard deviation of historical forecast errors of world GDP growth predictions and a normality distribution assumption. Unweighted and weighted CRPS scores show that the BCEF fan charts are more accurate and reliable than the "naïve" density forecasts. In terms of forecast accuracy and reliability, the results show there is an incentive to invest in a method like the BCEF presented in this paper.

Furthermore, by construction, the BCEF results include variance and skewness decompositions of the density forecasts. For instance, the August 2015 density forecast indicates that around 68 percent of the world GDP growth uncertainty—measured via the variance of the predictive distribution—is directly explained with the four selected risk factors, whereas "other" risk factor components explain the rest of this predictive variance. In terms of the skewness decomposition, crude oil prices and "other" risk factor components are the elements that contribute more to the direction and magnitude of the world GDP growth predictive skewness across the three forecast horizons.

We describe the BCEF procedure in the following sections of the paper. Econometric considerations, unobserved parameters, distributional assumptions, prior information, and the BCEF minimization problem is explained in Section 2. The BCEF method is tested in Section 3, producing density forecasts to predict world GDP growth. The practical exercise targets world GDP growth, which is a variable that faces—*ex-ante* and ex-post—concerns of being incorrectly known, incorrectly quantified, or both. Risk factors affecting world GDP growth, prior values, the estimated density forecast series, and their evaluation scores are also discussed in Section 3. Finally, the main conclusions of this study are presented in Section 4.

2. The Model

The introduced BCEF technique allows for the quantification of risks in the prediction of a studied variable. The BCEF procedure can be applied to predict economic variables; however, it is also suitable in other fields where unobserved information prevails. Examples of studied or target variables are GDP growth, poverty headcount rates, population growth, consumer price inflation, stock market prices or returns, consumption or investment levels, among others. The generation of BCEF density forecasts uses two categories of constraints: behavioral and statistical ones. The behavioral restrictions rely on assumptions about the statistical shape and parameter performance of the predictive density forecasts. The statistical—or consistency—constraints rely on formal probability statements that govern random variables.

2.1 The linear assumption

The BCEF procedure is a non-linear inversion procedure. It provides a basis for information recovery while helping to make conservative inferences about an unknown and unobservable number, vector, or function (Golan, Judge, and Miller 1996). In the case of the estimation of predictive density functions, the target variable—and its statistics—might be unobserved and indeed not accessible for direct measurement. The assembled system incorporates not only ex-post information but also *ex-ante* knowledge about dispersion and skewness measures of associated risk factors of the target random variable.

The forecasted variable, Y, is assumed to be a function of a subset of risk factors, $\boldsymbol{Z} \subseteq \boldsymbol{\Theta}$. The selection of the subset of risk factors, $\boldsymbol{Z}$**,** might depend on empirical results and theoretical underpinnings associated with Y. Note that $\boldsymbol{\Theta}$ denotes the complete set of risk factors; $\boldsymbol{\Theta} = \{\boldsymbol{Z}, \boldsymbol{Z}^c\}$, where $\boldsymbol{Z}^c$ states for the complement of $\boldsymbol{Z}$. The forecasted dispersion and skewness measures of Y are assumed to depend on—ex-post and *ex-ante*—information of $\boldsymbol{Z}$. The basic idea is that markets perform not only using ex-post information, but they also behave incorporating *ex-ante* information once it is made available. The uncertainty of Y is approximated using dispersion and skewness statistics. Both statistics are modeled in separate forms to support their specific constraints. For instance, while the variance is always a positive number, the third central moment statistic can take negative and positive values.

The BCEF procedure looks to exploit the potential uneven and noisy trajectories of $\boldsymbol{Z}$ and Y using a minimization information-theoretical setup. The combination of the ex-post and *ex-ante* information of $\boldsymbol{Z}$ and the ex-post information of Y are critical inputs in the BCEF procedure, which estimates *ex-ante* uncertainty on Y. The *ex-ante* information of $\boldsymbol{Z}$**,** compiled from a variety of theoretical models and different survey sources, is critical too—and complements—the BCEF procedure in estimating the *ex-ante* uncertainty on Y.

Let us denote the set of risk factors indices by $\mathscr{R} = \{r_1, r_2, \ldots, R | r_1 \in \mathbb{N}_{\geq 1}, \ldots, R \in \mathbb{N}_{\geq 1}, r_1 < r_2 < \cdots < R\}$. The predicted variable, Y, is assumed to be a function of R number of risk factors: $Y = f(\boldsymbol{Z})$; $\boldsymbol{Z} = \left(Z_{r_1} \;\; Z_{r_2} \ldots Z_{r_R}\right) \;=\; (Z_1 \;\; Z_2 \ldots Z_R)$. For econometric convenience, the behavioral assumption that links Y to the exogenous risk factors, $\boldsymbol{Z}$**,** is linear, as stated in Equation (1). This Equation restricts the reactions of the variable Y to risk factor movements.

(1) $\qquad Y_t = a_1 Z_{1,t} + \cdots + a_R Z_{N_R,t} + \varepsilon_t.$

To describe ex-post and *ex-ante* categories of information, let us denote N_T as the current period. The three main types of information in the BCEF procedure are i) ex-post information of the target variable, $\{Y_t\}_{t=t_1}^{T}$, ii) ex-post information of the associated risk factors, $\{\boldsymbol{Z}_t\}_{t=t_1}^{T}$, and iii) *ex-ante* available information of the risk factors, $\left\{\boldsymbol{Z}_{t+h}\right\}_{t=t_1}^{T}$, for some horizons $h \in \mathscr{H}$, where $\mathscr{H}$ is a strictly monotonically increasing sequence, $\mathscr{H} = \left\{h_1, h_2, \ldots, H | h_1 \in \mathbb{N}_{\geq 0}, \ldots, H \in \mathbb{N}_{\geq 0}, h_1 < h_2 < \cdots < H\right\}$.

The process of the predicted variable, Y, is defined on a complete probability space, $(\Omega, \mathscr{F}, P)$, which is filtered using an increasing family of σ-fields in discrete-time over the sequence $\mathscr{T} = \left\{t_1, t_2, \ldots, T | t_1 \in \mathbb{N}_{\geq 1}, \ldots, T \in \mathbb{N}_{\geq 1}, t_1 < t_2 < \cdots < T\right\}$, i.e., the probability space $\left(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 1}, P\right)$. Moreover, at any period t, the information set $\mathscr{F}_t$ includes all the relevant information of the target variable, Y_t, and the associated risk factors, $\boldsymbol{Z}_t$. This assumption implies that the information set $\mathscr{F}_t$ comprises all the available information, both ex-post and *ex-ante*.

The essential advantage of the BCEF method involves the combination of ex-post data of Y with all the available *ex-ante* and ex-post information of $\boldsymbol{Z}$ at any period $t \in \mathscr{T}$ and for any available horizons, $h \in \mathscr{H}$. The information set $\mathscr{F}_T$ contains—at least—the following three series: $\{Y_t\}_{t=t_1}^{T}$, $\{\boldsymbol{Z}_t\}_{t=t_1}^{T}$, and $\left\{\boldsymbol{Z}_{t+h}\right\}_{t=t_1}^{T}$, for any horizon $h \in \mathscr{H}$. Note that $\left\{\boldsymbol{Z}_{r,t+h}\right\}_{t=t_1}^{T}$ denotes the *ex-ante* series of the risk factor $Z_r \in \boldsymbol{Z}$, for any horizon $h \in \mathscr{H}$.

Besides, the predictive variance and skewness for the risk factors are denoted by $var\left(\boldsymbol{Z}_{r,t+h}\right)$ and $\gamma\left(\boldsymbol{Z}_{r,t+h}\right)$ for any period $t \in \mathscr{T}$, and horizon $h \in \mathscr{H}$. The conditional variance and skewness of Y given all available information at any period t and for any horizon h are represented by $var\left(Y_{t+h}|\mathscr{F}_t\right)$ and $\gamma\left(Y_{t+h}|\mathscr{F}_t\right)$, respectively. Note that $var\left(\boldsymbol{Z}_{r,t+h}\right)$ and $\gamma\left(\boldsymbol{Z}_{r,t+h}\right)$ are not conditional on $\mathscr{F}_t$ because we want to make explicit that our statistics of interest, $var\left(Y_{t+h}|\mathscr{F}_t\right)$ and $\gamma\left(Y_{t+h}|\mathscr{F}_t\right)$, rigorously include all available information: ex-post and *ex-ante*.

Note that the BCEF procedure does not focus on predicting $\mathbb{E}\left(Y_{t+h}|\mathscr{F}_t\right)$ or $mode\left(Y_{t+h}|\mathscr{F}_t\right)$; these are taken as given. We argue that many models and experts do a good job predicting central measures of economic variables under structural assumptions and statistical constraints. The BCEF method focuses hence on predicting the conditional variance and skewness given all available information at specific period t, $var\left(Y_{t+h}|\mathscr{F}_t\right)$ and $\gamma\left(Y_{t+h}|\mathscr{F}_t\right)$, respectively. The BCEF method complements and polishes central measure predictions as the $mode\left(Y_{t+h}|\mathscr{F}_t\right)$ adding information of the uncertainty around the point forecasts via the predicted variance and skewness, $var\left(Y_{t+h}|\mathscr{F}_t\right)$ and $\gamma\left(Y_{t+h}|\mathscr{F}_t\right)$. These forward-looking conditional variance and skewness are econometrically approximated as presented in Proposition 1 and Proposition 2. These two propositions also denote decompositions of the forward-looking variance and skewness of the study variable Y.

PROPOSITION 1: Under the linear assumption in (1), the predicted conditional variance of the study variable is a linear combination of the variance of the risk factors plus an error term, as defined in (2).

$$(2) \qquad \underbrace{var(Y_{t+h}|\mathcal{F}_t)}_{\substack{conditional\\ variance\ of\\ the\ study\\ variable}} = \sum_{r\subseteq R} {a_{r,h}}^2 \underbrace{var(Z_{r,t+h})}_{\substack{variance\\ of\\ predicted\\ risk-factors}} + \underbrace{\epsilon_{\sigma^2,t+h}}_{\substack{variance\\ components\\ from\\ cross-moments\\ and\ other\\ risk-factors}}{}^2 \quad , \qquad \forall t \in \mathcal{T}, h \in \mathcal{H} \ .$$

PROPOSITION 2: Under the linear assumption in (1), the predicted conditional skewness of the study variable is a linear combination of the risk factor skewness plus an error term as defined in (3).

$$(3) \qquad \underbrace{\gamma(Y_{t+h}|\mathcal{F}_t)}_{\substack{conditional\\ skewness\\ of\ the\\ study\\ variable}} = \sum_{r\subseteq R} {a_{r,h}}^3 \underbrace{\gamma(Z_{r,t+h})}_{\substack{skewness\\ of\\ predicted\\ risk-factors}} + \underbrace{\epsilon_{\gamma,t+h}}_{\substack{skewness\\ components\\ from\ cross-\\ moments\ and\\ other\ risk-factors}}{}^3 \quad , \qquad \forall t \in \mathcal{T}, h \in \mathcal{H} \ .$$

Note that $\gamma(\bullet)$ denotes the skewness operator. The predicted conditional variance and skewness of Y are a weighted average of the conditional dispersion and skewness statistics of the accounted risk factors. The error terms $\epsilon_{\sigma^2,t+h}$ and $\epsilon_{\gamma,t+h}$ account for cross expected moments and "other" risk factor components. The proof of Proposition 1 and Proposition 2 can be seen in Appendix A1 and A2, respectively.

Besides, note that the same coefficients $\{a_1, a_2, \ldots, a_R\}$ that come from the econometric assumption in (1) constrain the estimation of Equation (2) and (3). Thus, for consistency purposes, (2) and (3) must be estimated simultaneously. As presented in Subsection 2.5, the BCEF procedure is a constrained non-linear programming model. The simultaneous recovery of the coefficients in the set A restricts the number of values that the parameters can take. While the coefficients in A estimate the direct effects on the statistics of the forecasted variables, the cross-effects of the risk factors on the target variable, Y, are also captured and added up in the error terms, ${\epsilon_\sigma}^2$ and ${\epsilon_\gamma}^3$. For analytical purposes, the coefficients in Equations (2) and (3) should be read as first-order effects of the risk statistics over the forecasted variable, Y.

2.2 Latent statistics

The BCEF procedure looks to predict the variance and skewness of Y for period $T + h$, for any horizon $h \in \mathscr{H}$.[7] However, the information of these statistics is not observable or uncertain, i.e., the forecaster does not have data on the series $\left\{var\left(Y_{t+h}|\mathscr{F}_t\right)\right\}_{t=t_1}^{T}$ and $\left\{\gamma\left(Y_{t+h}|\mathscr{F}_t\right)\right\}_{t=t_1}^{T}$. In our empirical exercise, the series of predicted world GDP growth variance and skewness are not available. This lack of information makes complicated the econometric estimation of (2) and (3). While simultaneously estimating (2) and (3), the BCEF method recovers the i) the set of risk-share coefficients $A = \{a_1, a_2, \ldots, a_R\}$; ii) the cross-moment effects and "other" risk-factor components affecting the predictive variance and skewness, $\left\{\epsilon_{\sigma^2,t+h}\right\}_{t=t_1}^{T}$ and $\left\{\epsilon_{\gamma,t+h}\right\}_{t=t_1}^{T}$, and iii) the conditional variance and skewness of the study variable, $\left\{var\left(Y_{t+h}|\mathscr{F}_t\right)\right\}_{t=t_1}^{T}$ and $\left\{\gamma\left(Y_{t+h}|\mathscr{F}_t\right)\right\}_{t=t_1}^{T}$.

2.3 Distributional assumptions

Even though several parametric distributions are available to model density forecast, we selected a distribution that mainly facilitates the estimation of variance and skewness statistics as we are interested in studying nonsymmetric behaviors. Thus, to generate the predictive probability density function of the study variable, Y, we assume the two-piece normal (TPN) function.[8] The variable Y is predicted using three statistics: a central measure like the mode, a dispersion measure like the variance, and a skewness measure like the third central moment statistic. The TPN assumption (Wallis 2014) is flexible enough to allow skewed distributional shapes of the predicted random variable. In some cases, the TPN assumption can be a restriction. This distributional constraint can be especially stringent in variables with very high or shallow levels of excess kurtosis, i.e., leptokurtic and platykurtic shapes, respectively.

[7] Note that the actual or most recent period is denoted with T.

[8] We acknowledge that the distributional assumption of the TPN distribution is restrictive. For instance, bi-modal distributions are excluded in this assumption. However, we consider that the non-symmetric option that the TPN distribution brings is still of much value for density forecasting purposes and forward-looking uncertainty analysis.

Under the TPN assumption, the shape of the distribution of the forecasted variable, Y, can be defined via the mode, a dispersion, and a skewness statistic.[9] The functional form and further details of the TPN distribution are presented in Appendix A5. In the BCEF method, we use the following two equations to denote the TPN distribution of the density forecast.

(4) $$var\left(Y_{t+h}|\mathcal{F}_t\right) = \left(\tfrac{\pi-2}{\pi}\right)\left(\sigma_{2,Y,t+h} - \sigma_{1,\mathrm{Y},t+h}\right)^2 + \sigma_{1,\mathrm{Y},t+h}\,\sigma_{2,\mathrm{Y},t+h}, \qquad \forall t \in \mathcal{T}, h \in \mathcal{H}.$$

(5) $$\gamma\left(Y_{t+h}|\mathcal{F}_t\right)\sqrt{\tfrac{\pi}{2}}\left(\sigma_{2,Y,t+h} - \sigma_{1,\mathrm{Y},t+h}\right)^{-1} = \left(\tfrac{4-\pi}{\pi}\right)\left(\sigma_{2,Y,t+h} - \sigma_{1,\mathrm{Y},t+h}\right)^2 + \sigma_{1,\mathrm{Y},t+h}\sigma_{2,\mathrm{Y},t+h}, \qquad \forall t \in \mathcal{T}, h \in \mathcal{H}.$$

Ex-ante dispersion and skewness information are inputs to identify the forecast uncertainty dimensions of Y. This ex-ante information is complemented with an already elicited forecast mode—or most likely outcome of Y—and the TPN distributional assumption. The magnitude of the *ex-ante variance* that the forecaster elicits, $var\left(Y_{t+h}|\mathcal{F}_t\right)$, reflects his level of knowledge and confidence in the likelihood around central expected outcomes. The *ex-ante skewness* of the predicted variable, $\gamma\left(Y_{t+h}|\mathcal{F}_t\right)$, summarizes the forecaster's knowledge and beliefs that the most likely outcome would materialize on one or the other side of the distribution.

There are two complementary angles to the usefulness of eliciting dispersion and skewness measures: technical and policy-oriented. On the technical side, the assessment of dispersion and skewness statistics is central to accurately quantify the upside and downside uncertainties of the most likely outcome of the predicted variable Y. The technical side also allows the evaluation of uncertainty in recursive form and the testing of the predictive power of the elicited density forecasts. From the policy analysis angle, the predictive dispersion and skewness statistics provide the forecaster's sentiment around the most likely outcome. In policy analysis, these two statistics offer a fundamental understanding of the future behavior of Y, complementing the decision-making process under uncertainty.

[9] Even though the study variable Y is assumed to follow a TPN distribution, its density forecast cannot be estimated via maximum likelihood as we only collect few elicited central-measure predictions of Y.

2.4 Prior information

The econometric steps for estimating Equations (2) and (3) deal with two main restrictions. First, the *ex-ante variance* and *ex-ante skewness* of the predicted variable are unobserved or uncertain. Second, there might be empirical evidence, previous literature, or beliefs affecting the sign of the coefficients in Equations (2) and (3). To deal with these two significant constraints, the BCEF procedure solves a minimization information-theoretical setup where prior information is integrated into the setting of the problem. In specific, our BCEF method follows a minimum cross-entropy setting, as presented in Golan, Judge, and Miller (1996), and Judge and Mittelhammer (2011).

The BCEF procedure incorporates prior information related to $\boldsymbol{Z}$ and Y. The use of prior information is an implication of the principles that one should fully specify to make probabilistic statements of what is unknown (Geweke and Whiteman 2006). The subjective probability distribution that a decision-maker assesses upon a variable should embody a synthesis of all his or her information relevant to its outcome (Bunn 1984, chapter seven). In this regard, we incorporate prior sampling information of the underlying phenomena affecting Y as well as the perceptions of the forecaster. Note that this personal information about the distribution of a phenomenon can be symmetric or skewed, although in any case, the prior assessment should always be coherent (De Finetti 1937).

The BCEF method needs three steps to set prior information. Each of these three steps corresponds to three classes of unobserved parameters, which are components of Equations(2)–(3). The three categories of unobserved parameters are risk factor coefficients, error terms of the variance and skewness equations, and the predicted variance and skewness statistics. Furthermore, we assume that all the unobserved parameters behave in linear form including additive error terms for all periods $t \in \mathscr{T}$ and horizons $h \in \mathscr{H}$ as presented in Equations (6)–(10).

Risk-factor coefficients:

(6) $$a_r = a_r^0 + \xi_{a_r}.$$

Error terms on behavioral equations:

(7) $\epsilon_{\sigma^2} = {\epsilon_{\sigma^2}}^0 + \xi_{\epsilon_{\sigma^2}}$.

(8) $\epsilon_\gamma = {\epsilon_\gamma}^0 + \xi_{\epsilon_\gamma}$.

Variance and skewness statistics:

(9) $var\left(Y_{t+h}|\mathcal{F}_t\right) = var^0\left(Y_{t+h}|\mathcal{F}_t\right) + \xi_{\sigma^2,t+h}$.

(10) $\gamma\left(Y_{t+h}|\mathcal{F}_t\right) = \gamma^0\left(Y_{t+h}|\mathcal{F}_t\right) + \xi_{\gamma,t+h}$.

In the BCEF method, the set $\beta = \left\{a_r, \epsilon_{\sigma^2}, \epsilon_\gamma, var\left(Y_{t+h}|\mathcal{F}_t\right), \gamma\left(Y_{t+h}|\mathcal{F}_t\right)\right\}$ denotes posterior parameters. The prior values are represented in the set $\beta^0 = \left\{a_r^0, {\epsilon_{\sigma^2}}^0, {\epsilon_\gamma}^0, var^0\left(Y_{t+h}|\mathcal{F}_t\right), \gamma^0\left(Y_{t+h}|\mathcal{F}_t\right)\right\}$. The set of error terms of these linear specifications in Equations (6)–(10) is depicted by $\Xi = \left\{\xi_{a_r}, \xi_{\epsilon_{\sigma^2}}, \xi_{\epsilon_\gamma}, \xi_{\sigma^2,t+h}, \xi_{\gamma,t+h}\right\}$. All the elements in Ξ are mean values with a corresponding prior probability set $\pi(\Xi) = \left\{\pi\left(\xi_{a_r}\right), \pi\left(\xi_{\epsilon_{\sigma^2}}\right), \pi\left(\xi_{\epsilon_\gamma}\right), \pi\left(\xi_{\sigma^2,t+h}\right), \pi\left(\xi_{\gamma,t+h}\right)\right\}$. Furthermore, the posterior probability set of the error terms in Equations (6)–(10) is denoted by $\pi(\Xi|\mathcal{F}_T) = \left\{\pi\left(\xi_{a_r}|\mathcal{F}_T\right), \pi\left(\xi_{\epsilon_{\sigma^2}}|\mathcal{F}_T\right), \pi\left(\xi_{\epsilon_\gamma}|\mathcal{F}_T\right), \pi\left(\xi_{\sigma^2,t+h}|\mathcal{F}_T\right), \pi\left(\xi_{\gamma,t+h}|\mathcal{F}_T\right)\right\}$.

The first step is related to the risk-factor coefficients. We do the setting of the prior mean values for each of the risk factor coefficients needed in the estimation of Equations (2) and (3). The prior mean values of these risk-factor coefficients are denoted by the set A^0, such that $A^0 = \left\{{a_1}^0, {a_2}^0, \ldots, {a_R}^0\right\}$.

The prior mean values in A^0 are assigned by the forecaster based on empirical pieces of evidence, previous literature, or beliefs, among other procedures. Strong beliefs or empirical evidence affect and constraint the sign of the posterior coefficients in A. For instance, if the prior evidence or theory suggests that the risk factor Z_r is positively associated with the target variable, Y, then, the coefficient a_r is expected to be positive. Thus, implying that ${a_r}^0$ should have positive values as well. This positive sign consideration also has implications on the values that the error term ξ_{a_r} can take. For instance, to accomplish the restriction ${a_r}^0 + \xi_{a_r} > 0$, the mean error term, ξ_{a_r}, is constrained into the open space $\left(-\left|{a_r}^0\right|, \infty^+\right)$, i.e., $\xi_{a_r} \in \left(-\left|{a_r}^0\right|, \infty^+\right)$, for all selected prior ${a_r}^0 \in \mathbb{R}^1$.

The second step corresponds to the error terms in the behavioral equations. We set the priors of the error terms in Equations(2)–(3) in this step. We denote these prior mean values by $\left\{{\epsilon_{\sigma,t+h}}^0\right\}_{t=t_1}^T$ and $\left\{{\epsilon_{\gamma,t+h}}^0\right\}_{t=t_1}^T$, for all horizons $h \in \mathscr{H}$. As a general rule, we suggest that the mean error terms, ${\epsilon_{\sigma,t+h}}^0$ and ${\epsilon_{\gamma,t+h}}^0$, are set to zero—unless there is information that adjusts this assumption. Besides, in general, ϵ_{σ^2} , $\epsilon_\gamma \in \left(\infty^-, \infty^+\right)$ which implies that ${\epsilon_{\sigma^2}}^0, {\epsilon_\gamma}^0, \xi_{\epsilon_{\sigma^2}}, \xi_{\epsilon_\gamma}$ belong to the real numbers $\mathbb{R}^1$.

In the third step, we select the prior mean values of the conditional variance and skewness statistics across the period t_1–T, and for all horizons $h \in H$. These series are described by $\left\{var^0\left(Y_{t+h}|\mathscr{F}_t\right)\right\}_{t=t_1}^T$, and $\left\{\gamma^0\left(Y_{t+h}|\mathscr{F}_t\right)\right\}_{t=t_1}^T$. In Appendix A3, we show a constrained ordinary least squares (OLS) procedure that can be used as an alternative to proxy the prior mean values of the mentioned series. There are multiple methods, however, that the forecaster can use to get his or her priors (Golan 2017).

2.5 Bayesian cross-entropy estimation

The BCEF procedure consists of solving a minimization information-theoretical problem. The posterior mean components to be estimated from Equations (2) and (3) were already defined in the set β, while the corresponding priors and error terms of Equations (6)–(10) are denoted by the sets β^0 and ξ. The corresponding sets of prior and posterior probabilities of the ξ error terms are denoted by $\pi(\xi)$ and $\pi(\xi|\mathscr{F}_T)$.

For simplicity, we assume discrete and coherent prior and posterior probability distributions for each error term in the set Ξ, $\pi\left(\xi_j\right)$ and $\pi\left(\xi_j|\mathscr{F}_T\right)$, respectively. Each error term in the set Ξ are described using the support set $V_j = \left\{v_{1_j}, v_{2_j}, \ldots, v_{K_j} |\ K_j \subseteq \mathbb{N}_{\geq 3}\right\}$, $\forall j \in \Xi$. For symmetric purposes of these prior and posterior probability mass distributions, the integer K_j is constrained as an odd natural number. $K_j = \left\{2k_j + 1 \colon k_j \in \mathbb{Z}^+\right\}$, $\forall j \in \Xi$. Prior and posterior probability weight sets are defined as in Equations (11) and (12) for all elements j in Ξ .

(11) $\Pi_j^0 = \left\{ \pi_{1_j}^0, \dots, \pi_{K_j}^0 | \pi_{i_j}^0 \in [0,1], K_j \subseteq \mathbb{N}_{\geq 3} \right\}.$

(12) $\Pi_j = \left\{ \pi_{1_j}, \dots, \pi_{K_j} | \pi_{i_j} \in [0,1], K_j \subseteq \mathbb{N}_{\geq 3} \right\}.$

For coherence, the probability weights add to one, as stated in Equations (13) and (14).

(13) $\sum_{k_j=1}^{K_j} \pi_{k_j} = 1.$

(14) $\sum_{k_j=1}^{K_j} \pi_{k_j}^0 = 1.$

As previously stated, unless further evidence arises, the BCEF procedure assumes that the prior mean values of the error terms are zero; i.e., ${\xi_j}^0 = \sum_i {\pi_{i_j}}^0 \upsilon_{i_j} \equiv 0.$[10] The posterior mean of the error terms is defined in standard form as presented in Equation (15).

(15) $\xi_j = \sum_{k_j=1}^{K_j} \pi_{k_j} \nu_{k_j}.$

Note that both the error support V_j, the probability masses, Π_j^0 and Π_j, and the coherence conditions depict the prior and posterior discrete probability distributions, $\pi\left(\xi_j\right)$ and $\pi\left(\xi_j | \mathscr{F}_T\right)$ for each element j in Ξ.

To obtain the parameters of interest β, the BCEF procedure solves a minimization problem adopting the Kullback-Leibler (KL) objective function $\sum_{\forall j \in \Xi} D_{KL}\left(\pi\left(\xi_j | \mathscr{F}_T\right) \parallel \pi\left(\xi_j\right)\right)$ subject to some consistency and behavioral restrictions as presented in Equation (16).[11,12]

---

[10] Informative and non-informative symmetric prior distributions for the error terms ${\xi_j}^0$ are discussed in Go, Lofgren, Mendez-Ramos, and Robinson (2016).

[11] An example of the cross-entropy setting applied to social accounting matrices can be seen in (Golan, Judge, and Robinson 1994).

[12] In the language of Bayesian inference, $D_{KL}\left(\pi\left(\xi_j | \mathscr{F}_t\right) \parallel \pi\left(\xi_j\right)\right)$ is a measure of the information gain in moving from the prior distribution to the posterior distribution: $\pi\left(\xi_j\right) \rightarrow \pi\left(\xi_j | \mathscr{F}_t\right)$.

$$(16) \qquad \min_{\{\Pi_{\mathrm{j}},\forall j\in\Xi\}} \left\{ \sum_j \sum_i \pi\left(\xi_j = \upsilon_{i_j} \mid \bullet\right) \ln\left(\frac{\pi\left(\xi_j=\upsilon_{i_j}\mid\bullet\right)}{\pi\left(\xi_j=\upsilon_{i_j}\right)}\right)\right\} = \min_{\{\Pi_{\mathrm{j}},\forall j\in\Xi\}} \left\{ \sum_j \sum_i \pi_{i_j} \ln\left(\frac{\pi_{\mathrm{i_j}}}{\pi^0_{i_j}}\right)\right\}$$

Subject to behavioral and consistency constraints.

While the behavioral constraints—Equations (2)–(5)—correspond to assumptions regarding the shape of the density forecasts, the consistency restrictions are related to probability distribution properties embedded in the BCEF method as stated in Equations (6)–(15).

3. Prediction of world GDP growth

In this section, the BCEF procedure is applied to estimate density forecasts for world GDP growth. The use of density forecasts is an alternative to add valuable information to point estimates of predicted growth. Whereas economic growth density forecasts still provide central measures—the point forecasts itself, they can also help to explain second and third moments of the predictive density: upside and downside risk magnitudes. Besides, the implemented BCEF method reveals the sizes of those second and third moments by corresponding risk factor shares as presented in Subsection 2.1. The ex-post used information span the period of October 2005–August 2015, while the *ex-ante* information covers October 2005–August 2017. An evaluation of the generated BCEF predictions is implemented using the continuous ranked probability score.[13] Besides, we construct *naïve*—or rule of thumb—density predictions to compare the performance of the created BCEF forecasts.[14]

The use of national accounting systems started to become predominant across nations in the twentieth century. During this period, the prediction of economic growth quickly gained importance and surged. Given its importance in the current economic framework, often, predicting global and cross-country economic growth receives a lot of attention and resources. Despite this, history has shown that forecasters continually elicit point forecasts that are close

[13] See Appendix A1 for further details of proper scoring rules.

[14] Note that the *naïve* forecasts are symmetric by construction These naïve density forecasts follow a normal distribution with mean equal to the most likely outcome and standard deviation equal to the—sample—standard deviation of historical forecasts errors.

to the realization, but that fail to foresee the precise economic growth of a country. In recent years, evidence shows a wide gap in inaccuracies of a variety of forecasters in predicting the economic growth of several countries (Figure 1).

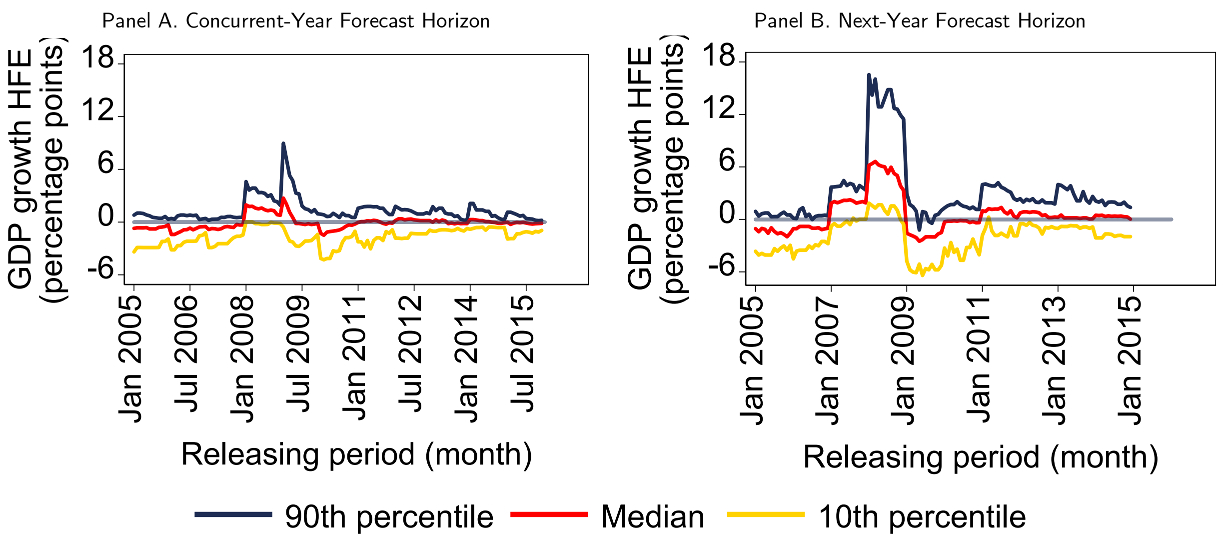

FIGURE 1. HISTORICAL FORECAST ERRORS OF REAL GDP GROWTH RATES FOR A VARIETY OF COUNTRIES

Source: Consensus Economics, Global Economic Prospects, and World Development Indicators.
Note: Monthly observations over January 2005–August 2015 period. Forty-one economies with real GDP growth forecasts are considered: Argentina, Australia, Bulgaria, Brazil, Canada, Switzerland, Chile, China, Colombia, Czech Republic, Germany, Euro area, Spain, Estonia, France, the United Kingdom, Hong Kong SAR-China, Croatia, Indonesia, India, Italy, Japan, the Republic of Korea, Lithuania, Latvia, Mexico, Malaysia, Netherlands, Norway, New Zealand, Peru, Poland, Romania, the Russian Federation, Singapore, Slovak Republic, Sweden, Thailand, Turkey, the United States, and Venezuela. Historical forecast error is defined as predicted minus realized value. A positive HFE value indicates overprediction. The next-year horizon historical forecast errors (Panel B) shows that the country growth point forecasts made during January 2007–December 2008 were mostly overconfident.

The most likely outcome—mode—of world GDP growth is taken from predictions made in a structural model that includes inputs from experts on country economic growth (World Bank 2016). However, note that the looking-forward conditional variance and skewness of world GDP growth are assumed not observed. The *ex-ante* information on GDP growth and its risk factors uses forward-looking information covering the scope of annual predictions spanning the period of 2005–2017. The BCEF method recovers predictive densities of global growth over the full monthly period of October 2005–August 2015 with three annual horizons:

the current year, next year, and two-year ahead horizons. For instance, the density forecasts estimated in August 2015 predicts world GDP growth for 2015, 2016, and 2017.

3.1 Risk factors and priors

We select a subset of risk factors, $\boldsymbol{Z}$, directly affecting world GDP growth. Four random variables help us to account for substantial risks of world GDP growth. These variables account for absolute deviations from consumer price inflation targets, global energy prices proxied by crude oil prices, global financial conditions quantified by the S&P 500 index price, and world term spread conditions measured by country-individual term spreads.

The density predictions for GDP growth are constructed reflecting upon basic economic mechanisms between output growth and underlying risks. The linear relationship of GDP growth, Y, and its associated risk factors, $\boldsymbol{Z}$, imply a constrained behavior of the variance and skewness as discussed in Proposition 1 and 2. In this application, the assumed responses of global growth to the associated risk factor movements are supported using previous studies.[15] The sign restrictions on the responses of World GDP growth to the changes of its global risk factors are presented in Table 1. The sign of the risk-coefficients in the linear Equation (1) defines the reactions of the skewness to predicted global growth, as seen in Equation (3). In this regard, the global growth skewness is marginally synchronized with the directional change of an associated risk factor.

Table 1 – Sign Restrictions of the Responses of World GDP to Movements of Global Conditions

	World Term Spreads	World (CPI) Inflation: Absolute Deviations from Targets	World Energy Prices	S&P 500 Index Prices
World GDP growth	(+)	(-)	(-)	(+)

15 Apart from the literature revision, we also estimated variety of econometric representations for ex-post world GDP growth using OLS and robust White-Huber standard errors. These OLS estimations obtain coefficients of the risk factors with the same signs as the ones presented in Table 1. The OLS estimates are available upon request.

The general financial conditions affecting global growth are represented using cross-country term spreads and the S&P 500 index prices. High-interest rate term spreads are assumed positively linked with output growth. This is in line with the following authors: Harvey (1989), Estrella and Hardouvelis (1991), Haubrich and Dombrosky (1996), Stock and Watson (2003), Wheelock and Wohar (2009), Aretz and Peel (2010), and Kao, Kuan, and Chen (2013), among others. The stock market performance measured by the S&P 500 index prices—and other equity indexes—is assumed to be positively associated with anticipated output growth; as per Mitchell and Burns (1938), Fama (1981), Fischer and Merton (1984), Harvey (1989), Jay Choi, Hauser, and Kopecky (1999), Hassapis and Kalyvitis (2002), and Tsouma (2009), among others.

Energy sector performance—measured via crude oil and gas prices—is assumed to be negatively associated with output growth. This association is in line with Hamilton (2005), Jimenez-Rodriguez and Sánchez (2005), Blanchard and Gali (2007), Blanchard and Riggi (2013), among others. Absolute deviations of CPI inflation targets are negatively associated with GDP growth performance. In this empirical application, the assumed benchmark inflation target is fixed at two percent across the study countries.

Based on data origination, the four selected risk factors affecting world GDP growth are divided into two categories: i) survey-based; and ii) market-based. The survey-based risk factors are a) the term spread, and b) the CPI inflation. The survey-based observations are collected by Consensus Economics. The market-based risk factor information is collected using Bloomberg. implied volatilities of option prices. This information is put together for the S&P 500 index, and the West Texas Intermediate (WTI) and Brent crude oils. As previously mentioned, elicited most likely forecasts of world GDP growth are obtained from World Bank (2016).

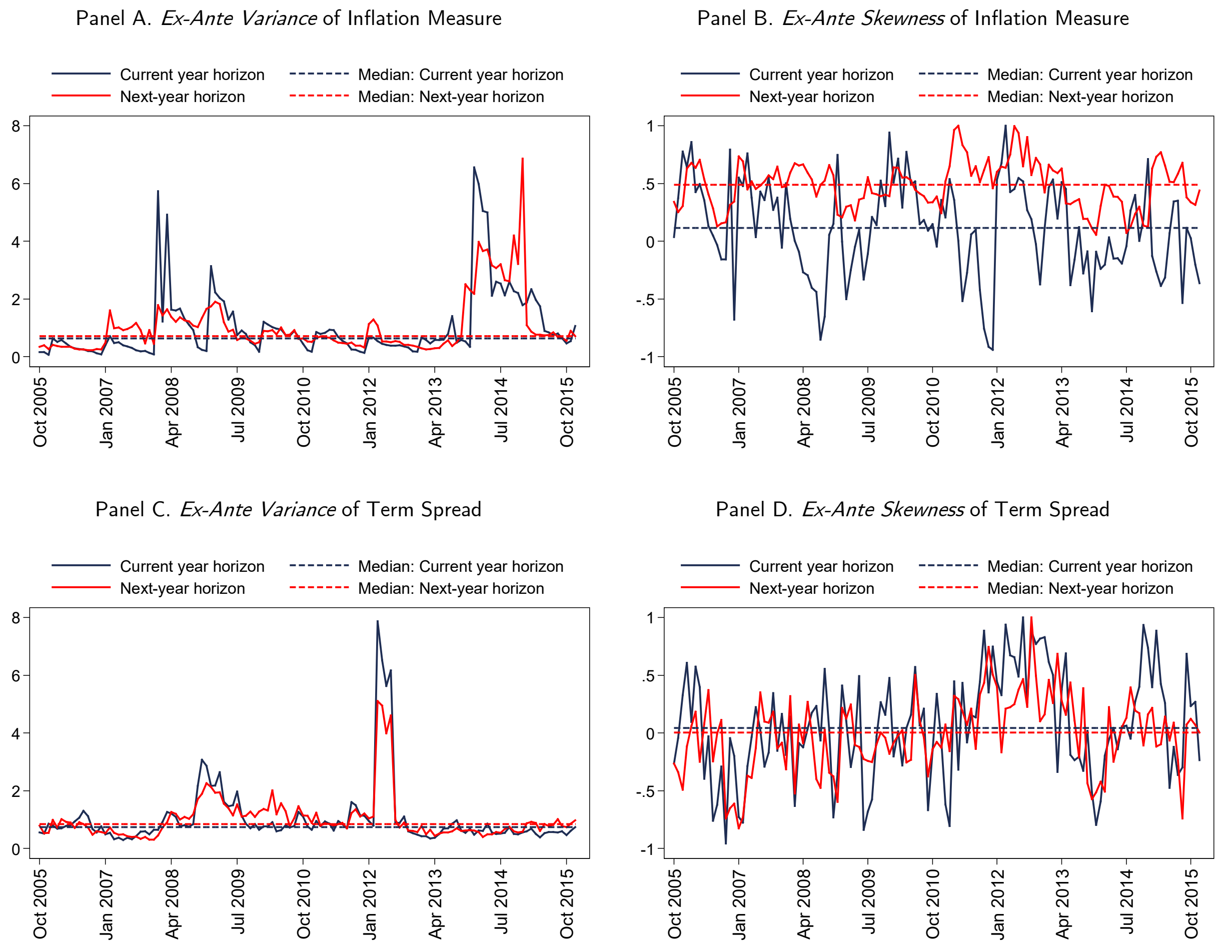

FIGURE 2. *FORWARD-LOOKING CONDITIONAL* VARIANCE AND SKEWNESS OF TERM SPREAD AND ABSOLUTE DEVIATIONS FROM INFLATION TARGETS

Source: Consensus Forecasts and author's calculations.

Note: The global term spread statistics are constructed using fifteen countries: 1) Australia; 2) Canada; 3) France; 4) Germany; 5) India; 6) Indonesia; 7) Italy; 8) Japan; 9) the Republic of Korea; 10) the Netherlands; 11) Spain; 12) Sweden; 13) Switzerland; 14) the United Kingdom; and 15) the United States. The global inflation statistics are derived using the twenty most significant economies in terms of GDP size: 1) Australia; 2) Brazil; 3) Canada; 4) China; 5) France; 6) Germany; 7) India; 8) Indonesia; 9) Italy; 10) Japan; 11) the Republic of Korea; 12) Mexico; 13) the Netherlands; 14) the Russian Federation; 15) Spain; 16) Sweden; 17) Switzerland; 18) Turkey; 19) the United Kingdom; and 20) the United States. The global variance series are scaled down using the sample standard deviation of the original series. The global skewness series is scaled down using the maximum value of the corresponding original series.

Ex-ante variance and skewness information of the term spread and absolute deviations of inflation targets from a variety of countries are used to proxy the corresponding global measures (Figure 2). Implied volatilities of options contracts are the source of information for estimating the *ex-ante variance* and *ex-ante skewness* of the S&P 500 and crude oil prices (Figure 3).

The selected implied volatilities come from option contracts with 3, 12, and 24-month maturities to represent the uncertainty for the current year, next year, and two-year ahead horizons. The procedure for recovering the variance and skewness of the equity and commodity prices from implied volatility series—assuming the Black-Scholes-Merton option pricing model—is discussed in Appendix A6.

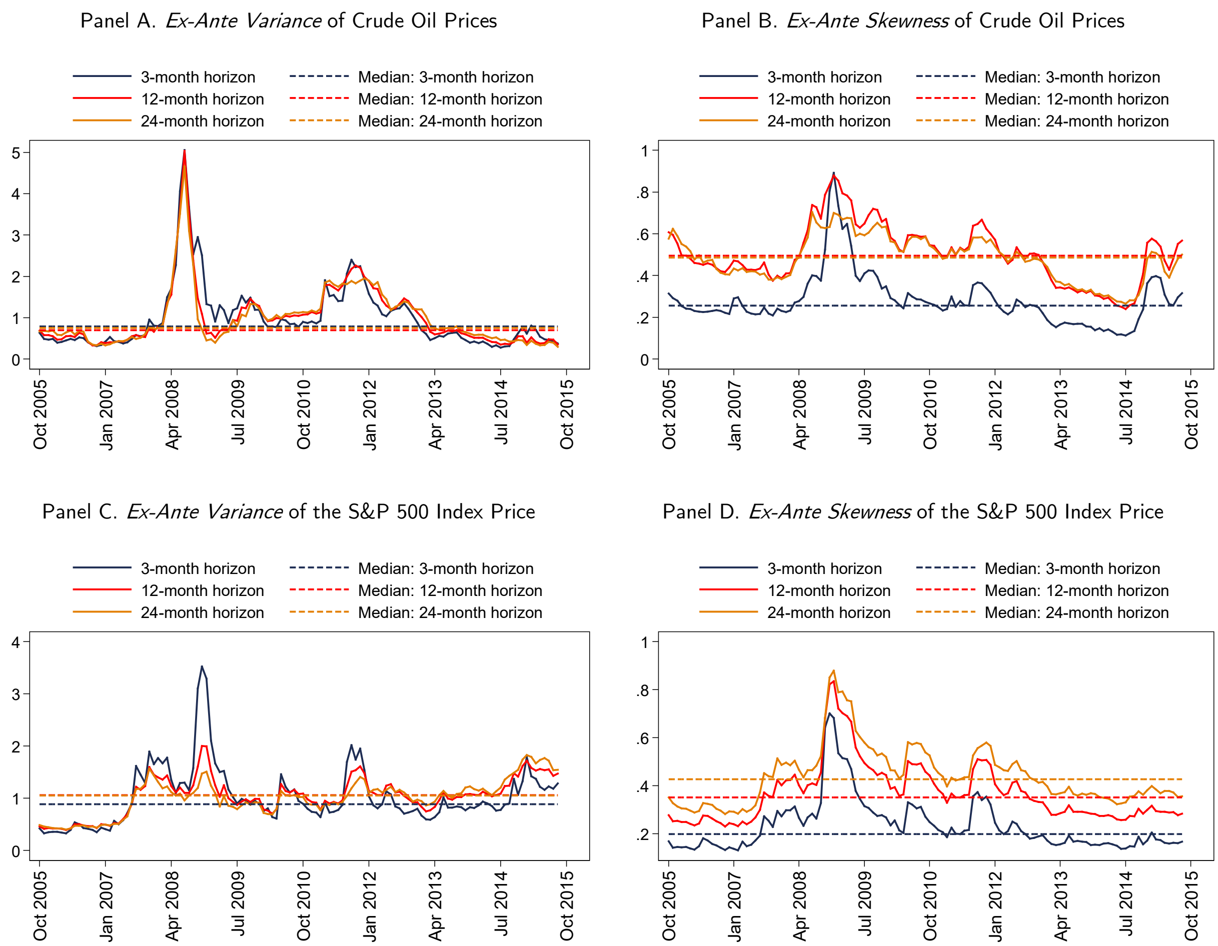

FIGURE 3. *EX-ANTE* VARIANCE AND SKEWNESS OF CRUDE OIL AND S&P 500 INDEX PRICES

Source: Bloomberg and author's calculations.
Note: Ex-ante variance and skewness of crude oil and the S&P 500 index prices are estimated via implied volatilities of option contract prices. Bloomberg data uses a variation of the option pricing model of Black-Scholes-Merton. The option contracts series are at the money. The variance series is scaled down using the sample standard deviation of the original series. The skewness series is scaled down using the maximum value of the corresponding original series.

A challenging step of the BCEF method involves the identification of prior series for the unobserved *ex-ante variance* and *ex-ante skewness* of the world GDP growth, Y. In this regard, we modeled the prior *ex-ante variance* and *ex-ante skewness* of world GDP growth as pointed out in the Appendix A3[16] using the sign restrictions presented in Table 1 and the available forward-looking variance and skewness series of the risk factors.

Furthermore, based on point forecasts of world GDP growth (World Bank 2016), Table 2 shows that there is generally an overprediction pattern in the elicited predictions. Thus, opposite to the overpredicted forecasts, the initial guess for the prior global *ex-ante skewness* is to assign more weight to the downside section of the world GDP growth predicted distribution; i.e., to assign negative skewness values.

TABLE 2 – HISTORICAL FORECAST ERRORS OF WORLD GDP GROWTH PREDICTIONS

Horizons:	Concurrent Year (T)	T+1	T+2	Combined Horizons
Mean	0.22	0.98	2.31	0.82
Median	0.16	0.73	1.49	0.51
Std. Dev.	1.20	1.85	1.78	1.66
Observations	13.00	11.00	4.00	28.00

Source: Global Economic Prospects (World Bank 2016).
Note: Realizations of world GDP growth based on WDI. Historical forecast error defined as predicted minus realized value. A positive HFE value indicates overprediction.

In this specific world GDP growth forecast, the mechanics of optimistic baseline forecasts assessed with more significant downside than upside risks are a summary and a statistical translation of the well-known phrase "hope for the best and prepare for the worst." Finally, once priors for the alpha coefficients and the future variance and skewness of real GDP are selected, the posterior distributions that solve the cross-entropy optimization model presented in Subsection 2.5 are ready to be estimated.

[16] There are multiple methods to obtain priors (Golan 2017). The procedure presented in Appendix A3 to proxy the prior distributions of the study parameters is only an alternative approximation.

3.2 Results

The main results of the application to calculate the density forecasts for world GDP growth based on the BCEF method are summarized in Table 3 and Figure 4. Table 3 shows the prior and posterior $a_{r,h}$ coefficients that solve the BCEF problem stated in Equation (16). Our priors of the $a_{r,h}$ coefficients at the concurrent period, $h = 0$ weight more—in absolute terms—the crude oil prices and the world term-spreads, while the measure of inflation target deviations and the S&P 500 index are the risk factors with coefficients with smaller magnitudes. At horizons $h = 1$ and $h = 2$, these prior values for the a_r coefficients are more homogeneous. The posterior estimates of a_r are quite heterogenous between risk factors and across time horizons (Table 3).

Table 3 – Posterior and Prior Average Direct Responses of World GDP to Movements Of Global risks

a_r Coefficients	Horizon	World Term Spread	World (CPI) Inflation Absolute Deviations from Target	World Energy (Crude Oil) Prices	S&P 500 Index
Posterior	$h = 0$	0.35	-0.29	-0.51	0.30
Posterior	$h = 1$	0.43	-0.46	-0.50	0.47
Posterior	$h = 2$	0.36	-0.36	-0.53	0.53
Prior	$h = 0$	0.51	-0.41	-0.69	0.24
Prior	$h = 1$	0.51	-0.55	-0.63	0.62
Prior	$h = 2$	0.54	-0.53	-0.66	0.69

Note: These coefficients are used in Equations (2), and (3) to shape the parametric TPN predictive distribution of world GDP growth.

These posterior estimates of the risk factor coefficients a_r feed into Equations (2) and (3) to recover the average series of *ex-ante* variance and skewness of GDP growth for three forecasting horizons. As seen in Panel A of Figure 4, the estimated *ex-ante* variance of world GDP growth has peaks around June 2008 and April 2012 across all the three studied time horizons. The series representing the *ex-ante* skewness of world GDP growth is in general negative, highlighting the bias of the density forecasts to the downside section of the distribution. These *ex-ante* skewness series vary substantially between horizons.

The estimates for the *ex-ante* variance and skewness of world GDP growth (Figure 4) are elements that feed into Equations (4) and (5) to generate density forecasts as the ones presented in Figure 5. This figure shows asymmetric confidence intervals of 40, 75, and 90 percent around the baseline—most likely outcome—for the current, next-year, and two-year ahead horizons. Panels A and B of Figure 5 are illustrations of hypothetically elicited density forecast at the end of March and August 2015.

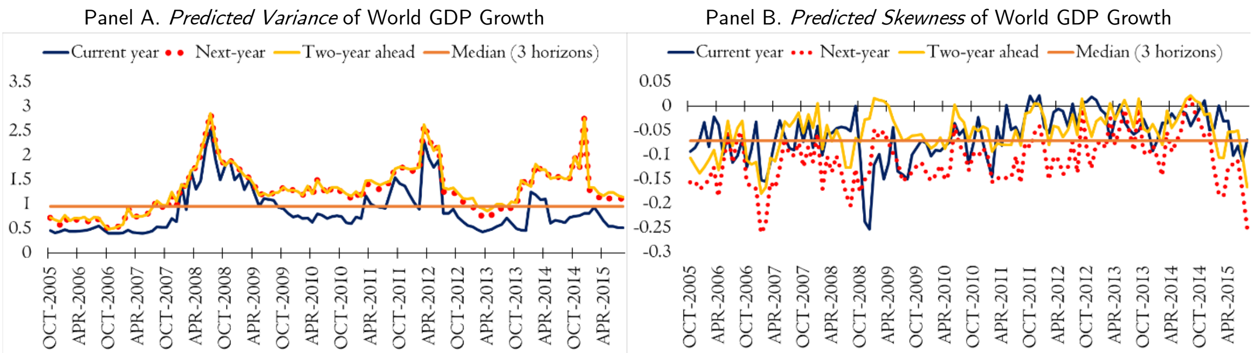

FIGURE 4. POSTERIOR ESTIMATES OF *EX-ANTE* VARIANCE AND SKEWNESS OF WORLD GDP GROWTH

Note: The series presented in Panel A and B are the posterior mean values of the BCEF estimates.

The BCEF estimates also provide substantial information about the distribution of risks. Table 4 shows the variance decomposition and the balance of risks (the skewness decomposition) of the density forecasts presented in Panels A and B of Figure 5. The results of the variance decomposition for the August 2015 density forecasts indicate that term spread in 2015 and crude oil prices in 2016 and 2017 are the risk factors with smaller share of the size of the variance of the predictive distribution. On the contrary, "other" risk factor components explain more than 32 percent of the variation across the three studied forecasting horizons. In terms of the skewness decomposition, crude oil prices are the risk factor that solely contributes more to the skewness magnitude, while "other" risk factor components also play a substantial role in the balance of risks in all the studied horizons (Table 4).

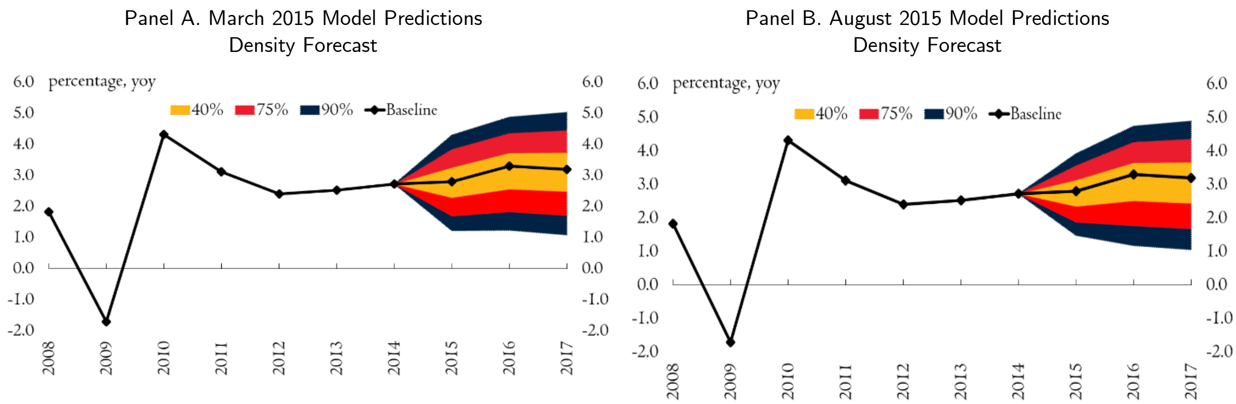

FIGURE 5. GDP GROWTH DENSITY FORECAST AND DECOMPOSITION OF UNCERTAINTY

Note: The monthly data spans the period of October 2005–August 2015. Panel A and B report the results corresponding to elicited density forecasts in March 2015, while panel B reports the August 2015 results.

TABLE 4 – VARIANCE AND SKEWNESS DECOMPOSITIONS OF WORLD GDP GROWTH FORECASTS

	Variance Decomposition (percentage)			Balance of Risks (percentage)		
Risk factor/horizon	2015	2016	2017	2015	2016	2017
March 2015 Elicited Density Forecast						
Term spread	7.6	14.2	9.1	-58.5	4.2	4.1
S&P 500	15.6	29.7	37.7	-15.3	-16.6	-50.3
Inflation target deviation	20.5	13.9	7.9	-29.8	40.1	33.0
Crude oil	22.7	9.1	8.2	162.5	36.9	67.1
Other	33.6	33.0	37.0	41.1	35.4	46.1
August 2015 Elicited Density Forecast						
Term spread	13.1	16.8	11.1	18.5	23.0	20.1
S&P 500	22.6	29.6	37.0	-6.6	-11.7	-31.0
Inflation target deviation	12.9	12.8	7.5	-18.5	25.4	18.6
Crude oil	19.0	7.9	7.1	59.3	28.0	43.9
Other	32.5	33.0	37.3	47.3	35.2	48.4

Note: The monthly data spans the period from October 2005 to August 2015. These results use the predictions obtained via the BCEF method. The balance of risks stands for the skewness decomposition. The sections denoting "Other" includes corresponding cross-moments of variance and skewness of the risk factors as well as "other" risk factor components not considered in the exercise. By using Equation (3), the contribution of the risk factor Z_r to the global skewness is estimated as $\frac{a_{r,h}{}^3\gamma(Z_{r,t+h})}{\gamma(Y_{t+h}|\mathcal{F}_t)}$.

3.3 EVALUATION

In this section, we generated naïve density forecasts to benchmark the BCEF predictions of world GDP growth. The purpose of the benchmarks only looks to show how bad or good the BCED density forecasts are doing concerning an elementary and time-saving form of constructing confidence intervals around the baseline. The naïve- density forecasts are built using historical World Bank semiannual point forecast series of world GDP growth (World

Bank 2016). The symmetric confidence interval is constructed attaching standard deviations of the historical forecast error (HFE) to the baseline series. In specific, the naïve density forecast follows a normal distribution with a mean equal to the most likely GDP growth outcome and a variance size equal to the corresponding HFE standard deviation. Two rules of thumb are used to generate these naïve density forecasts: i) horizon specific HFE standard deviations, and ii) average HFE standard deviations across horizons. Illustrations of these naïve-generated benchmark forecasts can be seen in Figure 6.

Panel A. Varying HFE Standard Deviation across Horizons

Panel B. Fixing HFE Standard Deviation across Horizons

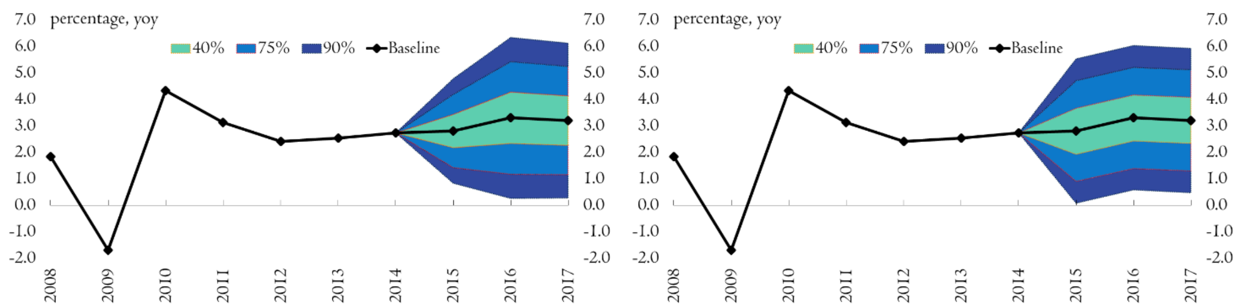

FIGURE 6. NAÏVE-GENERATED DENSITY FORECASTS OF GDP GROWTH

Note: The forecasts are symmetric and follow a normal distribution. Panel A shows a density forecast which varies given horizon-specific standard deviations of historical forecast errors. Panel B shows a density forecast that uses a single average historical forecast error standard deviation across horizons.

We use the CRPS proper scoring rule to evaluate and compare the reliability of the BCEF and naïve density forecasts. Unweighted and weighted averages of the individual CRPS scores are also estimated. The formulation of the weighted average and its interaction with the CRPS score is introduced in Subsection A4.2 in Appendix A4. A summary of the ex-post performance of the BCEF and the naïve density forecast series spanning the period 2005–2015 is presented in Table 5.

TABLE 5 – CONTINUOUS RANKED PROBABILITY SCORE OF DENSITY FORECASTS

	Average for TPN Density Forecasts (BCEF method)	Weighted Average for TPN Density Forecasts (BCEF method)	Average for Normal Density Forecasts (Varying HFE Standard Deviation Across Horizons)	Weighted Average for Normal Density Forecasts (Varying HFE Standard Deviation Across Horizons)
Panel A. Monthly Average Forecast Scores Across the Period of 2005–2015 and Three Annual Time-Horizons $h = \{0,1,2\}$.				
January	1.6	0.55	1.79	0.62
February	1.55	0.58	1.79	0.67
March	1.51	0.61	1.79	0.72
April	1.55	0.66	1.79	0.77
May	1.51	0.69	1.79	0.82
June	1.45	0.7	1.76	0.85
July	1.44	0.73	1.76	0.89
August	1.47	0.79	1.76	0.94
September	1.46	0.82	1.79	1
October	1.46	0.86	1.76	1.03
November	1.47	0.9	1.76	1.08
December	1.45	0.93	1.76	1.12
Panel B. Annual Average Forecast Scores Across Monthly Observations Over 2005–2015 and Three Annual Time-Horizons $h = \{0,1,2\}$.				
2005	1.04	0.62	1.5	0.9
2006	1.26	0.61	1.6	0.78
2007	2.6	0.92	2.65	0.99
2008	2.69	1.43	2.82	1.46
2009	1.6	1.03	1.97	1.15
2010	1.27	0.63	1.55	0.77
2011	1.23	0.56	1.47	0.68
2012	1.07	0.49	1.38	0.65
2013	0.89	0.4	1.32	0.62
2014	0.93	0.59	1.29	0.83
2015	0.72	0.57	0.97	0.77
Panel C. Horizon Average Forecast Scores Across Monthly Observations Over 2005–2015.				
$h = 0$	1.15	0.97	1.34	1.14
$h = 1$	1.73	0.89	2.07	1.07
$h = 2$	1.65	0.30	1.95	0.36
Panel D. Average Forecast Scores Across Monthly Observations Over 2005 – 2015 and Three Annual Time-Horizons $h = \{0,1,2\}$.				
Aggregated Score	490.90	243.15	583.41	289.45

Note: The weighted scores use masses derived following the rule presented in Equation (32). The weights penalize the forecasts under the following principle: when a forecast is elicited, it is expected that it has better accuracy if it is produced closer to the realization period than when it is released with a further horizon. In this sense, the closer to the realization period, the more punished (weighted) the elicited forecast. HFE stands for historical forecast error. The monthly estimated density forecasts cover the period of October 2005–August 2015. Forecasts generated in 2014 are evaluated only with the realizations of 2014 and 2015. Predictions made in 2015 are scored using realizations of 2015. Scores of naïve density forecasts generated using the rule of thumb of a fixed HFE standard deviation across the three forecasting horizons (results are available upon request) were estimated and not presented in this table. These latter scoring results remained the same as the ones shown in this table: BCEF density forecasts outperform naïve density predictions.

The CRPS scores overwhelmingly show more accurate density forecasts—lower score values—for the BCEF density forecasts than for the naïve predictions. The unweighted and weighted CRPS results are robust over four views: across horizons, months, years, and in the aggregated form. In this regard, the CRPS scores also indicate that the BCEF fan charts are more reliable to measure the uncertainty around the studied world GDP growth baseline than the naïve—normal distributed—density forecasts. The takeaway point here is that the BCEF forecasts at least overperform, in terms of prediction accuracy, the simple, inexpensive, and time-saving symmetric normal distributed density forecasts. Although in some settings, this will be a minimal consideration, the CRPS scores show that forecast accuracy and reliability are incentives to invest in a method like the BCEF presented in this paper.

4. Concluding remarks

Informed decisions require to consider potential upside and downside outcomes. For policymaking and investment purposes, predictive models that consider the modeling of asymmetric density forecasts provide superior reliability and decision-advantage. As per the procedure presented in this paper, the generation of forecasting procedures that allow the incorporation and update of prior information are encouraged to obtain more flexible models to enhance the reliability of elicited density forecasts. Revision of established prequential systems is essential in practice to improve forecast scores. The proper scoring rule framework is also encouraged to track the reliability and transparency of the predictions.

In this paper, we introduce a minimum cross-entropy procedure to generate asymmetric density forecasts in systematic form. The methodology relies on prior information of parameters and unobserved statistics, expectations of risk factors statistics, and the two-piece normal distribution assumption. Despite its constraints, the TPN distribution provides some flexibility to assess the upside and downside sections of the predictive density forecasts. The BCEF method allows the selected risk factors to depict the shape of the density forecasts, i.e., the description of the uncertainty around the most likely outcomes of the study variable.

The introduced BCEF method models a variance and skewness decomposition of the predictive distributions. The combination of the study variable and selected risk factors offers a simple linear regression framework to exploit *ex-ante* information regarding predictive *variance* and skewness. This *ex-ante* information is used to predict a confidence interval around the most likely outcome of a study variable. By construction, the BCEF procedure gives optimal results in the sense that it minimizes information losses when moving from the prior distribution assumptions to the posterior distribution estimates subject to consistency and behavioral constraints, and the data.

In the empirical application of predicting world GDP growth, the associated *ex-ante* market information emerges from term spreads, absolute deviations from inflation targets, energy prices, and the S&P 500 index prices. The recovered posterior series of the global GDP growth variance and skewness are used to construct hypothetical BCEF density forecasts. These BCEF world GDP growth forecasts are monthly-elicited spanning the period of October 2005–August 2015 and are evaluated using the continuous-ranked probability score. The unweighted and weighted average CRPS scores show that the Bayesian entropy fan charts outperform symmetric normal distributed—naïve—density forecasts, which are generated only using historical forecast error information. Noticeably, the estimated monthly BCEF-generated density forecasts for world GDP growth show predominantly negative values of skewness—downside skewed density forecasts—across the period of October 2005–August 2015. These negative skewness spotlight two aspects of the model: its easiness to incorporate prior information such as historical overprediction of world GDP growth, and its ability to minimize noise under the relative entropy setup.

Implementing proper scores—like the continuous-ranked probability score—provides strong incentives for the forecaster to elicit his or her best and honest density predictions. Evaluation of density forecasts through proper scoring rules provides transparency and the opportunity to distinguish the heterogeneity of forecasting model accuracy and forecaster skill. The recording of elicited predictions and their scoring rule evaluations will allow the comparison of forecasters and methods that aim to improve better investment and policymaking decisions.

5. References

Al-Najjar, N. I., and E. Shmaya. 2015. "Uncertainty and Disagreement in Equilibrium Models." *Journal of Political Economy* 123 (4): 778-808.

Aretz, K., and D. A. Peel. 2010. "Spreads Versus Professional Forecasters as Predictors of Future Output Change." *Journal of Forecasting* 29 (6): 517-522.

Bessler, D. A., and J. L. Kling. 1990. "Prequential Analysis of Cattle Prices " *Applied Statistics-Journal of the Royal Statistical Society Series C* 39 (1): 95-106.

Bickel, J. E. 2007. "Some Comparisons among Quadratic, Spherical, and Logarithmic Scoring Rules." *Decision Analysis* 4 (2): 49-65.

Blanchard, O. J., and J. Gali. 2007. "*The Macroeconomic Effects of Oil Shocks: Why are the 2000s so Different from the 1970s?*": National Bureau of Economic Research.

Blanchard, O. J., and M. Riggi. 2013. "Why are the 2000s so different from the 1970s? A structural Interpretation of Changes in the Macroeconomic Effects of Oil Prices." *Journal of the European Economic Association* 11 (5): 1032-1052.

Brier, G. W. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review* 78 (1): 1.

Brown, T. A. 1974. "Admissible Scoring Systems for Continuous Distributions." *The Rand Corporation*.

Bunn, D. W. 1984. *Applied Decision Analysis*. McGraw-Hill Book Company.

Clark, T. E., M. W. McCracken, and E. Mertens. 2018. "Modeling Time-Varying Uncertainty of Multiple-Horizon Forecast Errors." *The Review of Economics and Statistics*.

Dawid, A. P. 1982. "The Well-Calibrated Bayesian." *Journal of the American Statistical Association* 77 (379): 605-610.

----------. 1984. "Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach." *Journal of the Royal Statistical Society. Series A (General)* 147 (2): 278-292.

De Finetti, B. 1937. *Foresight: Its Logical Laws, its Subjective Sources* New York: John Wiley and Sons.

Diebold, F. X., J. Hahn, and A. S. Tay. 1999. "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange." *Review of Economics and Statistics* 81 (4): 661-673.

Epstein, E. S. 1969. "A Scoring System for Probability Forecasts of Ranked Categories." *Journal of Applied Meteorology* 8 (6): 985-987.

Estrella, A., and G. A. Hardouvelis. 1991. "The Term Structure as a Predictor of Real Economic Activity." *The Journal of Finance* 46 (2): 555-576.

Fama, E. F. 1981. "Stock Returns, Real Activity, Inflation, and Money." *The American Economic Review*: 545-565.

Fischer, S., and R. C. Merton. 1984. *Macroeconomics and Finance: The Role of the Stock Market*. Elsevier.

Geweke, J., and C. Whiteman. 2006. "Bayesian Forecasting." *Handbook of Economic Forecasting* 1: 3-80.

Giordani, P., and P. Söderlind. 2003. "Inflation Forecast Uncertainty." *European Economic Review* 47 (6): 1037-1059.

Gneiting, T. 2011. "Quantiles as Optimal Point Forecasts." *International Journal of Forecasting* 27 (2): 197-207.

Gneiting, T., F. Balabdaoui, and A. E. Raftery. 2007. "Probabilistic Forecasts, Calibration and Sharpness." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 69 (2): 243-268.

Gneiting, T., and A. E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102 (477): 359-378.

Gneiting, T., A. E. Raftery, A. H. Westveld III, and T. Goldman. 2005. "Calibrated Probabilistic Forecasting using Ensemble Model Output Statistics and Minimum CRPS Estimation." *Monthly Weather Review* 133 (5): 1098-1118.

Gneiting, T., and T. L. Thorarinsdottir. 2010. "Predicting Inflation: Professional Experts versus No-Change Forecasts." *arXiv preprint arXiv:1010.2318*.

Go, D. S., H. Lofgren, F. Mendez-Ramos, and S. Robinson. 2016. "Estimating Parameters and Structural Change in CGE Models using a Bayesian Cross-Entropy Estimation Approach." *Economic Modelling* 52: 790-811.

Golan, A. 2017. *Foundations of Info-Metrics: Modeling and Inference with Imperfect Information*. Oxford University Press.

Golan, A., G. Judge, and S. Robinson. 1994. "Recovering Information from Incomplete or Partial Multisectoral Economic Data." *The Review of Economics and Statistics*: 541-549.

Golan, A., G. G. Judge, and D. Miller. 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Wiley New York.

Granger, C. W. J. 1986. "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts: Comment." *Journal of Business & Economic Statistics* 4 (1): 16-17.

Grimit, E. P., T. Gneiting, V. Berrocal, and N. A. Johnson. 2006. "The Continuous Ranked Probability Score for Circular Variables and its Application to Mesoscale Forecast Ensemble Verification." *Quarterly Journal of the Royal Meteorological Society* 132 (621C): 2925-2942.

Hamilton, J. D. 2005. "Oil and the Macroeconomy." *The New Palgrave Dictionary of Economics*: 201-228.

Harvey, C. R. 1989. "Forecasts of Economic Growth from the Bond and Stock Markets." *Financial Analysts Journal* 45 (5): 38-45.

Hassapis, C., and S. Kalyvitis. 2002. "Investigating the Links Between Growth and Real Stock Price Changes with Empirical Evidence from the G-7 Economies." *The Quarterly Review of Economics and Finance* 42 (3): 543-575.

Haubrich, J. G., and A. M. Dombrosky. 1996. "Predicting Real Growth using the Yield Curve." *Economic Review-Federal Reserve Bank of Cleveland* 32: 26-35.

Hersbach, H. 2000. "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems." *Weather and Forecasting* 15 (5): 559-570.

Jay Choi, J., S. Hauser, and K. J. Kopecky. 1999. "Does the stock market predict real activity? Time series evidence from the G-7 countries." *Journal of Banking & Finance* 23 (12): 1771-1792.

Jimenez-Rodriguez, R., and M. Sánchez. 2005. "Oil Price Shocks and Real GDP Growth: Empirical Evidence for Some OECD Countries." *Applied Economics* 37 (2): 201-228.

Johnstone, D. 2007a. "Economic Darwinism: Who has the Best Probabilities?" *Theory and Decision* 62 (1): 47-96.

----------. 2007b. "The Value of a Probability Forecast from Portfolio Theory." *Theory and Decision* 63 (2): 153-203.

Johnstone, D., V. R. R. Jose, and R. L. Winkler. 2011. "Tailored Scoring Rules for Probabilities." *Decision Analysis* 8 (4): 256.

Judge, G. G., and R. C. Mittelhammer. 2011. *An Information Theoretic Approach to Econometrics*. Cambridge University Press.

Kao, Y.-C., C.-M. Kuan, and S. Chen. 2013. "Testing the Predictive Power of the Term Structure Without Data Snooping Bias." *Economics Letters* 121 (3): 546-549.

Kling, J. L. 1987. "Predicting the Turning Points of Business and Economic Time Series." *The Journal of Business* 60 (2): 201-238.

Kling, J. L., and D. A. Bessler. 1989. "Calibration-Based Predictive Distributions: An Application of Prequential Analysis to Interest Rates, Money, Prices, and Output." *The Journal of Business* 62 (4): 477-499.

Kohonen, J., and J. Suomela. 2006. *Lessons Learned in the Challenge: Making Predictions and Scoring Them*. Springer.

Krämer, W. 2006. "Evaluating Probability Forecasts in Terms of Refinement and Strictly Proper Scoring Rules." *Journal of Forecasting* 25 (3): 223-226.

Lichtenstein, S., B. Fischhoff, and L. D. Phillips. 1982. "Calibration of Probabilities: The State of the Art to 1980." In *Judgement under Uncertainty: Heuristics and Biases* edited by D. Kahneman, P. Slovic, and A. Tversky, 306-334. Cambridge University Press.
Matheson, J. E., and R. L. Winkler. 1976. "Scoring Rules for Continuous Probability Distributions." *Management Science* 22 (10): 1087-1096.
Mitchell, W. C., and A. F. Burns. 1938. "Statistical Indicators of Cyclical Revivals." In *Statistical Indicators of Cyclical Revivals*, 1-12. NBER.
Murphy, A. H. 1971. "A Note on the Ranked Probability Score." *Journal of Applied Meteorology* 10 (1): 155-156.
Murphy, A. H., and E. S. Epstein. 1967. "A Note on Probability Forecasts and “Hedging”." *Journal of Applied Meteorology* 6 (6): 1002-1004.
Nelson, R. G., and D. A. Bessler. 1989. "Subjective Probabilities and Scoring Rules: Experimental Evidence." *American Journal of Agricultural Economics* 71 (2): 363-369.
Rich, R., and J. Tracy. 2010. "The Relationships Among Expected Inflation, Disagreement, and Uncertainty: Evidence from Matched Point and Density Forecasts." *The Review of Economics and Statistics* 92 (1): 200-207.
Sanders, F. 1963. "On Subjective Probability Forecasting." *Journal of Applied Meteorology* 2 (3): 191-201.
Savage, L. J. 1971. "Elicitation of Personal Probabilities and Expectations " *Journal of the American Statistical Association* 66 (336): 783-&.
Schervish, M. J. 1989. "A General Method for Comparing Probability Assessors " *Annals of Statistics* 17 (4): 1856-1879.
Stock, J. H., and M. W. Watson. 2003. "Forecasting Output and Inflation: The Role of Asset Prices." *Journal of Economic Literature* 41 (3): 788-829.
Tsouma, E. 2009. "Stock Returns and Economic Activity in Mature and Emerging Markets." *The Quarterly Review of Economics and Finance* 49 (2): 668-685.
Von Holstein, C. 1977. *The Continuous Ranked Probability Score in Practice*. Springer.
Wallis, K. F. 2014. "The Two-Piece Normal, Binormal, or Double Gaussian Distribution: Its Origin and Rediscoveries." *Statistical Science* 29 (1): 106-112.
Wheelock, D. C., and M. E. Wohar. 2009. "Can the Term Spread Predict Output Growth and Recessions? A Survey of the Literature." *Federal Reserve Bank of St. Louis Review* 91 (5 Part 1): 419-440.
Winkler, R. L. 1986. "On Good Probability Appraisers." In *Bayesian Inference and Decision Techniques* edited by P. G. a. A. Zellner, 265-278. Amsterdam: Elsevier.
Winkler, R. L. 1996. "Scoring Rules and the Evaluation of Probabilities." *TEST* 5 (1): 1-60.
Winkler, R. L., and A. H. Murphy. 1968. ""Good" Probability Assessors." *Journal of Applied Meteorology* 7: 751-758.
Winkler, R. L., and A. H. Murphy. 1970. "Nonlinear Utility and the Probability Score." *Journal of Applied Meteorology* 9 (1): 143.
World Bank (A. Kose, and F. Ohnsorge). 2016. "*Spillovers amid Weak Growth*." Washington, DC: World Bank Group.

Appendix A1. Proof of Proposition 1

The conditional variance of Y_{t+h} using the linear assumption in (1) becomes:

$$var\left(Y_{t+h}|\mathscr{F}_t\right) = var\left(a_1 Z_{1,t+h} + \cdots + a_{N_R} Z_{N_R,t+h} + \varepsilon_{t+h}|\mathscr{F}_t\right)$$

$$= \mathbb{E}\left[\left(\varepsilon_{t+h} + \sum_{i=1}^{R} a_i Z_{i,t+h}\right)^2 | \mathscr{F}_t\right] - \mathbb{E}^2\left[\left(\varepsilon_{t+h} + \sum_{i=1}^{R} a_i Z_{i,t+h}\right) | \mathscr{F}_t\right]$$

$$= \mathbb{E}\left[\left(\varepsilon_{t+h}^2 + 2\sum_{i=1}^{R} a_i Z_{i,t+h}\varepsilon_{t+h} + \sum_{i=1}^{R} a_i{}^2 Z_{i,t+h}{}^2 + \sum_{\substack{j=1 \\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j Z_{i,t+h} Z_{j,t+h}\right) | \mathscr{F}_t\right]$$

$$- \left[\mathbb{E}^2\left[\varepsilon_{t+h} | \mathscr{F}_t\right] + 2\sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h} | \mathscr{F}_t\right] \mathbb{E}\left[\varepsilon_{t+h} | \mathscr{F}_t\right] + \sum_{i=1}^{R} a_i{}^2 \mathbb{E}^2\left[Z_{i,t+h} | \mathscr{F}_t\right]\right.$$

$$\left. + \sum_{\substack{j=1 \\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j \mathbb{E}\left[Z_{i,t+h} | \mathscr{F}_t\right] \mathbb{E}\left[Z_{j,t+h} | \mathscr{F}_t\right]\right]$$

$$= \sum_{i=1}^{R} a_i{}^2 \mathbb{E}\left[Z_{i,t+h}{}^2 | \mathscr{F}_t\right] - \sum_{i=1}^{R} a_i{}^2 \mathbb{E}^2\left[Z_{i,t+h} | \mathscr{F}_t\right]$$

$$+ \sum_{\substack{j=1 \\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j \mathbb{E}\left[Z_{i,t+h} Z_{j,t+h} | \mathscr{F}_t\right] - \sum_{\substack{j=1 \\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j \mathbb{E}\left[Z_{i,t+h} | \mathscr{F}_t\right] \mathbb{E}\left[Z_{j,t+h} | \mathscr{F}_t\right] +$$

$$+ \mathbb{E}\left[\varepsilon_{t+h}^2 | \mathscr{F}_t\right] - \mathbb{E}^2\left[\varepsilon_{t+h} | \mathscr{F}_t\right]$$

$$+ 2\sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}\varepsilon_{t+h} | \mathscr{F}_t\right] - 2\sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h} | \mathscr{F}_t\right] \mathbb{E}\left[\varepsilon_{t+h} | \mathscr{F}_t\right]$$

$$= \sum_{i=1}^{R} a_i{}^2 \mathbb{E}\left[Z_{i,t+h}{}^2\right] - \sum_{i=1}^{R} a_i{}^2 \mathbb{E}^2\left[Z_{i,t+h}\right]$$

$$+ \sum_{\substack{j=1 \\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j \mathbb{E}\left[Z_{i,t+h} Z_{j,t+h}\right] - \sum_{\substack{j=1 \\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j \mathbb{E}\left[Z_{i,t+h}\right] \mathbb{E}\left[Z_{j,t+h}\right]$$

$$+ \mathbb{E}\left[\varepsilon_{t+h}^2\right] - \mathbb{E}^2\left[\varepsilon_{t+h}\right] + 2\sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}\varepsilon_{t+h}\right] - 2\sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}\right] \mathbb{E}\left[\varepsilon_{t+h}\right]$$

$$= \sum_{i=1}^{R} a_i{}^2\ var\left[Z_{i,t+h}\right] + \sum_{\substack{j=1 \\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j cov\left[Z_{i,t+h} Z_{j,t+h}\right] + var\left[\varepsilon_{t+h}\right]$$

$$+ 2\sum_{i=1}^{R} a_i cov\left[Z_{i,t+h}\varepsilon_{t+h}\right].$$

Denoting $\epsilon_{\sigma^2,t+h}{}^2 = \sum_{\substack{j=1 \\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j cov\left[Z_{i,t+h} Z_{j,t+h}\right] + var\left[\varepsilon_{t+h}\right] + 2\sum_{i=1}^{R} a_i cov\left[Z_{i,t+h}\varepsilon_{t+h}\right]$

gives us Equation (2). ∎

Appendix A2. Proof of Proposition 2

Observe that for a constant a and a random variable X, the third central moment of aX becomes:

$$\gamma(aX) = \mathbb{E}\left[(aX - a\mathbb{E}X)^3\right] \tag{17}$$

$$= \mathbb{E}\left[a^3 X^3 - 3a^3 X^2 \mathbb{E}X + 3a^3 X \mathbb{E}^2 X - a^3 \mathbb{E}^3 X\right]$$

$$= a^3\left[\mathbb{E}X^3 - 3\mathbb{E}X \bullet var(X) - \mathbb{E}^3 X\right].$$

Using the results of $var(Y_{t+h}|\mathscr{F}_t)$ from Appendix and the linear assumption in (1), the conditional skewness of the *ex-ante* value of Y at period t and horizon h becomes:

$$\text{(18)}\quad \gamma(Y_{t+h}|\mathscr{F}_t) = \mathbb{E}\left[\left(Y_{t+h} - \mathbb{E}(Y_{t+h})\right)^3|\mathscr{F}_t\right]$$

$$= \underbrace{\mathbb{E}\left[\left(\varepsilon_{t+h} + \sum_{i=1}^{R} a_i Z_{i,t+h}\right)^3|\mathscr{F}_t\right]}_{section\ A}$$

$$\underbrace{-3\mathbb{E}\left[\left(\varepsilon_{t+h} + \sum_{i=1}^{R} a_i Z_{i,t+h}\right)|\mathscr{F}_t\right] var\left(\left(\varepsilon_{t+h} + \sum_{i=1}^{R} a_i Z_{i,t+h}\right)|\mathscr{F}_t\right)}_{section\ B}$$

$$- \underbrace{\mathbb{E}^3\left[\left(\varepsilon_{t+h} + \sum_{i=1}^{R} a_i Z_{i,t+h}\right)|\mathscr{F}_t\right]}_{section\ C}.$$

Expanding section A of (18):

$$\text{(19)}\quad \underbrace{\mathbb{E}\left[\left(\varepsilon_{t+h} + \sum_{i=1}^{R} a_i Z_{i,t+h}\right)^3|\mathscr{F}_t\right]}_{section\ A} = \mathbb{E}\left[\varepsilon_{t+h}^3|\mathscr{F}_t\right] + \sum_{i=1}^{R} a_i^3 \mathbb{E}\left[Z_{i,t+h}^3|\mathscr{F}_t\right]$$

$$+3\mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right] \sum_{i=1}^{R} a_i^2 \mathbb{E}\left[Z_{i,t+h}^2|\mathscr{F}_t\right]$$

$$+3\mathbb{E}\left[\varepsilon_{t+h}^2|\mathscr{F}_t\right] \sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}|\mathscr{F}_t\right]$$

$$= \mathbb{E}\left[\varepsilon_{t+h}^3\right] + \sum_{i=1}^{R} a_i^3 \mathbb{E}\left[Z_{i,t+h}^3\right]$$

$$+3 \sum_{i=1}^{R} a_i^2 \mathbb{E}\left[Z_{i,t+h}^2\right] \bullet \mathbb{E}\left[\varepsilon_{t+h}\right]$$

$$+3 \sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}\right] \bullet \mathbb{E}\left[\varepsilon_{t+h}^2\right].$$

Expanding section B of (18):

$$\text{(20)}\quad \underbrace{3\,\mathbb{E}\left[\left(\varepsilon_{t+h} + \sum_{i=1}^{R} a_i Z_{i,t+h}\right)|\mathscr{F}_t\right] var\left(\left(\varepsilon_{t+h} + \sum_{i=1}^{R} a_i Z_{i,t+h}\right)|\mathscr{F}_t\right)}_{section\ B}$$

$$= 3\left(\mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right] + \sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}|\mathscr{F}_t\right]\right)$$

$$\left(\sum_{i=1}^{R} a_i^{\,2}\ var\left[Z_{i,t+h}\right] + \sum_{\substack{j=1\\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j cov\left[Z_{i,t+h} Z_{j,t+h}\right]\right.$$

$$\left. + var\left[\varepsilon_{t+h}\right] + 2\sum_{i=1}^{R} a_i cov\left[Z_{i,t+h}\varepsilon_{t+h}\right]\right)$$

$$= 3\,\mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right] \sum_{i=1}^{R} a_i^{\,2}\ var\left[Z_{i,t+h}\right]$$

$$+3\,\mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right]\ \sum_{\substack{j=1\\ j\neq i}}^{R}\sum_{i=1}^{R} a_i a_j cov\left[Z_{i,t+h}Z_{j,t+h}\right]$$
$$+3\,\mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right] var\left[\varepsilon_{t+h}\right]$$
$$+3\bullet 2\bullet \mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right]\sum_{i=1}^{R} a_i cov\left[Z_{i,t+h}\varepsilon_{t+h}\right]$$
$$+3\bullet \sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}|\mathscr{F}_t\right]\bullet \sum_{i=1}^{R} a_i{}^2\bullet var\left[Z_{i,t+h}\right]$$
$$+3\bullet \sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}|\mathscr{F}_t\right]\bullet \sum_{\substack{j=1\\ j\neq i}}^{R}\sum_{i=1}^{R} a_i a_j cov\left[Z_{i,t+h}Z_{j,t+h}\right]$$
$$+3\bullet \sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}|\mathscr{F}_t\right]\bullet var\left[\varepsilon_{t+h}\right]$$
$$+3\bullet 2\bullet \sum_{i=1}^{N_R} a_i \mathbb{E}\left[\hat{Z}_{i,t+h}|\mathscr{F}_t\right]\bullet \sum_{i=1}^{R} a_i cov\left[Z_{i,t+h}\varepsilon_{t+h}\right]$$
$$= 3\bullet \mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right]\sum_{i=1}^{R} a_i{}^2\bullet var\left[Z_{i,t+h}\right]$$
$$+3\bullet \mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right]\sum_{\substack{j=1\\ j\neq i}}^{R}\sum_{i=1}^{R} a_i a_j cov\left[Z_{i,t+h}Z_{j,t+h}\right]$$
$$+3\bullet \mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right]\bullet var\left[\varepsilon_{t+h}\right]$$
$$+6\bullet \mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right]\sum_{i=1}^{R} a_i cov\left[Z_{i,t+h}\varepsilon_{t+h}\right]$$
$$+3\bullet \sum_{i=1}^{R}\sum_{j=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}|\mathscr{F}_t\right]\bullet a_j{}^2\bullet var\left[Z_{j,t+h}\right]$$
$$+3\bullet \sum_{k=1}^{R}\sum_{\substack{j=1\\ j\neq i}}^{R}\sum_{i=1}^{R} a_i a_j a_k\bullet cov\left[Z_{i,t+h}Z_{j,t+h}\right]\bullet \mathbb{E}\left[Z_{k,t+h}|\mathscr{F}_t\right]\bullet$$
$$+3\bullet \sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}|\mathscr{F}_t\right]\bullet var\left[\varepsilon_{t+h}\right]$$
$$+6\bullet \sum_{i=1}^{R}\sum_{j=1}^{R} a_i a_j\bullet \mathbb{E}\left[Z_{i,t+h}|\mathscr{F}_t\right]\bullet cov\left[Z_{j,t+h}\varepsilon_{t+h}\right]$$
$$= 3\bullet \mathbb{E}\left[\varepsilon_{t+h}\right]\sum_{i=1}^{R} a_i{}^2\bullet var\left[Z_{i,t+h}\right]$$
$$+3\bullet \mathbb{E}\left[\varepsilon_{t+h}\right]\sum_{\substack{j=1\\ j\neq i}}^{R}\sum_{i=1}^{R} a_i a_j cov\left[Z_{i,t+h}Z_{j,t+h}\right]$$
$$+3\bullet \mathbb{E}\left[\varepsilon_{t+h}\right]\bullet var\left[\varepsilon_{t+h}\right]$$
$$+6\bullet \mathbb{E}\left[\hat{\varepsilon}_{t+h}\right]\sum_{i=1}^{R} a_i cov\left[Z_{i,t+h}\hat{\varepsilon}_{t+h}\right]$$
$$+3\bullet \sum_{i=1}^{R}\sum_{j=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}\right]\bullet a_j{}^2\bullet var\left[Z_{j,t+h}\right]$$
$$+3\bullet \sum_{k=1}^{R}\sum_{\substack{j=1\\ j\neq i}}^{R}\sum_{i=1}^{R} a_i a_j a_k\bullet cov\left[Z_{i,t+h}Z_{j,t+h}\right]\bullet \mathbb{E}\left[Z_{k,t+h}\right]$$
$$+3\bullet \sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}\right]\bullet var\left[\varepsilon_{t+h}\right]$$
$$+6\bullet \sum_{i=1}^{R}\sum_{j=1}^{R} a_i a_j\bullet \mathbb{E}\left[Z_{i,t+h}\right]\bullet cov\left[Z_{j,t+h}\varepsilon_{t+h}\right]$$
$$= 3\bullet \mathbb{E}\left[\varepsilon_{t+h}\right]\sum_{i=1}^{R} a_i{}^2\bullet var\left[Z_{i,t+h}\right]$$
$$+3\bullet \mathbb{E}\left[\varepsilon_{t+h}\right]\sum_{\substack{j=1\\ j\neq i}}^{R}\sum_{i=1}^{R} a_i a_j cov\left[Z_{i,t+h}Z_{j,t+h}\right]$$
$$+3\bullet \mathbb{E}\left[\varepsilon_{t+h}\right]\bullet var\left[\varepsilon_{t+h}\right]$$
$$+6\bullet \mathbb{E}\left[\varepsilon_{t+h}\right]\sum_{i=1}^{R} a_i cov\left[Z_{i,t+h}\varepsilon_{t+h}\right]$$
$$+3\bullet \sum_{\substack{j=1\\ j=i}}^{R}\sum_{i=1}^{R} a_j \mathbb{E}\left[Z_{j,t+h}\right]\bullet a_i{}^2\bullet var\left[Z_{i,t+h}\right]$$
$$+3\bullet \sum_{\substack{j=1\\ j\neq i}}^{R}\sum_{i=1}^{R} a_j \mathbb{E}\left[Z_{j,t+h}\right]\bullet a_i{}^2\bullet var\left[Z_{i,t+h}\right]$$
$$+3\bullet \sum_{k=1}^{R}\sum_{\substack{j=1\\ j\neq i}}^{R}\sum_{i=1}^{R} a_i a_j a_k\bullet cov\left[Z_{i,t+h}Z_{j,t+h}\right]\bullet \mathbb{E}\left[Z_{k,t+h}\right]$$

$+3 \bullet \sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}\right] \bullet var\left[\varepsilon_{t+h}\right]$
$+6 \bullet \sum_{i=1}^{R} \sum_{j=1}^{R} a_i a_j \bullet \mathbb{E}\left[Z_{i,t+h}\right] \bullet cov\left[Z_{j,t+h}\varepsilon_{t+h}\right]$

Expanding section C of (18):

$$(21) \quad \underbrace{\mathbb{E}^3\left[\left(\varepsilon_{t+h} + \sum_{i=1}^{R} a_i Z_{i,t+h}\right)|\mathscr{F}_t\right]}_{section\ C} = \left(\mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right] + \mathbb{E}\left[\sum_{i=1}^{R} a_i Z_{i,t+h}\,|\mathscr{F}_t\right]\right)^3$$

$$=\mathbb{E}^3\left[\varepsilon_{t+h}|\mathscr{F}_t\right]$$
$$+\mathbb{E}^3\left[\sum_{i=1}^{R} a_i Z_{i,t+h}\,|\mathscr{F}_t\right]$$
$$+\ 3 \bullet \mathbb{E}\left[\varepsilon_{t+h}|\mathscr{F}_t\right] \bullet \mathbb{E}^2\left[\sum_{i=1}^{R} a_i Z_{i,t+h}\,|\mathscr{F}_t\right]$$
$$+\ 3 \bullet \mathbb{E}^2\left[\varepsilon_{t+h}|\mathscr{F}_t\right] \bullet \mathbb{E}\left[\sum_{i=1}^{R} a_i Z_{i,t+h}\,|\mathscr{F}_t\right].$$
$$=\mathbb{E}^3\left[\varepsilon_{t+h}\right]$$
$$+\sum_{i=1}^{R} a_i^3 \mathbb{E}^3\left[Z_{i,t+h}\right]$$
$$+\ 3 \bullet \mathbb{E}\left[\varepsilon_{t+h}\right] \bullet \sum_{i=1}^{R} a_i^2 \mathbb{E}^2\left[Z_{i,t+h}\right]$$
$$+\ 3 \bullet \mathbb{E}^2\left[\varepsilon_{t+h}\right] \bullet \sum_{i=1}^{R} a_i \mathbb{E}\left[Z_{i,t+h}\right].$$

Let us define the error term, ${\epsilon_{\gamma,t+h}}^3$ in (22):

$$(22) \quad {\epsilon_{\gamma,t+h}}^3 = \gamma\left[\varepsilon_{t+h}\right] -3 \bullet \mathbb{E}\left[\varepsilon_{t+h}\right] \bullet \sum_{\substack{j=1\\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j \bullet cov\left[Z_{i,t+h} Z_{j,t+h}\right]$$
$$-6 \bullet \mathbb{E}\left[\varepsilon_{t+h}\right] \bullet \sum_{i=1}^{R} a_i \bullet cov\left[Z_{i,t+h}\varepsilon_{t+h}\right]$$
$$-3 \bullet \sum_{\substack{j=1\\ j\neq i}}^{R} \sum_{i=1}^{R} a_j \mathbb{E}\left[Z_{j,t+h}\right] \bullet {a_i}^2 \bullet var\left[Z_{i,t+h}\right]$$
$$-3 \bullet \sum_{k=1}^{R} \sum_{\substack{j=1\\ j\neq i}}^{R} \sum_{i=1}^{R} a_i a_j a_k \bullet cov\left[Z_{i,t+h} Z_{j,t+h}\right] \bullet \mathbb{E}\left[Z_{k,t+h}\right]$$
$$-6 \bullet \sum_{i=1}^{R} \sum_{j=1}^{R} a_i a_j \bullet \mathbb{E}\left[Z_{i,t+h}\right] \bullet cov\left[Z_{j,t+h}\varepsilon_{t+h}\right].$$

Inserting the results from (19), (20), and (21) into (18) and defining the error term ${\epsilon_{\gamma,t+h}}^3$ as in (22) gives (3). ∎

Appendix A3. (Optional estimation and online appendix) Prior predictive variance and skewness

In this section, we present a procedure to proxy the prior mean values of the predictive variance and skewness series, $\left\{var^0\left(Y_{t+h}|\mathscr{F}_t\right)\right\}_{t=t_1}^{T}$ and $\left\{\gamma^0\left(Y_{t+h}|\mathscr{F}_t\right)\right\}_{t=t_1}^{T}$. Although this step is not central in the BCEF method, the procedure could help to polish or frame some heterogeneity in the mentioned predictive series. This optional approximation involves the estimation of a system of equations via constrained ordinary least squares (OLS). As a first step, we select preliminary prior mean values for the coefficients embedded in Equations (2) and (3) and which we denote with $\tilde{A}^0 = \{\tilde{a}_1{}^0, \tilde{a}_2{}^0, \ldots, \tilde{a}_R{}^0\}$. Then, we proxy preliminary series of the prior variance and skewness that we denote with $\{\widetilde{var}^0(Y_{t+h}|\mathscr{F}_t)\}_{t=t_1}^{T}$ and $\{\tilde{\gamma}^0(Y_{t+h}|\mathscr{F}_t)\}_{t=t_1}^{T}$ using the Equations (23) and (24) for all periods $t \in \mathscr{T}$ and horizons $h \in \mathscr{H}$.

(23) $\widetilde{var}^0(Y_{t+h}|\mathscr{F}_t) = \sum_{r\subseteq\mathscr{R}} (\tilde{a}_{r,h}{}^0)^2 var(Z_{r,t+h}|\mathscr{F}_t)$.

(24) $\tilde{\gamma}^0(Y_{t+h}|\mathscr{F}_t) = \sum_{r\subseteq\mathscr{R}} (\tilde{a}_{r,h}{}^0)^3 \gamma(Z_{r,t+h}|\mathscr{F}_t)$.

Once the prior series $\{\widetilde{var}^0(Y_{t+h}|\mathscr{F}_t)\}_{t=t_1}^{T}$ and $\{\tilde{\gamma}^0(Y_{t+h}|\mathscr{F}_t)\}_{t=t_1}^{T}$ are obtained, we solve a constrained OLS problem as defined through Equations (25)–(29).

$$(25) \qquad \underset{\left\{\substack{\{var^0(Y_{t+h}|\mathscr{F}_t)\}_{t=t_1}^{T};\\ \{\gamma^0(Y_{t+h}|\mathscr{F}_t)\}_{t=t_1}^{T};\, a_{r,h}{}^0}\right\}}{Min} \left\{\begin{matrix} \sum_h \sum_t \left(var^0(Y_{t+h}|\mathscr{F}_t) - \widetilde{var}^0(Y_{t+h}|\mathscr{F}_t)\right)^2 \\ + \\ \sum_h \sum_t \left(\gamma^0(\hat{Y}_{t+h}|\mathscr{F}_t) - \tilde{\gamma}^0(\hat{Y}_{t+h}|\mathscr{F}_t)\right)^2 \end{matrix}\right\}.$$

Subject to

(26) $var^0(Y_{t+h}|\mathscr{F}_t) = \sum_{r\subseteq R} (a_{r,h}{}^0)^2 var(Z_{r,t+h}|\mathscr{F}_t), \quad \forall t \in \mathscr{T}, h \in \mathscr{H}$.

(27) $\gamma^0(\hat{Y}_{t+h}|\mathscr{F}_t) = \sum_{r\subseteq R} (a_{r,h}{}^0)^3 \gamma(Z_{r,t+h}|\mathscr{F}_t), \quad \forall t \in \mathscr{T}, h \in \mathscr{H}$.

(28) $a_{1,h}{}^0 \lesseqgtr 0, \ldots, a_{R,h}{}^0 \lesseqgtr 0$.

(29) $(a_{r,h}{}^0)^2 var(Z_{r,t+h}|\mathscr{F}_t) \leq b_{r,h}, \quad \forall r \in \mathscr{R},\ t \in \mathscr{T}, h \in \mathscr{H}$.

Appendix A4. Scores (Online Appendix)

For reliability purposes and transparency, predictive densities need to be evaluated. Without proper evaluation of the predictive outcomes, the forecast exercise becomes meaningless and unreliable. The prequential forecasting system framework of Dawid (1984) combined with the statistical and economic concept of proper scoring rules of Savage (1971) is applied in this research to assess the dynamic performance of the Bayesian entropy-generated density forecasts. While the revision of established forecasting systems is critical in enhancing the quality of predictions, there is no incentive in doing so if there are no transparent mechanisms to evaluate the reliability of elicited forecasts. In this paper, the evaluation of elicited density forecasts is made via scoring rules. A scoring rule is a statistic that quantifies in an ex-post manner the accuracy of released forecasts. A scoring rule is defined as a numerical ranking based on the predictive probability distribution, and the materialized event. A scoring rule could be proper or improper. Proper scoring rules provide the incentive for the forecaster to elicit her best and accurate predictions. On this matter, the implementation of transparent scoring rule frameworks is encouraged to distinguish forecasting skill heterogeneity. Also, the sequential compilation of scoring rule outcomes will allow for future forecast enhancements.

The objective of the proper scoring rules framework is twofold: to evaluate the density forecasts and to incentivize the generation of better—probabilistic—predictions. In terms of evaluation, the scoring rule measures the quality of elicited density forecasts. It also rewards the forecaster for his or her probability predictions and ranks density forecast outcomes. From the perspective of the generation of forecasts, a scoring rule has two purposes: to encourage the forecaster to make careful assessments, and to ensure that the forecaster retains his or her honesty during the process, i.e., to ensure that the forecaster elicits his or her real and best density forecast.

The literature of assessments of probabilistic predictions comes from a long-standing tradition and with a wealth of examples. The pioneering work of Brier (1950) establishes the quadratic score to evaluate probabilistic forecasts using a raining example. Brier (1950) looks into the principle of the proper scoring rules to devise a verification scheme that cannot influence the forecaster in an undesirable way. Even using the proper quadratic score (Brier), Sanders (1963) studies various weather elements and provides evidence that skillful probability statements can indeed be formulated subjectively. Winkler and Murphy (1968) argue that a subjective probability statement quantifies the confidence of an individual in a correct and particular proposition, e.g., an economist's confidence in the statement that "next-year inflation will be two percent." The assessment of the elicitation of probabilities and the discussion of proper scoring rules are widely debated in Savage (1971), which concludes that scoring rules enable us, in principle, to discover people's opinions.

As an alternative to proper scoring rules, the quality, evaluation, and improvement of density forecasts can be assessed through probabilistic forecast calibration procedures. The calibration method makes use of the cumulative probability integral transform to evaluate elicited density forecasts in an ex-post manner. Although calibration and scoring rules are complementary procedures in assessing the quality of forecasts, the scoring rule framework still allows the forecasters to be rewarded for their probability predictions and to directly rank forecasters or forecasting systems. Lichtenstein, Fischhoff, and Phillips (1982) summarize the literature on the calibration of point, and density forecasts up to the 1980s. Dawid (1982) posits a theorem to assess calibration and also discusses recalibration processes and coherence versus calibration. In a classic work, Bunn (1984) also explains the calibration and recalibration procedures of discrete predictive densities, as well as the evaluation of probabilistic forecasts through scoring rules. Examples of calibration and re-calibration assessments to improve univariate density forecasts are found in Kling (1987), Kling and Bessler (1989), Schervish (1989), Bessler and Kling (1990), etcetera. More recently, Diebold, Hahn, and Tay (1999) have extended the univariate calibration assessment to a multivariate framework using examples that involve high-frequency exchange rate density forecasts.

A4.1 Proper scoring rules

The reliability of the predicted density forecasts can be evaluated using proper scoring rules. The scoring rules allow the ranking of density forecast accuracy in assigning a numerical score. The score is based on two characteristics: a) the predictive distribution; and b) the event, state of nature, or numerical value that is realized. Proper scoring rules incentivize the forecaster to elicit his or her best predictive density forecasts and to be honest (Murphy and Epstein 1967). The continuous-ranked probability score (CRPS) evaluates predictive cumulative density functions as proper scoring rules (Gneiting and Raftery 2007). Properties of the proper scoring rules and the definitions of the CRPS are presented in the following paragraphs.

The actual forecaster's predictive probability for the target random variable Y is denoted with $\boldsymbol{p}$ and the forecaster's reported predictive probability with $\boldsymbol{u}$. Both $\boldsymbol{p}$ and $\boldsymbol{u}$ can be probability density functions or probability mass distributions. The scoring rule is a function $S(\boldsymbol{p}, y)$ that assigns a score and maps into real numbers $\mathbb{R}$. y denotes the realization of the random variable, Y. A scoring rule is a (weakly) proper scoring rule—in an *ex-ante* sense—if the forecaster's expected score eliciting his real predictive probability—density or mass—function $\boldsymbol{p}$, $\mathbb{E}_p S(\boldsymbol{p}, Y) = \int S(\boldsymbol{p}, y)p(y)dy$, is smaller than his expected score eliciting any another probability function $\boldsymbol{u}$, $\mathbb{E}_p S(\boldsymbol{u}, Y) = \int S(\boldsymbol{u}, y)p(y)dy$, e.g., Equation (30) will hold for all probability functions $\boldsymbol{p}$ and $\boldsymbol{u}$. Finally, the scoring rule is strictly proper when the equality holds if and only if $\boldsymbol{p} = \boldsymbol{u}$ almost surely (Gneiting 2011).

(30) $\quad \mathbb{E}_p S(\boldsymbol{p}, Y) \leq \mathbb{E}_p S(\boldsymbol{u}, Y).$

If the forecaster prefers to use his true density function $\boldsymbol{p}$, instead of cheating and reporting any other function $\boldsymbol{u}$, the forecaster will be enforced by the proper scoring rule to set $\boldsymbol{u} = \boldsymbol{p}$, i.e., the forecaster obtains his best-expected score for reporting his true predictive probability. Besides that, the negative orientation of (30) concerning the forecast accuracy originates from a conventional change of minimizing the forecaster's expected penalty to maximizing the forecaster's expected score (Bunn 1984). Thus, the lower the score in (30), the better the forecaster is compensated. Hence, the desire to minimize the expected score from a strictly proper scoring rule motivates the forecaster to provide probabilities that are well-calibrated and sharp (Winkler 1986). Proper scoring rules motivate the forecasters to minimize their expected scores only by reporting individual probabilistic forecasts honestly. The use of proper scoring rules in forecast density evaluations minimizes the concerns that assessed subjective probabilities are, in some sense, arbitrary.

The objective of the scoring rule can also be interpreted in an *ex-post sense.* Assume that Y has N_T sequential realizations, $\left\{y_{t+h}\right\}_{t=t_1}^{T}$, for a variety of horizons $h \in \mathscr{H}$. Two different prequential forecasting systems (PFSs) for Y collect the issued probability forecast sequences $\left\{\hat{\boldsymbol{p}}_{t+h}\right\}_{t=t_1}^{T}$, and $\left\{\hat{\boldsymbol{u}}_{t+h}\right\}_{t=t_1}^{T}$, $\forall h \in \mathscr{H}$. They can be compared using the scoring rule $S(\bullet)$ by estimating their average scores, $\bar{S}_T^{\hat{p}}(\bullet) = \frac{1}{T}\sum_{t=t_1}^{T} S(\hat{\boldsymbol{p}}_{t+h}, y_{t+h})$ and $\bar{S}_T^{\hat{u}}(\bullet) = \frac{1}{T}\sum_{t=t_1}^{T} S(\hat{\boldsymbol{u}}_{t+h}, y_{t+h})$, $\forall h \in \mathscr{H}$. Then, the subject evaluating the forecast sequences would prefer the PFS p to the u ($p \succcurlyeq u$) if and only if $\bar{S}_T^{\hat{p}} \leq \bar{S}_T^{\hat{u}}$, at any period $T \in \mathbb{N}_{\geq 1}$ and for all horizon $h \in \mathscr{H}$, and u otherwise (Gneiting 2011).

By definition, a proper scoring rule assumes that the forecaster is risk-neutral. The basic framework of the proper scoring rule assumes a linear utility reward for the agent (Nelson and Bessler 1989). Some rules, such as the quadratic, logarithmic, and spherical ones, are motivated by this risk-neutral assumption. Derivations of the proper scoring rules under the risk-neutral forecaster hypothesis can be seen in Winkler and Murphy (1968), Winkler (1996), Bunn (1984), Krämer (2006), Bickel (2007), Gneiting, Balabdaoui, and Raftery (2007), among others. Non-risk-neutral forecasters might have different implications on the forecaster's incentives to elicit his or her real and honest probability predictions. Examples of risk-averse and risk-lover forecasters are found in Winkler and Murphy (1970), Winkler (1996), Bickel (2007), Johnstone (2007a), Johnstone (2007b), and Johnstone, Jose, and Winkler (2011).

A4.2 Weighted averages

As an alternative to using the arithmetic mean score, $\bar{S}_{N_T}(\bullet)$, the evaluation of density forecasts across time can be studied via the weighted average of individual scores, $\bar{S}_{w,T}(\bullet)$, $\forall T \in \mathbb{N}_{\geq 1}$. The weighted average score of the sequence of issued density forecasts $\{\hat{\boldsymbol{u}}_{t+h}\}_{t=t_1}^{T}$, is denoted by $\bar{S}_{w,T}^{\hat{u}}(\bullet)$, $\forall T \in \mathbb{N}_{\geq 1}$, $h \in \mathscr{H}$ and stated in Equation (31). While cycles and forecast horizons are essential components in the forecasting exercises, weighted average scores incorporate these sequential behaviors providing a personalized mass for each score.

(31) $$\bar{S}_{w,T}^{\hat{u}}\left(\{\hat{\boldsymbol{u}}_{t+h}\}_{t=t_1}^{T}, \{y_{a,t+h}\}_{t=t_1}^{T}\right) = \sum_{t=t_1}^{T} w_{t+h}\, S(\hat{\boldsymbol{u}}_{t+h}, y_{a,t+h}),\ \forall T \in \mathbb{N}_{\geq 1},\ h \in \mathscr{H}.$$

The weights w_{t+h} can be estimated considering some cyclicality or seasonality depending on two primary considerations: i) the exact date when the forecast is released—seasonal component—; and ii) how many horizons ahead the forecast was made for. For instance, a density forecast elicited in January 2017 to predict the annual private consumption of República Bolivariana de Venezuela during 2017 could be penalized or rewarded differently than a density forecast released in December 2017 to predict the same annual event. The December 2017 density forecast would have incorporated useful information of the first three quarters of 2017, while the January 2017 density forecast would lack the use of this information. Using this logic, we assumed that the closer the event is to being materialized, the more the score will be penalized; and the further the event is to being realized, the more the score will be rewarded. In the next subsection, we provide an example of the estimation of weights used for Equation (31).

Monthly elicited forecasts for an annual event. In this subsection, we present an example of the weighted penalization of negatively oriented scores in the CRPS context incorporating monthly-elicited forecasts and annual materialized events. Note that the event materializes at the end of the year. We denote $w_{m,i+h}$ as the weight for the forecast released in month $m \in M = \{1,2,\ldots,12\}$ of year i for horizon $h \in \mathscr{H}$, such that $m = 1$ denotes January and $m = 12$ denotes December. To normalize the weights, we assume that $\sum_m \sum_h w_{m,i+h} = 1$ for all year i. Under this setup, a simple rule to estimate the weights $w_{m,i+h}$, which are used in Equation (31), is presented in (32). Following Equation (32), Figure 7 presents the values of the weights for the sequence of months $\{m_1,\ldots,m_{12}\}$ and a three-year horizons set $\mathscr{H} = \{0,1,2\}$.

$$(32) \qquad w_{m,i+h} = \frac{1+\frac{m-12(h+1)}{12(N_H+1)}}{\sum_{j=1}^{12}\sum_{k=1}^{N_H}\left(1+\frac{j-12(k+1)}{12(N_H+1)}\right)}, \quad \forall i, m \in \{1,2,\ldots,12\}, h \in \left\{0,1,\ldots,H \mid H \in \mathbb{N}_{\geq 0}\right\}.$$

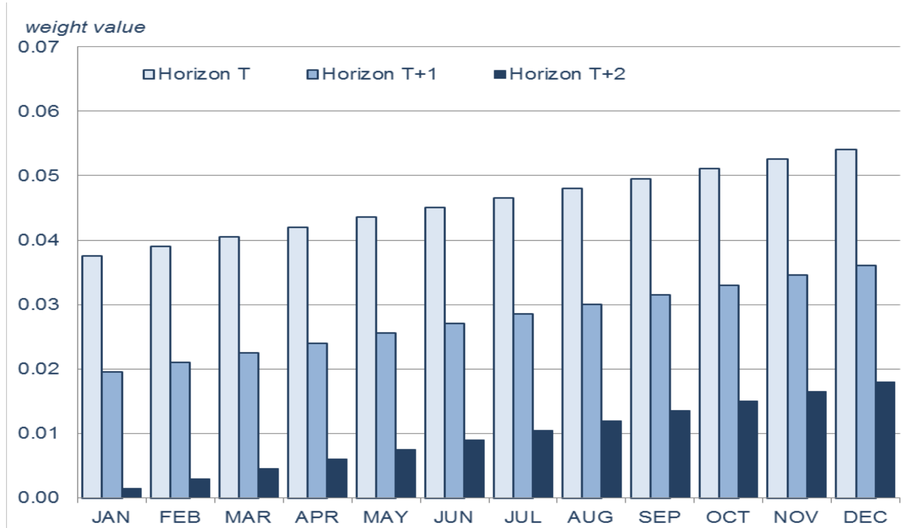

FIGURE 7. EXAMPLE OF WEIGHTS TO ASSESS MONTHLY RELEASED FORECASTS

Note: The weights are for elicited monthly forecasts with three annual horizons. The penalization (weights) to the predictions decreases as the horizon increases. December will be the more penalized month if the forecaster does not elicit accurate annual predictions.

A4.3 THE CONTINUOUS RANKED PROBABILITY SCORE

The continuous-ranked probability score (CRPS) is a proper scoring rule that allows the evaluation and ranking of density forecasts. The literature addressing the CRPS and some of its properties are Epstein (1969), Murphy (1971), Brown (1974), Matheson and Winkler (1976), Von Holstein (1977), Hersbach (2000), Gneiting, Raftery, Westveld III, and Goldman (2005), Kohonen and Suomela (2006), Grimit, Gneiting, Berrocal, and Johnson (2006), and Gneiting, Balabdaoui, and Raftery (2007), among others. For descriptive purposes, the predictive probability density function (PDF) of Y is denoted by $\rho(y)$. It is assumed that any predictive PDF $\rho(y)$ is generated by a PFS. The cumulative distribution function (CDF) for the value y of the random variable Y is denoted by $P(y) = P(Y \leq y)$.

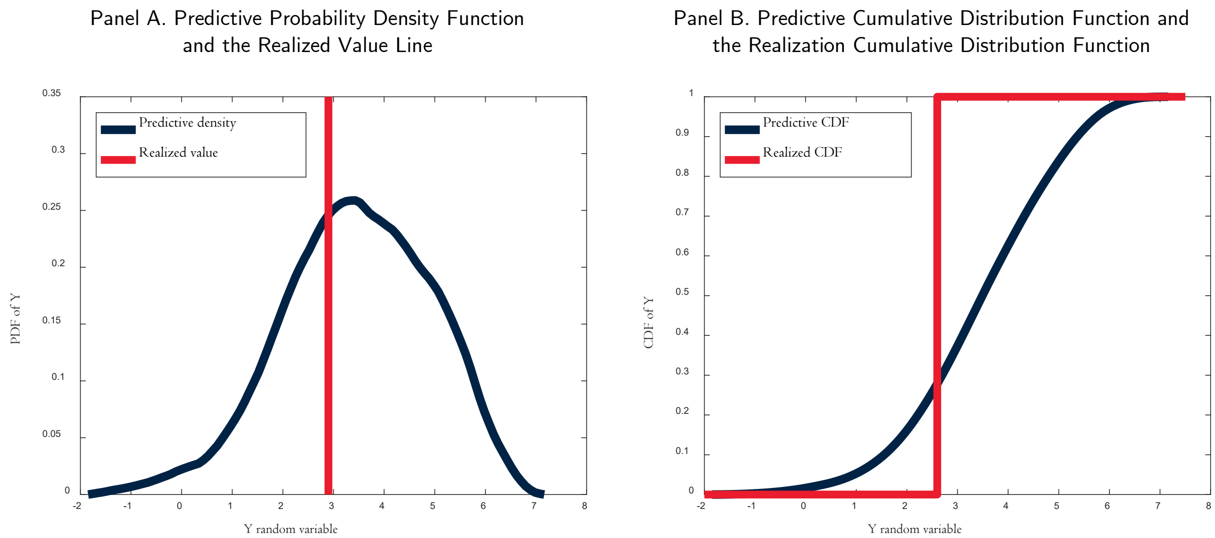

FIGURE 8. PROBABILITY FUNCTIONS OF AN ISSUED DENSITY FORECAST AND THE REALIZED VALUE

On the other hand, the CDF for the value y of the realized event-variable Y_a is denoted as $P_a(y) = P(Y_a \leq y)$. Note that $P_a(y)$ is the indicator function $1_{\{(y-y_a)\geq 0\}}$. Under this notation, the CRPS definition can be stated as presented in (33). Figure 8 shows an example of elicited probability density and cumulative distribution functions for the released predicting variable Y and the realized variable Y_a.

(33) $$CRPS = CRPS(P, y_a) = \int_{-\infty}^{\infty} \left[P(y) - P_a(y)\right]^2 dy.$$

The CRPS ranks the distance between the issued event probabilities, $P(y)$ and the materialized event probability, $P_a(y)$. Then, the CRPS measures the difference between the predicted and the occurred cumulative distributions. The CRPS best and minimal value of zero is only achieved when $P(y) = P_a(y)$, that is, in the case of a perfect categorical (deterministic) foresight. The CRPS has the dimension of the random variable Y, which is entered via the integration over dy. Besides, the CRPS can be understood as the limit of a ranked probability score (RPS)—Epstein (1969) and Murphy (1971)—with an infinite number of classes and each with zero width.

The orientation of the CRPS is negative; that is, we prefer the scores with the smallest values. The lower the CRPS score, the better the prediction characteristics of the forecaster. Most importantly, the CRPS is sensitive to the shape of the distribution outside the realized value y_a. The CRPS score is strictly proper, in the sense that a risk-neutral forecaster will maximize his expected score if his released distribution $P(y)$ agrees with his actual predictive distribution, $G(y)$. Moreover, the evaluation of several density forecasts across time can be studied via the weighted average of individual scores, as pointed out in Equation (31), i.e., $\overline{CRPS}\left(\left\{P(y)_{t+h}\right\}_{t=t_1}^{T}, \left\{y_{a,t+h}\right\}_{t=t_1}^{T}\right) = \sum_{t=t_1}^{T} w_{t+h}\, CRPS\left(P(y)_{t+h}, y_{a,t+h}\right)$ for any period $T \in \mathbb{N}_{\geq 1}$, and any horizon $h \in \mathscr{H}$.

Appendix A5. The two-piece normal distribution (Online Appendix)

The TPN distribution for the random variable Y is denoted as $Y \sim TPN\left(\mu_Y, \sigma_{1,Y}, \sigma_{2,Y}\right)$. The TPN assumption allows the modeling of asymmetric distributional behaviors. Under this distribution, the variable Y is fully depicted using three parameters. For instance, the TPN can be entirely described by the mode, μ, the variance, σ, and the skewness, γ. Alternatively, it can be defined via the mode, μ, the downside standard deviation, σ_1, and the upside standard deviation, σ_2. Note that $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^{++}$, $\gamma \in \mathbb{R}$, $\sigma_1 \in \mathbb{R}^{++}$, and $\sigma_2 \in \mathbb{R}^{++}$. The TPN probability density function for a random variable Y is defined in Equation (34), while its mean, variance, and skewness are stated in (35), (36), and (37), respectively. Under (34), when $\sigma_1 > \sigma_2$, the distribution of Y is skewed to the left (negative or downside skewness), which also means $Prob(X \leq \mu) > 0.5$. When the distribution of Y is skewed to the right (positive or upside skewness), this means that $\sigma_1 < \sigma_2$ and $Prob(X \leq \mu) < 0.5$.

$$(34) \qquad f(y;\mu,\sigma_1,\sigma_2) = \begin{cases} f_1(\bullet) = \sqrt{2/\pi}\,(\sigma_1+\sigma_2)^{-1} \exp\left(-\frac{(y-\mu)^2}{2\sigma_1^2}\right) I_{\{(y-\mu)<0\}}. \\ f_2(\bullet) = \sqrt{2/\pi}\,(\sigma_1+\sigma_2)^{-1} \exp\left(-\frac{(y-\mu)^2}{2\sigma_2^2}\right) I_{\{(y-\mu)\geq 0\}}. \end{cases}$$

$$(35) \qquad \mathbb{E}[Y] = \mu + \sqrt{2/\pi}\,(\sigma_2 - \sigma_1).$$

$$(36) \qquad var[Y] = (1 - 2/\pi)(\sigma_2-\sigma_1)^2 + \sigma_1\sigma_2.$$

$$(37) \qquad \gamma[Y] = \mathbb{E}\left[(Y-\mathbb{E}[Y])^3\right] = \sqrt{2/\pi}\,(\sigma_2-\sigma_1)\left[\left(\tfrac{4}{\pi}-1\right)(\sigma_2-\sigma_1)^2 + \sigma_1\sigma_2\right].$$

The TPN distribution easily allows the estimation of the probability of Y between the values a_1 and a_2, $Prob(a_1 \leq Y \leq a_2)$. This probability is presented in (38), where $\Phi(\bullet)$ is the standard normal cumulative distribution function, and $I(\bullet)$ is an indicator function.

$$(38) \qquad Prob(a_1 \leq Y \leq a_2) = \begin{cases} \frac{2\sigma_1}{(\sigma_1+\sigma_2)}\left[\Phi\left(\frac{a_2-\mu}{\sigma_1}\right) - \Phi\left(\frac{a_1-\mu}{\sigma_1}\right)\right] I_{\{a_1 \leq a_2 \leq \mu\}}. \\ \frac{2\sigma_2}{(\sigma_1+\sigma_2)}\left[\Phi\left(\frac{a_2-\mu}{\sigma_2}\right) - \Phi\left(\frac{a_1-\mu}{\sigma_2}\right)\right] I_{\{\mu \leq a_1 \leq a_2\}}. \\ \left(\int_{a_1}^{\mu} \quad f_1(\bullet)dy + \int_{\mu}^{a_2} \quad f_2(\bullet)dy\right) I_{\{a_1 \leq \mu \leq a_2\}} \ . \end{cases}$$

Under the TPN assumption, the CRPS rule takes on the closed form presented in (39). This closed representation is introduced by Gneiting and Thorarinsdottir (2010) where $\rho(\bullet)$ stands for the standard normal probability density function.

$$(39) \qquad CRPS(P, y_a) = \int_{-\infty}^{\infty}\left[P(y) - P_a(y)\right]^2 dy.$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{y} f(x)dx - I_{\{(y-y_a)\geq 0\}}\right]^2 dy.$$

$$= \int_{-\infty}^{\infty}\left(\frac{2\sigma_1}{(\sigma_1+\sigma_2)}\left[\Phi\left(\frac{y-\mu}{\sigma_1}\right)\right] I_{\{y \leq \mu\}}.\right.$$

$$\left. + \frac{2\sigma_2}{(\sigma_1+\sigma_2)}\left[1 - \Phi\left(\frac{y-\mu}{\sigma_2}\right)\right] I_{\{\mu \leq y\}} - I_{\{(y-y_a)\geq 0\}}\right)^2 dy$$

$$= \left[\frac{4\sigma_1^2}{\sigma_1+\sigma_2}\left[\frac{y_a-\mu}{\sigma_1}\Phi\left(\frac{y_a-\mu}{\sigma_1}\right) + \rho\left(\frac{y_a-\mu}{\sigma_1}\right)\right] - (y_a - \mu)\right.$$

$$\left. + \frac{2}{\sqrt{\pi}}\ \frac{\sqrt{2}\sigma_2(\sigma_2^2-\sigma_1^2)-\left(\sigma_1^3+\sigma_2^3\right)}{(\sigma_1+\sigma_2)^2}\right] I_{\{y_a \leq \mu\}}$$

$$+ \left[\frac{4\sigma_2^2}{\sigma_1+\sigma_2}\left[\frac{y_a-\mu}{\sigma_2}\Phi\left(\frac{y_a-\mu}{\sigma_2}\right) + \rho\left(\frac{y_a-\mu}{\sigma_2}\right)\right] + \frac{\left[(\sigma_1-\sigma_2)^2-4\sigma_2^2\right](y_a-\mu)}{(\sigma_1+\sigma_2)^2}\right.$$

$$\left. + \frac{2}{\sqrt{\pi}}\ \frac{\sqrt{2}\sigma_1(\sigma_1^2-\sigma_2^2)-\left(\sigma_1^3+\sigma_2^3\right)}{(\sigma_1+\sigma_2)^2}\right] I_{\{\mu \leq y_a\}}.$$

APPENDIX A6. DATA TRANSFORMATIONS AND LOG-NORMALITY ASSUMPTIONS (ONLINE APPENDIX)

The original series of absolute deviation of inflation targets and term spread are denoted via X_π and X_{TS}, respectively. The original series are assumed to behave with means μ_{X_π}, $\mu_{X_{TS}}$, standard deviations σ_{X_π}, $\sigma_{X_{TS}}$, and skewness (third central moment) γ_{X_π}, $\gamma_{X_{TS}}$, respectively. For smoothing and variability comparison purposes, we construct the variables that enter the Equation (1) as $Z_\pi = \frac{X_\pi}{\hat{\sigma}_{X_\pi}}$, and $Z_{TS} = \frac{X_{TS}}{\hat{\sigma}_{X_{TS}}}$. The statistic $\hat{\sigma}_s$ for any $s \in \{X_\pi, X_{TS}\}$ can be approximated with the corresponding sample standard deviation.

The original price return between periods t_0 and t_t for the oil and the S&P 500 index stock series, r_{oil} and r_{SPX} , are assumed to follow a normal distribution, i.e., $r_{oil} = \ln \frac{S_{oil,t_f}}{S_{oil,t_0}} \sim N\left(\left(\mu_{r_{oil}} - \frac{\sigma^2_{r_{oil}}}{2}\right)(t_f - t_0), \sigma^2_{r_{oil}}(t_f - t_0)\right)$ and $r_{SPX} = \ln \frac{S_{SPX,t_f}}{S_{SPX,t_0}} \sim N\left(\left(\mu_{r_{SPX}} - \frac{\sigma^2_{r_{SPX}}}{2}\right)(t_f - t_0), \sigma^2_{r_{SPX}}(t_f - t_0)\right)$. Thus, $\ln S_{k,t_f}$ is distributed as $N\left(\ln S_{k,t_0} + \left(\mu_{r_k} - \frac{\sigma^2_{r_k}}{2}\right)(t_f - t_0), \sigma^2_{r_k}(t_f - t_0)\right)$ for all $k \in \{oil, SPX\}$. This price return distributional assumption implies that the stock prices can be approximated with log-normal distribution, $S_{k,t_f} \sim logN\left(\ln S_{k,t_0} + \left(\mu_{r_k} - \frac{\sigma^2_{r_k}}{2}\right)(t_f - t_0), \sigma^2_{r_k}(t_f - t_0)\right)$ for all $k \in \{oil, SPX\}$. The stock prices behave following the moments specified in (40)–(42).

$$(40) \qquad \mathbb{E}\left[S_{k,t_f}\right] = e^{\left(\ln S_{k,t_0} + \left(\mu_{r_k} - \frac{\sigma^2_{r_k}}{2}\right)(t_f - t_0) + \frac{\sigma^2_{r_k}}{2}(t_f - t_0)\right)}$$

$$= S_{k,t_0}\, e^{\left(\mu_{r_k}(t_f - t_0)\right)} \qquad \text{for all } k \in \{oil, SPX\}.$$

$$(41) \qquad var\left[S_{k,t_f}\right] = \left[e^{\left(\sigma^2_{r_k}(t_f - t_0)\right)} - 1\right] \bullet e^{\left(2 \ln Z_{k,t_0} + 2\left(\mu_{r_k} - \frac{\sigma^2_{r_k}}{2}\right)(t_f - t_0) + \sigma^2_{r_k}(t_f - t_0)\right)}$$

$$= S_{k,t_0}^2 \bullet \left[e^{\left(\sigma_{r_k}^2 (t_f - t_0)\right)} - 1\right] \bullet e^{\left(2\,\mu_{r_k} (t_f - t_0)\right)} \qquad \text{for all } k \in \{oil, SPX\}.$$

$$(42) \qquad \gamma\left(S_{k,t_f}\right) = \left[e^{\left(\sigma_{r_k}^2 (t_f - t_0)\right)} + 2\right] \sqrt{e^{\left(\sigma_{r_k}^2 (t_f - t_0)\right)} - 1} \qquad \text{for all } k \in \{oil, SPX\}.$$

The Black-Scholes-Merton model for option pricing assumes the log-normality distribution of the stock prices. This assumption allows us to consistently proxy the *ex-ante variance* and *ex-ante skewness* of the stock prices. The series of *ex-ante variance* and *ex-ante skewness* of the stock prices are recovered using Equations (41) and (42). First, note that the implied-volatility $\sigma_{imp,T+h}$ estimated from the Black-Scholes-Merton model of option pricing is the *ex-ante* standard deviation of the stock price return $\sigma_{r,T+H}$. Bloomberg assumes the Black-Scholes-Merton model to computationally proxy the implied volatility from option prices. As a consequence, $\sigma_{r,T+H}$ is approached using the implied volatility of the stock prices provided by Bloomberg. Second, the Black-Scholes-Merton model assumes a riskless portfolio consisting of a position in the option and a position in the stock prices. In the absence of arbitrage opportunities, the return from the portfolio must be the risk-free interest rate, i. Thus, the *ex-ante* expected return—implied return—$\mu_{r_k,T+h}$, is used as a proxy via the risk-free interest rate, i, which is taken as the 5-year Treasury Note Yield at constant maturity collected from Haver Analytics.

Finally, we transform the stock price series as in the case of the construction of the series Y, Z_π, Z_{TS}. This transformation aims to smooth and benchmark the variation of the risk factors in Equation (1) as explicitly stated in (2). The risk factor series of crude oil and the S&P 500 index prices, Z_{oil} and Z_{SPX}, which enter the Equation (1) are generated using the transformation presented in (43).

$$(43) \qquad Z_k = \frac{S_k}{\sqrt{\widehat{var}\left[S_k\right]}} \,.$$

Where $\widehat{var}\left[S_k\right]$ is the sample mean of the series $\left\{var\left[S_{k,t}\right]\right\}_{t=t_0}^{T}$.