Assessing Forecast Uncertainty

An Information Bayesian Approach

Fabian Mendez-Ramos
**Abstract**

Regardless of the field, forecasts are widely used and yet assessments of the embedded uncertainty—the magnitude of the downside and upside risks of the prediction itself—are often missing. Particularly in policy-making and investment, accounting for these risks around baseline predictions is of outstanding importance for making better and more informed decisions. This paper introduces a procedure to assess risks associated with a random phenomenon. The methodology assigns probability distributions to baseline-projections of an economic or social random variable—for example gross domestic product growth, inflation, population growth, poverty headcount, and so forth—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 methodology also decomposes the variance and skewness of the predictive distribution accounting for the shares of selected risk factors. The procedure relies on a Bayesian information-theoretical approach, which allows the inclusion of judgment and forecaster expertise. For reliability purposes and transparency, the paper also evaluates the constructed density forecasts assigning a score. The continuous ranked probability score is used to assess the prediction accuracy of elicited density forecasts. The selected score incentivizes the forecaster to provide its true and best predictive distribution. An empirical application to forecast world gross domestic product growth is used to test the Bayesian entropy methodology. Predictive variance and skewness of world gross domestic product growth are associated with ex-ante information of four risk factors: term spreads, absolute deviations of headline inflation targets, energy prices, and the Standard and Poor’s 500 index prices. The Bayesian entropy technique is benchmarked with naïve-generated density forecasts that utilize information from historical forecast errors. The results show that the Bayesian density forecasts outperform the naïve-generated benchmark 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.
Assessing Forecast Uncertainty: An Information Bayesian Approach

Fabian Mendez-Ramos
The World Bank

JEL Classification: C11, C53, C54

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

I am grateful to Israel Osorio Rodarte, Raul Ibarra Ramirez, 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. Contact address: Fabian Mendez Ramos, Development Research Group, The World Bank, Level 3, Sasana Kijang, No. 2, Jalan Dato’ Onn, 50480, Kuala Lumpur, Malaysia; Email: fmendezramos@worldbank.org.
2 From here onwards, density forecast, probabilistic forecast, or fan chart are used interchangeably.
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 true and best density forecast.

A density forecast is a risk management tool that helps depict and quantify risks, expected returns, or measures related to the risks associated with the forecasted variable. Density forecasts vary in their scope and application. Density—probabilistic—forecasts are more informative and more valuable to decision makers than categorical—“interest rates will not rise this year”—or mean forecasts given that they explicitly recognize and quantify uncertainty (Winkler 1996). For instance, in modern financial risk management, tools like portfolio Value-at-Risk (VaR) forecasts and conditional VaR assessments make use of predictive densities; e.g., in Andersen, Bollerslev, Diebold and Labys (2003). For policy-making purposes, the density forecasts help contextualize the risks associated with a policy variable and its outcomes. The hazard is that the decision-takers might not process the information embedded in the generated probabilistic forecasts in proper forms, e.g., they could take decisions with very large or relatively very small risk.

Although a few institutions and companies have achieved some progress in generating probabilistic forecasts, they are still scarce and—when available—are not clearly evaluated. Some central banks and financial institutions have implemented predictive densities to assess inflation, GDP growth, market volatility, exchange rates, etc. Examples include the International Monetary Fund (Kannan and Elekdag 2009), Bank of England (2013, 2014, Inflation Report, several volumes), Sveriges Riksbank (2014, Monetary Policy Report, several volumes), the Reserve Bank of India (Banerjee and Das 2011), European Central Bank (2013, 2014, Macroeconomic Projections, several volumes), Banco Central de Chile (2014, Informe de Política Monetaria, several volumes), Bank of Mexico (2015, Quarterly Inflation Report, several volumes), Banco Central Do Brasil (2014, Inflation Report, several volumes), Central Bank of the Republic of Turkey (2014, Inflation Report, several volumes), among many other central banks and forecasting agencies.

Information extracted from surveys on expectation and implied statistics of forecasting models can be useful in assessing the uncertainty in predictive density forecasts. While the use of ex-ante information to predict uncertainty 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. On the other hand, using the same SPF data, Giordani and Söderlind (2003) find that disagreement among forecasters is a good proxy for inflation and for output growth uncertainties.

This Bayesian framework allows the incorporation of subjective and prior information on some parameters and statistics. It is known that forecasters disagree in beliefs and in assigning prior distributions. These disagreements in beliefs might be interpreted as a reflection of underlying uncertainty (Al-Najjar and Shmaya 2015). On that account, the methodology is flexible to exploit heterogeneity in beliefs embedded in prior probability distributions evaluating the accuracy of generated predictive density forecasts. 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 (Dawid 1984) combined with statistical “proper scoring rules” (Savage 1971) is applied in this research to assess the dynamic performance of the Bayesian entropy-generated density forecasts.

The literature of subjective probabilistic assessments 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 any undesirable way. Using the quadratic (Brier) score for various weather elements, Sanders (1963) provides evidence that skillful probability statements can indeed be formulated subjectively. Winkler and Murphy (1968) discuss properties of proper scoring rules to evaluate probability forecasts. Winkler and Murphy (1968) argue that a subjective probability statement quantifies the confidence of an individual in a true and particular proposition, e.g., an economist’s confidence in the statement that “next-year inflation will be two percent”. The assessment of 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 technique makes use of the cumulative probability integral transform. 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. Bunn (1984) 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), etc. More recently, Diebold, Hahn and Tay (1999) have extended the univariate calibration assessment to a multivariate framework using an example which involves high frequency exchange rate density forecasts.

In this paper, the author introduces the Bayesian entropy (BEN) procedure to generate density forecasts for solving an information-theoretical problem, which uses ex-post and ex-ante information of the studied variable and its risk factors. The empirical application generates density forecasts of world GDP growth associating selected risk factors: terms spreads, absolute deviations of headline inflation targets, energy prices, and the S&P 500 index prices. There are two key advantages of using the BEN procedure. First, prior information of the underlying unknown and unobserved parameters—or latent variables—can be incorporated to help the system find less noisy levels of information. Second, the Bayesian entropy procedure recovers and processes information when the underlying sampling model is partly or incorrectly known, and the data is limited, partial or incomplete (Golan, Judge and Miller 1996).

The BEN procedure is described in the following sections of the paper. Distributional assumptions, econometric considerations, unobserved statistics, prior information, and the Bayesian entropy setup are explained in Section 2 of this paper. Proper scoring rules that underpin the evaluation of density forecasts using the continuous ranked probability score are presented in Section 3. The BEN technique is tested in Section 4 through an empirical application that generates predictive world GDP growth densities. The practical exercise uses a variable that faces—ex-ante and ex-post—concerns of being incorrectly known,
incorrectly quantified, or both. Risk factors, prior values, generated density forecast series, and their scores are also described in Section 4. Finally, the main conclusions of this study and further research are presented in Section 5.

2 The Model

The presented BEN technique allows for the quantification of risks in the prediction of a studied variable. The BEN 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, etc. The generation of the BEN density forecasts uses two categories of constraints: behavioral and statistical ones. The behavioral constraints rely on the assumption that parametric distributions depict the density forecasts. The statistical—consistency—constraint relies on formal probability statements that model random variables. Even though several parametric and non-parametric distributions are available, the selected distribution should particularly facilitate the estimation of dispersion and skewness statistics.

2.1 Distributional assumptions

For the purpose of generating the predictive probability density function of the target variable, the author relies on the two-piece normal (TPN) function. The forecasted 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. Furthermore, the TPN assumption permits and facilitates the dynamic modeling and estimation of predictive density forecasts under a proposed minimal-noise optimization setup. While the TPN allows the modeling of the bias in the distribution, the TPN assumption could be—in some cases—a real restriction, especially for variables with very high or very low levels of excess kurtosis, i.e., leptokurtic and platykurtic shapes, respectively. The functional form and details of the TPN distribution are presented in Appendix 1.

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. Ex-ante dispersion and skewness information are inputs to identify the forecast uncertainty dimensions of $Y$. This is complemented with an assessed forecast mode—or most likely outcome of $Y$—and the TPN distributional assumption. The ex-ante dispersion magnitude that the forecaster elicits reflects his level of knowledge and confidence on the likelihood around central expected outcomes. The ex-ante skewness of the predicted variable, $Y$, 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

---

3 Alternatively, the forecasted variable can be called target, focus, or predicted variable.
central to properly quantifying the upside and downside uncertainty 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 provide a key understanding of the future behavior of $Y$, complementing the decision-making process under uncertainty.

2.2 The linear assumption

The BEN 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 information 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, $Z \subseteq \Theta$. The selection of the subset of risk factors, $Z$, might depend on empirical results and theoretical underpinnings associated with $Y$. Note that $\Theta$ denotes the complete set of risk factors; $\Theta = \{Z, Z^c\}$, where $Z^c$ states for the complement of $Z$. The forecasted dispersion and skewness measures of $Y$ are assumed to depend on—ex-post and ex-ante—information of $Z$. The basic idea is that markets perform not only using ex-post information, they 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 BEN procedure looks to exploit the potential uneven and noisy trajectories of $Z$ and $Y$ using a minimization information-theoretical setup. The combination of the ex-post and ex-ante information of $Z$ and the ex-post information of $Y$ are key inputs in the BEN procedure, which estimates ex-ante uncertainty on $Y$. The ex-ante information of $Z$, compiled from a variety of theoretical models and from assorted survey sources, is key to—and complements—the BEN procedure in estimating the ex-ante uncertainty on $Y$.

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

\begin{equation}
Y_t = a_1 Z_{1t} + \cdots + a_R Z_{N_Rt} + \varepsilon_t.
\end{equation}

To describe ex-post and ex-ante categories of information, let us denote $N_T$ as the current period. The three main categories of information in the BEN procedure are: i) ex-post information of the target variable, $\{Y_t\}_{t=t_1}^{N_T}$; ii) ex-post information of the associated risk factors, $\{Z_t\}_{t=t_1}^{N_T}$; and iii) ex-ante available
information of the risk factors, \( \{Z_{t+h}\}_{t=t_1}^{N_T} \), for some horizons \( h \in H \), where \( H \) is a strictly monotonically increasing sequence, \( H = \{h_1, h_2, ..., N_H | h_1 \in \mathbb{N}_{\geq 0}, ..., N_H \in \mathbb{N}_{\geq 0}, h_1 < h_2 < ... < N_H \} \).

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

The BEN procedure does not focus on predicting \( \mathbb{E}(Y_{t+h} | \mathcal{F}_t) \) or \( \text{mode}(Y_{t+h} | \mathcal{F}_t) \); these are taken as a given. The argument is that there are many models that do a good job of predicting central measures under structural and economic assumptions and statistical constraints. The BEN method hence focuses on predicting two measures of uncertainty of \( Y \): the ex-ante conditional variance and skewness, \( \text{var}(Y_{t+h} | \mathcal{F}_t) \) and \( \gamma(Y_{t+h} | \mathcal{F}_t) \), respectively.

The key advantage of the BEN method involves the combination of ex-post data of \( Y \) with all the available ex-ante and ex-post information of \( Z \) at any period \( t \in T \) and for any available horizons, \( h \in H \). The information set \( \mathcal{F}_{N_T} \) contains—at least—the following three series: \( \{Y_t\}_{t=t_1}^{N_T} \), \( \{Z_t\}_{t=t_1}^{N_T} \), and \( \{Z_{t+h}\}_{t=t_1}^{N_T} \), for any horizon \( h \in H \). Note that \( \{Z_{t+h}\}_{t=t_1}^{N_T} \) denotes the ex-ante series of the risk factor \( Z_r \in Z \), for any horizon \( h \in H \).

The ex-ante series of the variance and the skewness measure of the risk factors are denoted by \( \{\text{var}(Z_{t+h})\}_{t=t_1}^{N_T} \) for all \( h \in H \), and \( \{\gamma(Z_{t+h})\}_{t=t_1}^{N_T} \) for all \( h \in H \). The ex-ante conditional variance and skewness of \( Y \) are represented by \( \text{var}(Y_{t+h} | \mathcal{F}_t) \) and \( \gamma(Y_{t+h} | \mathcal{F}_t) \), respectively, for any period \( t \in T \), and horizon \( h \in H \). The conditional variance and skewness measures of \( Y \) are econometrically approximated as presented in Proposition 1 and Proposition 2. These propositions denote the decompositions of the dispersion and skewness of the random variable \( Y \).

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

\[
(2) \quad \text{var}(\hat{Y}_{t+h} | \mathcal{F}_t) = \sum_{r \in R} \alpha_r^2 \text{var}(Z_{t+h}) + \epsilon_{a_r^2}^2 + \epsilon_{\sigma_r^2}^2, \quad \forall t \in T, h \in H.
\]
PROPOSITION 2: Under the linear assumption in (1), the predicted conditional skewness of the target variable is a linear combination of the risk factor skewness plus an error term as defined in (3).

\[
\gamma(\tilde{Y}_{t+h} | F_t) = \sum_{r \subseteq R} a_r^3 \gamma(2_{t, t+h})^\text{ex-ante skewness measure of the target variable} + \epsilon_{\gamma, t+h}^3 \text{ex-ante skewness measure of the risk-factor } Z_r , \forall t \in T, h \in H.
\]

Note that \(\gamma(\cdot)\) 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 factors. See also the proof of Proposition 1 and Proposition 2 in the Appendix section.

2.3 Unobserved statistics

The BEN procedure mainly looks to predict the variance and skewness of \(Y\) for period \(N_T + h\), for any horizon \(h \in H\). Note that \(N_T\) is defined as the current period. However, the ex-ante information of these statistics is uncertain or unobserved, i.e., the modeler—the author—does not have data for the series \(\{\text{var}(\tilde{Y}_{t+h} | F_t)\}_{t=t_1}^{N_T-1}\) and \(\{\gamma(\tilde{Y}_{t+h} | F_t)\}_{t=t_1}^{N_T-1}\) for all horizon \(h \in H\). This restriction complicates the econometric estimation of (2) and (3). In this regard, the BEN procedure looks to recover three types of information while estimating (2) and (3): i) the set of coefficients \(A = \{a_1, a_2, \ldots, a_R\}\); ii) the ex-ante information \(\{\text{var}(\tilde{Y}_{t+h} | F_t)\}_{t=t_1}^{N_T-1}\) and \(\{\gamma(\tilde{Y}_{t+h} | F_t)\}_{t=t_1}^{N_T-1}\), \(\forall h \in H\); and iii) the ex-ante information on the current period \(N_T\), \(\text{var}(\tilde{Y}_{N_T+h} | F_{N_T})\) and \(\gamma(\tilde{Y}_{N_T+h} | F_{N_T})\), \(\forall h \in H\).

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). Clearly, for consistency purposes, (2) and (3) must be estimated simultaneously. Thus, the BEN 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, \(Y\).

2.4 Prior information and estimation steps

The BEN procedure incorporates prior information related to \(Z\) and \(Y\). The use of prior information is simply 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 the BEN procedure, prior knowledge is used to condition on what is known. Therefore, prior sampling information of the underlying phenomena is incorporated, as well as the
perceptions of the forecaster. This subjective information about the distribution of a phenomena could be symmetric or skewed, although in any case they should always be coherent (De Finetti 1937).

Under the above distributional and linear assumptions, the econometric steps for estimating (2) and (3) have to deal with two main restrictions: i) the ex-ante variance and skewness of the predicted variable are unobserved or uncertain; ii) there might be empirical evidence, previous literature, or beliefs affecting the sign of the coefficients in equations (2) and (3). To deal with these major constraints, the BEN procedure solves a minimization information-theoretical setup where prior information is integrated. The BEN follows an information-theoretical econometric method, as presented in Golan, Judge and Miller (1996), and Judge and Mittelhammer (2011). Five steps are designed to describe the solution of the minimization information-theoretical procedure:

1. Prior mean values are set for each of the risk factor coefficients in the set \( A^0 = \{ a_1^0, a_2^0, \ldots, a_R^0 \} \).
2. Prior ex-ante dispersion and ex-ante skewness series, \( \{ \text{var}^0(\hat{y}_{t+h}|F_t) \}_{t=t_1}^{N_T} \) and \( \{ \gamma^0(\hat{y}_{t+h}|F_t) \}_{t=t_1}^{N_T} \) are estimated via constrained OLS for all horizons \( h \in H \).
3. Prior mean values are selected for the error terms, \( \{ e_{\sigma_{t+h}} \}_{t=t_1}^{N_T} \) and \( \{ e_{\gamma_{t+h}} \}_{t=t_1}^{N_T} \), \( \forall h \in H \).
4. Prior probability mass distributions are set for error terms that account for the noise between the above-selected prior mean values and their corresponding posterior values.
5. Estimation of posterior mean values of the following coefficients, statistics and error terms via a minimization information-theoretical setup: i) \( A = \{ a_1, a_2, \ldots, a_R \} \); ii) \( \{ \text{var}(\hat{y}_{t+h}|F_t) \}_{t=t_1}^{N_T} \), \( \{ \gamma(\hat{y}_{t+h}|F_t) \}_{t=t_1}^{N_T} \), \( \forall h \in H \); and iii) \( \{ e_{\sigma_{t+h}} \}_{t=t_1}^{N_T} \) and \( \{ e_{\gamma_{t+h}} \}_{t=t_1}^{N_T} \), \( \forall h \in H \).

The first step implies the definition of the range of outcomes that the—posterior—coefficients can take. To accomplish this objective, the uncertain and unobserved parameters, \( A = \{ a_1, a_2, \ldots, a_R \} \), in (2) and (3), are modeled as posterior values. The BEN procedure assumes that the posterior value of each parameter, \( a_r \), depends on its estimated prior, \( \hat{a}_r^0 \), and an error mean term, \( \xi_{a_r} \). As presented in (4), the association of the prior coefficient and its error term is assumed to follow an additive manner.

\[(4) \quad a_r = \hat{a}_r^0 + \xi_{a_r} \]

Before getting the prior coefficients \( \{ \hat{a}_1^0, \hat{a}_2^0, \ldots, \hat{a}_R^0 \} \), initial prior mean values, \( A^0 = \{ a_1^0, a_2^0, \ldots, a_R^0 \} \), are assigned by the forecaster based on empirical evidences, previous literature, or beliefs. Strong beliefs or empirical evidence affect and constrain the sign of the coefficients \( A = \{ a_1, a_2, \ldots, a_R \} \). For instance, if the prior evidence or theory suggests that the risk factor \( Z_r \) is strongly and positively associated with the target variable, \( Y \), then, the posterior mean value of the coefficient \( a_r \) is expected to be positive. In this regard, for rationality, the posterior and the prior mean values of the coefficient \( a_r \) and \( \hat{a}_r^0 \), respectively, should be positive. In this example, to accomplish the restriction \( \hat{a}_1^0 + \xi_{a_1} > 0 \), the mean error term, \( \xi_{a_1} \), is constrained into the open space \( (-|\hat{a}_r^0|, \infty^+) \), i.e., \( \xi_{a_1} \in (-|\hat{a}_r^0|, \infty^+) \), for all selected prior \( \hat{a}_r^0 \in \mathbb{R}^1 \).

The prior trajectories of the ex-ante variance and skewness of the forecasted variable, \( Y \), are estimated in the second step. This step involves the estimation of a system of equations via constrained ordinary least
The restrictions are defined as prior mean value of the coefficients in the previous step, $A^0 = \{a_1^0, a_2^0, ..., a_R^0\}$, and play a major role in this stage. The prior mean values, $A^0 = \{a_1^0, a_2^0, ..., a_R^0\}$, are used as inputs to proxy the prior series $\left\{\text{var}^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ and $\left\{\gamma^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ for all horizons $h \in H$.

Before obtaining the estimated priors of the ex-ante statistics $\left\{\text{var}^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ and $\left\{\gamma^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$, the BEN procedure initially estimates the series $\left\{\text{var}^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ and $\left\{\gamma^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ using equations (5) and (6). The forecaster might have options to proxy initial prior mean values of $\left\{\text{var}^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ and $\left\{\gamma^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$, other than using equations (5) and (6). These additional options for guessing priors could originate from previous studies where ex-ante information, ex-post analysis, or combinations of ex-post and ex-ante information are considered.

$$\begin{align*}
\text{(5)} & \quad \text{var}^0(\hat{Y}_{t+h}|F_t) = \sum_{r \in R}(a_r^0)^2 \text{var}(\hat{Z}_{r,t+h}|F_t), \quad \forall t \in T, h \in H. \\
\text{(6)} & \quad \gamma^0(\hat{Y}_{t+h}|F_t) = \sum_{r \in R}(a_r^0)^3 \gamma(\hat{Z}_{r,t+h}|F_t), \quad \forall t \in T, h \in H.
\end{align*}$$

Once the prior series $\left\{\text{var}^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ and $\left\{\gamma^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ are approximated via equations (5) and (6), the BEN procedure solves the constrained ordinary least square problem defined through equations (7) – (11).

$$\begin{align*}
\text{(7)} & \quad \text{Min} \quad \left\{ \sum_{h=h_1}^{h_1} \sum_{t=t_1}^{t_1} \left( \text{var}^0(\hat{Y}_{t+h}|F_t) - \gamma^0(\hat{Y}_{t+h}|F_t) \right)^2 \\
& \quad + \sum_{h=h_1}^{h_1} \sum_{t=t_1}^{t_1} \left( \gamma^0(\hat{Y}_{t+h}|F_t) - \gamma^0(\hat{Y}_{t+h}|F_t) \right)^2 \right\}.
\end{align*}$$

Subject to

$$\begin{align*}
\text{(8)} & \quad \text{var}^0(\hat{Y}_{t+h}|F_t) = \sum_{r \in R}(a_r^0)^2 \text{var}(\hat{Z}_{r,t+h}|F_t), \quad \forall t \in T, h \in H. \\
\text{(9)} & \quad \gamma^0(\hat{Y}_{t+h}|F_t) = \sum_{r \in R}(a_r^0)^3 \gamma(\hat{Z}_{r,t+h}|F_t), \quad \forall t \in T, h \in H. \\
\text{(10)} & \quad a_1^0 \leq 0, ..., a_R^0 \leq 0. \\
\text{(11)} & \quad (a_r^0)^2 \text{var}(\hat{Z}_{r,t+h}|F_t) \leq b_r, \quad \forall r \in R, t \in T, h \in H.
\end{align*}$$

The prior estimated series $\left\{\text{var}^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ and $\left\{\gamma^0(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ are assumed to be linearly related with the posterior series $\left\{\text{var}(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ and $\left\{\gamma(\hat{Y}_{t+h}|F_t)\right\}_{t=t_1}^{NT}$ for any horizon $h \in \mathbb{N}_{\geq 1}$ as presented in equations (12) and (13). $\xi_{\sigma^2}$ and $\xi_{\gamma}$ denote error terms.

$$\begin{align*}
\text{(12)} & \quad \text{var}(\hat{Y}_{t+h}|F_t) = \text{var}^0(\hat{Y}_{t+h}|F_t) + \xi_{\sigma^2,t+h}, \quad \forall t \in T, h \in H; \text{ and}
\end{align*}$$
The prior mean values of the error terms in equations (2) and (3) are set in the third step. The posterior values of the error terms, \( \varepsilon_\sigma \) and \( \varepsilon_\gamma \), are assumed to depend linearly on its corresponding prior, \( \varepsilon_{\sigma^2}^0 \) and \( \varepsilon_\gamma^0 \), and on the error mean terms, \( \xi_{\varepsilon_\sigma^2} \) and \( \xi_{\varepsilon_\gamma} \) as presented in equations (14) and (15).

\[
\begin{align*}
\varepsilon_{\sigma^2} &= \varepsilon_{\sigma^2}^0 + \xi_{\varepsilon_\sigma^2}; \\
\varepsilon_\gamma &= \varepsilon_\gamma^0 + \xi_{\varepsilon_\gamma}.
\end{align*}
\]

Note that for simplicity the time and horizon subscripts were removed. (14) and (15) imply that while \( \varepsilon_{\sigma^2} \) and \( \varepsilon_\gamma \) are the residuals stated in (2) and (3), they are estimated as posterior values. \( \varepsilon_{\sigma^2}^0 \) and \( \varepsilon_\gamma^0 \) are set to zero—unless there is information that adjusts this assumption. Generally, \( \varepsilon_{\sigma^2}, \varepsilon_\gamma \in (\infty^-, \infty^+) \), which implies that \( \varepsilon_{\sigma^2}^0, \varepsilon_\gamma^0, \xi_{\varepsilon_\sigma^2}, \xi_{\varepsilon_\gamma} \) belongs to the real numbers \( \mathbb{R}^1 \).

In the fourth step, coherent prior probability mass distributions (PMD) are defined for the error terms \( \xi_{\varepsilon_{\varepsilon_{\sigma^2}}}, \xi_{\varepsilon_{\varepsilon_\gamma}}, \xi_{\varepsilon_{\varepsilon_\sigma^2}} \xi_{\varepsilon_{\varepsilon_\gamma}}, \) which were previously stated in equations (4), (12), (13), (14) and (15). Each of the error terms in the set \( \xi = \{\xi_{\varepsilon_{\varepsilon_{\sigma^2}}}, \xi_{\varepsilon_{\varepsilon_\gamma}}, \xi_{\varepsilon_{\varepsilon_\sigma^2}}, \xi_{\varepsilon_{\varepsilon_\gamma}}\} \) are discrete random variables with support \( V = \{v_1, v_2, \ldots, v_{K_j}\mid K_j \subseteq \mathbb{N}_{\geq 3}\}, \forall j = \{\varepsilon_{\sigma^2}, \gamma, \varepsilon_{\sigma^2}, \varepsilon_{\gamma}\} \). For symmetric purposes of the PMD of the error terms in the set \( \xi \), the integer \( K_j \) is constrained as a natural odd number. \( K_j = \{2k_j + 1: k_j \in \mathbb{Z}^+\}, \forall j \in J = \{\varepsilon_{\sigma^2}, \gamma, \varepsilon_{\sigma^2}, \varepsilon_{\gamma}\} \).

Prior and posterior probability weight sets are defined as

\[
W_0 = \{w_{j_1}^0, w_{j_2}^0, \ldots, w_{K_j}^0\mid w_{j_1}^0 \in [0, 1], K_j \subseteq \mathbb{N}_{\geq 3}\}, \forall j = \{\varepsilon_{\sigma^2}, \gamma, \varepsilon_{\sigma^2}, \varepsilon_{\gamma}\}.
\]

and

\[
W = \{w_{j_1}, w_{j_2}, \ldots, w_{K_j}\mid w_{j_1} \in [0, 1], K_j \subseteq \mathbb{N}_{\geq 3}\}, \forall j = \{\varepsilon_{\sigma^2}, \gamma, \varepsilon_{\sigma^2}, \varepsilon_{\gamma}\}.
\]

For coherence, the probability weights add to one, \( \sum_i w_{j_i} = 1 \), and \( \sum_i w_{j_i} = 1 \), \( \forall j \in J = \{\varepsilon_{\sigma^2}, \gamma, \varepsilon_{\sigma^2}, \varepsilon_{\gamma}\} \). Unless further evidence arises, the BEN procedure assumes that the prior mean values of the error terms are zero; i.e., \( \xi_{j_1}^0 = \sum_i w_{j_i}^0 v_{j_i} = 0, \forall j \in J = \{\varepsilon_{\sigma^2}, \gamma, \varepsilon_{\sigma^2}, \varepsilon_{\gamma}\} \). Further details of informative and non-informative symmetric prior distributions for the error terms \( \xi_{j_1}^0 \) are discussed in Go, Lofgren, Mendez-Ramos and Robinson (2016). The fifth step of the information-theoretical framework of the BEN procedure is described in the following Section 2.5.

### 2.5 Bayesian entropy estimation

The fifth step of the procedure consists of solving a minimization information-theoretical problem. The posterior mean values and error terms in (2) and (3) of the following parameters are simultaneously recovered: i) the risk factor coefficients, \( A = \{a_1, a_2, \ldots, a_R\} \); ii) the series of the ex-ante dispersion and skewness, \( \{\text{var} (\hat{Y}_{t+h})\}_{t=t_1}^{NT}, \{\gamma (\hat{Y}_{t+h})\}_{t=t_1}^{NT}, \forall h \in H \}; \) and iii) the noisy terms, \( \{\varepsilon_{\sigma^2,t+h}\}_{t=t_1}^{NT}, \{\varepsilon_{\gamma,t+h}\}_{t=t_1}^{NT}, \forall h \in H \). In this step, the BEN procedure minimizes the Kullback-Leibler (KL)—objective—
function subject to some consistency and behavioral restrictions. A summarized version of the information-theoretical problem is presented in (16)–(29).

\[
\begin{aligned}
\min_{\{w\}} \left( \alpha_R \sum_{r \in R} \sum_{k_{ar}=1}^{K_{ar}} w_{t+h,k_{ar}} \ln \left( \frac{w_{h,k_{ar}}}{w_{h,k_{ar}}} \right) + \alpha_F \sum_{l \in T} \sum_{f \in F} \sum_{k_{f}=1}^{K_f} w_{t+h,k_{f}} \ln \left( \frac{w_{t+h,k_{f}}}{w_{t+h,k_{f}}} \right) \right. \\
+ \left. \alpha_e \sum_{l \in T} \sum_{f \in F} \sum_{k_{ef}=1}^{K_{ef}} w_{t+h,k_{ef}} \ln \left( \frac{w_{t+h,k_{ef}}}{w_{t+h,k_{ef}}} \right) \right), \quad \forall h \in H.
\end{aligned}
\]

Subject to

Consistency property: posterior weights of the error-probability mass distributions

(17) \quad W_f = \left\{ w_{t+h,1}, \ldots, w_{t+h,K_f} \right\} \subseteq [0,1], K_f \subseteq \mathbb{N}_{\geq 3}, \quad \forall f \in F, t \in T, h \in H, j \in j.

(18) \quad \sum_{k_{j}=1}^{K_{j}} w_{t+h,k_{j}} = 1, \quad \forall f \in F, t \in T, h \in H, j \in j.

Consistency property: prior weights of the error-probability mass distributions

(19) \quad W^0_f = \left\{ w^0_{t+h,1}, \ldots, w^0_{t+h,K_f} \right\} \subseteq [0,1], K_f \subseteq \mathbb{N}_{\geq 3}, \quad \forall f \in F, t \in T, h \in H, j \in j.

(20) \quad \sum_{k_{ef}=1}^{K_{ef}} w^0_{t+h,k_{ef}} = 1, \quad \forall f \in F, t \in T, h \in H, j \in j.

Consistency property: posterior mean error terms

(21) \quad \xi_{f,t+h} = \sum_{k_{j}=1}^{K_{j}} w_{t+h,k_{j}} \nu_{t+h,k_{j}}, \quad \forall f \in F, t \in T, h \in H, j \in j.

Additive Bayesian update: posterior and prior mean values

(22) \quad \alpha_{r,h} = \tilde{\alpha}_{r,h}^{0} + \xi_{r,h}, \quad \forall r \in R, h \in H.

(23) \quad \var(Y_{t+h}|F_t) = \tilde{\var}(Y_{t+h}|F_t) + \xi_{2,r,t+h}^{2}, \quad \forall t \in T, h \in H.

(24) \quad \gamma(Y_{t+h}|F_t) = \tilde{\gamma}(Y_{t+h}|F_t) + \xi_{2,r,t+h}^{2}, \quad \forall t \in T, h \in H.

(25) \quad \epsilon_{f,t+h} = \epsilon_{f,t+h}^{0} + \xi_{f,t+h}, \quad \forall f \in F, t \in T, h \in H.

Consistency property: linear association of \( Y \) and \( Z \)

(26) \quad \var(Y_{t+h}|F_t) = \sum_{r \in R} \alpha_{r,h}^{2} \var(\tilde{Z}_{r,t+h}|F_t) + \epsilon_{2,r,t+h}^{2}, \quad \forall t \in T, h \in H.

(27) \quad \gamma(Y_{t+h}|F_t) = \sum_{r \in R} \alpha_{r,h}^{2} \gamma(\tilde{Z}_{r,t+h}|F_t) + \epsilon_{2,r,t+h}^{2}, \quad \forall t \in T, h \in H.

Behavioral assumption: TPN probability distribution of \( Y \)

(28) \quad \var(Y_{t+h}|F_t) = \frac{\pi-2}{\pi} \left( \sigma_{2,Y_{t+h}} - \sigma_{1,Y_{t+h}} \right)^2 + \sigma_{1,Y_{t+h}} \sigma_{2,Y_{t+h}}, \quad \forall t \in T, h \in H.

(29) \quad \gamma(Y_{t+h}|F_t) = \sqrt{\frac{\pi}{2}} \left( \sigma_{2,Y_{t+h}} - \sigma_{1,Y_{t+h}} \right)^{-1} = \frac{4-\pi}{\pi} \left( \sigma_{2,Y_{t+h}} - \sigma_{1,Y_{t+h}} \right)^2 + \sigma_{1,Y_{t+h}} \sigma_{2,Y_{t+h}}, \quad \forall t \in T, h \in H.
The set of parameters is defined as $J = \{a_r, \sigma^2, \gamma, \epsilon, \epsilon^2\}$. In previous sections of the paper the author defined the risk factor set as $R$, the time set as $T$, and the horizons set as $H$.

The objective function in (16) implies to minimize the noise (Kullback-Leibler pseudo-measure) between the posterior and the prior error distributions of the error terms on the set $\xi$. The consistency restrictions (17)–(20) assure coherence on the probability properties of the variety of involved random terms. The posterior error mean values of the variety of random terms are described in (21). The Bayesian update of the posterior mean value of the variety of random terms is done in additive form as pointed out in (22)–(25). Finally, the behavioral restrictions include the TPN probability distributional assumption of $Y$ and the linear association between the risk factors $Z$ and $Y$, as denoted in (26)–(29).

3 Scores

3.1 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 true forecaster’s predictive probability for the target random variable $Y$ is denoted with $p$ and the forecaster’s reported predictive probability with $u$. Both $p$ and $u$ can be probability density functions or probability mass distributions. The scoring rule is a function $S(p, y)$ 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 ex-ante sense—if the forecaster’s expected score eliciting his true predictive probability—density or mass—function $p$, $\mathbb{E}_p S(p, Y) = \int S(p, y)p(y)dy$, is smaller than his expected score eliciting any another probability function $u$, $\mathbb{E}_u S(u, Y) = \int S(u, y)p(y)dy$, e.g., equation (30) will hold for all probability functions $p$ and $u$. Finally, the scoring rule is strictly proper when the equality holds if and only if $p = u$ almost surely (Gneiting 2011).

\[ \mathbb{E}_p S(p, Y) \leq \mathbb{E}_p S(u, Y). \]  

If the forecaster prefers to use his true density function $p$, instead of cheating and reporting any other function $u$, the forecaster will be enforced by the proper scoring rule to set $u = p$, i.e., the forecaster obtains his best expected score for reporting his true predictive probability. Besides that, the negative orientation of (30) with respect to 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 individual 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, $\{y_{t+h} \}_{t=t_1}^{N_T}$, for a variety of horizons $h \in H$. Two different prequential forecasting systems (PFSs) for $Y$ collect the issued probability forecast sequences $\{p_{t+h} \}_{t=t_1}^{N_T}$ and $\{u_{t+h} \}_{t=t_1}^{N_T}$, $\forall h \in H$. They can be compared using the scoring rule $S(\cdot)$ by estimating their average scores, $\bar{S}_{N_T}^p(\cdot) = \frac{1}{N_T} \sum_{t=t_1}^{N_T} S(p_{t+h}, y_{t+h})$ and $\bar{S}_{N_T}^u(\cdot) = \frac{1}{N_T} \sum_{t=t_1}^{N_T} S(u_{t+h}, y_{t+h})$, $\forall h \in H$. Then, the subject performing the evaluation of the forecast sequences would prefer the PFS $p$ to the $u$ ($p \succeq u$) if and only if $\bar{S}_{N_T}^p \leq \bar{S}_{N_T}^u$, at any period $N_T \in \mathbb{N}_{\geq 1}$ and for all horizon $h \in H$, 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 true and honest probability predictions. Examples of risk-averse and risk-lover forecasters can be seen in Winkler and Murphy (1970), Winkler (1996), Bickel (2007), Johnstone (2007), Johnstone (2007), and Johnstone, Jose and Winkler (2011).

### 3.2 Weighted averages

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

$$
\bar{S}_{w,N_T}^u \left( \{u_{t+h} \}_{t=t_1}^{N_T}, \{y_{a,t+h} \}_{t=t_1}^{N_T} \right) = \sum_{t=t_1}^{N_T} w_{t+h} \cdot S(u_{t+h}, y_{a,t+h}), \quad \forall N_T \in \mathbb{N}_{\geq 1}, h \in H.
$$

The weights $w_{t+h}$ can be estimated taking into account some cyclical or seasonality depending on two main 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, in this paper, it is 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. Appendix 5 provides an example of the estimation of the weights used in (31).

3.3 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 $P(y) = P(Y \leq y)$.

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 (32). Figure 1 shows an example of elicited probability density and cumulative distribution functions for the released predicting variable $Y$ and the realized variable $Y_a$.

\[(32) \quad CRPS = CRPS(P, y_a) = \int_{-\infty}^{\infty} [P(y) - P_a(y)]^2 \, dy.\]

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 $dy$. 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 CRPS is oriented in negative form, that is, a low score is preferred to a high one. 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 true 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.,

$$CRPS \left( \left( P(y)_{t+h}^{N_T}, \{y_{a,t+h}\}_{t=t_1}^{N_T} \right) \right) = \sum_{t=t_1}^{N_T} w_{t+h} \ CRPS(P(y)_{t+h}, y_{a,t+h})$$

for any period $N_T \in \mathbb{N}_{\geq 1}$, and any horizon $h \in H$.

4 Application of the method: Prediction of world GDP growth

The use of national accounting systems started to become predominant across nations importantly in the twentieth century. During this period, the prediction of economic growth quickly gained importance and surged. Predicting economic growth accurately is an activity that many economists still struggle with. 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 constantly fail to foresee the precise economic growth of a country. Sometimes, these economic growth predictions fail dramatically. Recent evidence shows how a variety of forecasters have failed in predicting the economic growth of several countries (Figure 2). The January 2007–December 2008 historical forecast errors—for the next-year horizon—exemplify a period of high overconfident predictions, preceded and followed by under-confident forecast episodes.

In this section, the BEN procedure is applied to estimate world GDP growth density forecasts. Moreover, an evaluation of the generated forecasts is implemented. The CRPS proper scoring rule is used to evaluate the estimated density predictions. The assembled PFS follows the model stated in (16)–(29). To complement and benchmark the constructed BEN forecasts, naïve—or rule of thumb—density predictions are generated. The naïve forecasts are symmetric by construction. These benchmark predictions 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.

During the application, the ex-ante global growth density is assumed as partially unobserved. On one side, its most likely outcome—mode—is taken from predictions made in a structural model—with inputs from experts on country economic growth. On the other side, the conditional variance and the conditional third central moment of world GDP growth are not observed. The conditional variance and skewness measures complete the required statistics that fully depict the assumed TPN distribution stated in (28)–(29).

Monthly historical information about world GDP growth and risk factors is collected for the period of October 2005–August 2015 to generate the density forecasts. The ex-ante information of GDP growth and its risk factors uses predictive and implied data covering the scope of annual predictions over the period of 2005–2017. Predictive conditional variance and skewness series are estimated combining the historical and ex-ante information via the BEN procedure. The recovered conditional variance and skewness series are combined with historical information of world GDP growth and with predictive most likely outcomes—elicited in August 2015. This information allows the generation of predictive densities of global growth over the monthly time span of October 2005–August 2015 with an annual horizon $h = \{h_1 = N_T, h_2 = N_T +$
1, \( h_3 = N_T + 2 \), such that \( N_T = 2015 \). For instance, the density forecasts estimated for October 2006 predicted the world GDP growth in 2006, 2007, and 2008.

\[
N_T = 2015
\]

4.1 Risk factors and prior information

The subset of risk factors \( Z \) directly affecting global growth is shrunk to four variables. These random variables are: i) global absolute deviations from consumer price inflation targets; ii) global energy prices proxied by crude oil prices; iii) global financial conditions quantified by the S&P 500 index price; and iv) world term spreads measured by country-individual term spreads.

The density predictions for GDP growth are constructed reflecting upon the basic economic mechanisms between output growth and its underlying risks. These linear relationships of GDP growth, \( Y \), and its associated risk factors \( Z \) imply the constrained behavior of the variance and skewness. Although there is no general consensus on the sign of the linear associations, in this application, the assumed responses of global growth to the associated risk factor movements have been supported in previous studies.

<table>
<thead>
<tr>
<th>World Term Spread</th>
<th>World (CPI) Headline Inflation Absolute Deviations from Targets</th>
<th>World Energy Prices</th>
<th>S&amp;P 500 Index Prices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global GDP growth</td>
<td>(+)</td>
<td>(-)</td>
<td>(-)</td>
</tr>
</tbody>
</table>

The sign restrictions on the responses of World GDP growth to the movements of its global risk factors are presented in Table 1. The sign of the risk-coefficients in the linear equation (1) define the reactions of
the skewness to predicted global growth. The sign of the coefficients in equation (1) are preserved in the skewness equation (27). Thus—marginally—the global growth skewness is synchronized with the directional change of an associated risk factor. In addition, to identify the responses of GDP growth to changes in the selected risk factors, ordinary least squares (OLS) regressions using ex-post data under a variety of controls were estimated and presented in Appendix 6–Table 2.

International 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), etc. Absolute deviations of CPI headline 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 country sample; this negative relationship is supported by the results obtained in the OLS regressions presented in Table 2.

The four selected risk factors affecting world GDP growth are divided into two categories given their data source: 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 factors are collected by Consensus Economics. The market-based risk factor information is collected using Bloomberg. Implied volatilities of option prices are collected for the S&P 500 index, and for West Texas Intermediate (WTI) and Brent crude oils. Elicited most likely forecasts of world GDP growth are obtained from the World Bank.

Ex-ante global uncertainty statistics of term spread and the inflation measure are used as a proxy for the average aggregation of country-specific series weighted with corresponding real GDP information. The global term spread variance and skewness are constructed using the term spread data of 15 countries: Australia, Canada, France, Germany, India, Indonesia, Italy, Japan, the Republic of Korea, the Netherlands, Spain, Sweden, Switzerland, the United Kingdom, and the United States. The global inflation variance and skewness are aggregations from country data accounting for the 20 biggest economies in terms of GDP size: Australia, Brazil, Canada, China, France, Germany, India, Indonesia, Italy, Japan, the Republic of Korea, Mexico, the Netherlands, Russian Federation, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.

Implied volatilities of option contracts are the main source of information for estimating the ex-ante variance and skewness of the S&P 500 and crude oil prices. The selected implied volatilities come from option contracts with 3, 12 and 24-month maturities to represent the uncertainty for 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 4.
The most challenging section of the BEN method involves the identification of prior unobserved ex-ante variance and skewness of $Y$; real world GDP growth in this empirical application. Using the sign restrictions presented in Table 1, the author modeled the prior ex-ante variance and skewness of real world GDP growth as pointed out in Section 2.4—Prior information and estimation steps. Using data of world GDP growth baseline forecasts, Table 4 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. In this specific world GDP growth forecast context, the mechanics of “optimistic baseline forecasts assessed with bigger downside than upside risks” are a summary of the phrase “hope for the best, and prepare for the worst”. Finally, once priors for the alpha coefficients and the ex-ante variance and skewness of real GDP are selected, then the model presented in Section 2.5—Bayesian entropy estimation—can be applied to recover posterior estimates.

4.2 Estimated density forecasts

The semiannual elicited World Bank forecasts—most likely predicted outcomes—of real world GDP growth were combined with posterior monthly series of the ex-ante variance and skewness of real GDP growth to construct hypothetical monthly density forecasts. These generated monthly density forecasts are made from October 2005 to August 2015. The produced forecasts are called Bayesian entropy-generated forecasts—or BEN density forecasts.

Examples of hypothetically elicited BEN density forecasts of world GDP growth, its variance, and skewness decomposition are presented in Figure 6. This figure shows asymmetric confidence intervals of 40, 75, and 90 percent for the current, next-year, and two-year ahead horizons. The density forecast illustrations—hypothetically elicited at the end of March and August 2015—provide important information about the distribution of risks. For instance, the upside and downside risks can be easily compared across horizons. Table 3 shows the estimated posterior alpha coefficients used to derive the variance and skewness decompositions. Results of the posterior series of the ex-ante variance and skewness of GDP growth for three-year horizons are presented in Figure 7.

In addition, naïve density forecasts are generated to benchmark the BEN-generated density predictions. The naïve-density forecasts are constructed using historical semiannual World Bank point forecast series of real world GDP growth and attaching them an uncertainty of magnitude based on historical forecast error (HFE) standard deviations. The naïve density forecasts follow a normal distribution with a mean equal to the most likely GDP growth outcome and HFE standard deviations. 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 8.

4.3 Evaluation of density forecasts

The comparison of the Bayesian entropy-generated density forecast series with the naïve counterparts is done via the CRPS scoring rule. Unweighted and weighted averages of the individual CRPS scores are also
estimated. The formulation of the weighted average is presented in Appendix 5. The ex-post performance of the BEN and the naïve forecast series is presented in Table 5.

The results overwhelmingly show better scores—lower values—for the BEN density forecasts over the naïve-symmetric predictions across horizons, months, years, and in aggregated form. The BEN-generated density forecasts more accurately captured the uncertainty around world GDP growth. The CRPS scoring rule indicates that the BEN-generated fan charts are more reliable than the benchmark—naïve—density forecasts.

5. Concluding remarks

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

Assigning 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 scoring rule rankings will allow the comparison of future predictions: forecasters and methods. This compilation of scores will help incentivize forecasting improvements.

A Bayesian procedure was introduced in this paper to systematically generate asymmetric density forecasts. The methodology relies on prior information of parameters and unobserved statistics, expectations of risk factors statistics, and a two-piece normal distribution assumption. The TPN distribution provides flexibility for modeling the predictive densities. The method allows that selected risk factors, \( Z \), to depict the shape of the density forecasts, i.e., the description of the uncertainty around the most likely outcomes—predicted baseline—of the target variable, \( Y \).

The introduced Bayesian entropy methodology directly models the variance and skewness decompositions of the predictive density forecasts. The linear combination of the target variable, \( Y \), and its risk factors, \( Z \), offers a simple regression framework that exploits on the ex-ante variance and skewness information as a proxy for ex-ante uncertainty of the target variable. The Bayesian entropy procedure minimizes information noise among behavioral equations of the ex-ante global variance and skewness, the consistency 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 of 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 BEN-generated density forecasts. These world GDP growth BEN-generated forecasts are
monthly-elicited across the period of October 2005–August 2015, and are evaluated via the continuous ranked probability score. The CRPS is a proper scoring rule that measures the reliability of the resulting forecasts. The unweighted and weighted average CRPS scores show that the Bayesian entropy fan charts outperform the naïve density forecasts, which are generated by a rule of thumb using historical forecast error information and symmetric normal distribution assumptions. Noticeably, the estimated monthly BEN-generated density forecasts for world GDP growth show predominantly negative values of skewness—downside skewed density forecasts—across the time span of October 2005–August 2015. This spotlights 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.

This research could be expanded to include the testing of a variety of entropy measures to filter the information among the target variable and the selected risk factors. Future work could also investigate the quality and accuracy of the ex-ante variance and skewness information of the risk factors and its level of noise—usefulness. It is hoped that the incorporation of better and more accurate ex-ante prediction information will allow for the establishment of less noisy forecasting systems. Finally, the study of several scoring rules and the dynamic tracking of their score decompositions could enhance the discussion of forecaster skills.

References


Appendix

Appendix I. The two-piece normal distribution

The TPN distribution for the random variable $Y$ is denoted as $Y \sim TPN(\mu_Y, \sigma_{1,Y}, \sigma_{2,Y})$. 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, $\mu$, the variance, $\sigma$, and the skewness, $\gamma$. Or alternatively, it can be defined via the mode, $\mu$, the downside standard deviation, $\sigma_1$, and the upside standard deviation, $\sigma_2$. 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 (33), while its mean, variance, and skewness are stated in (34), (35), and (36), respectively. Under (33), 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 $Prob(X \leq \mu) < 0.5$.

$$f(y; \mu, \sigma_1, \sigma_2) = \begin{cases} f_1(\cdot) = \sqrt{2/\pi} (\sigma_1 + \sigma_2)^{-1} \exp \left(-\frac{(y-\mu)^2}{2\sigma_1^2}\right) \mathbb{I}\{y-\mu<0\}, \\ f_2(\cdot) = \sqrt{2/\pi} (\sigma_1 + \sigma_2)^{-1} \exp \left(-\frac{(y-\mu)^2}{2\sigma_2^2}\right) \mathbb{I}\{y-\mu\geq0\}. \end{cases}$$

(33)

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

(34)

$$\text{var}[Y] = (1 - 2/\pi)(\sigma_2 - \sigma_1)^2 + \sigma_1 \sigma_2.$$  

(35)

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

(36)

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 (37), where $\Phi(\cdot)$ is the standard normal cumulative distribution function, and $I(\cdot)$ is an indicator function.

$$Prob(a_1 \leq Y \leq a_2) = \begin{cases} 2\sigma_1 \over (\sigma_1 + \sigma_2) \left[\Phi\left(\frac{a_2-\mu}{\sigma_1}\right) - \Phi\left(\frac{a_1-\mu}{\sigma_1}\right)\right] \mathbb{I}\{a_1 \leq a_2 \leq \mu\}, \\ 2\sigma_2 \over (\sigma_1 + \sigma_2) \left[\Phi\left(\frac{a_2-\mu}{\sigma_2}\right) - \Phi\left(\frac{a_1-\mu}{\sigma_2}\right)\right] \mathbb{I}\{\mu \leq a_1 \leq a_2\}, \\ \left(f_1^\mu \right. \left.f_1(\cdot) \, \text{d}y + \int_{\mu}^{a_2} f_2(\cdot) \, \text{d}y \right) \mathbb{I}\{a_1 \leq \mu \leq a_2\}. \end{cases}$$

(37)
Under the TPN assumption, the CRPS rule takes on the closed form presented in (38). This closed representation is introduced by Gneiting and Thorarinsdottir (2010) where $\rho(\cdot)$ stands for the standard normal probability density function.

\begin{align}
CRPS(P, y_a) &= \int_{-\infty}^{\infty} [P(y) - P_a(y)]^2 dy. \\
&= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{y} f(x) dx - I_{(y-y_a)\geq 0} \right] dy. \\
&= \int_{-\infty}^{\infty} \left( \frac{2\sigma_1}{(\sigma_1+\sigma_2)} \Phi \left( \frac{y-a}{\sigma_1} \right) I(y \leq \mu) + \frac{2\sigma_2}{(\sigma_1+\sigma_2)} \left[ 1 - \Phi \left( \frac{y-a}{\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-a}{\sigma_1} \Phi \left( \frac{y-a}{\sigma_1} \right) + \rho \left( \frac{y-a}{\sigma_1} \right) \right] - (y-a) - \frac{2}{\sqrt{\pi}} \frac{\sqrt{\sigma_2} \left( \sigma_2^2 - \sigma_1^2 \right)}{(\sigma_1+\sigma_2)^2} I(y \leq \mu) \right] \\
&\quad + \left[ \frac{4\sigma_2^2}{\sigma_1+\sigma_2} \left[ \frac{y-a-a}{\sigma_2} \Phi \left( \frac{y-a}{\sigma_2} \right) + \rho \left( \frac{y-a}{\sigma_2} \right) \right] + \frac{4\sigma_1^2}{\sigma_1+\sigma_2} \left[ \frac{(\sigma_1-\sigma_2)^2 - 4\sigma_1^2}{\sigma_1+\sigma_2} \right] \frac{\left( y-a \right)}{(\sigma_1+\sigma_2)^2} \right] I(\mu \leq y). \\
\end{align}

**Appendix 2. Proof of Proposition 1**

The conditional variance of $\hat{Y}_{t+h}$—the ex-ante value of $Y$ at period $t$ and horizon $h$—using the linear assumption in (1) becomes:

\[
\text{var}(\hat{Y}_{t+h} | F_t) = \text{var}(a_1 \hat{Z}_{1,t+h} + \cdots + a_{N_R} \hat{Z}_{N_R,t+h} + \hat{\varepsilon}_{t+h} | F_t)
\]

\[
= \mathbb{E} \left[ (\hat{\varepsilon}_{t+h} + \sum_{i=1}^{N_R} a_i \hat{Z}_{i,t+h})^2 | F_t \right] - \mathbb{E}^2 \left[ (\hat{\varepsilon}_{t+h} + \sum_{i=1}^{N_R} a_i \hat{Z}_{i,t+h}) | F_t \right]
\]

\[
= \mathbb{E} \left[ \hat{\varepsilon}_{t+h}^2 + 2 \sum_{i=1}^{N_R} a_i \hat{Z}_{i,t+h} \hat{\varepsilon}_{t+h} + \sum_{i=1}^{N_R} a_i^2 \hat{Z}_{i,t+h}^2 + \sum_{j=1}^{N_R} \sum_{i=1}^{N_R} a_i a_j \hat{Z}_{i,t+h} \hat{Z}_{j,t+h} \right] | F_t \\
\quad - \left[ \mathbb{E}^2 [\hat{\varepsilon}_{t+h} | F_t] + 2 \sum_{i=1}^{N_R} a_i \mathbb{E} [\hat{Z}_{i,t+h} | F_t] \mathbb{E} [\hat{\varepsilon}_{t+h} | F_t] + \sum_{i=1}^{N_R} a_i^2 \mathbb{E}^2 [\hat{Z}_{i,t+h} | F_t] \\
\quad + \sum_{j=1}^{N_R} \sum_{i=1}^{N_R} a_i a_j \mathbb{E} [\hat{Z}_{i,t+h} | F_t] \mathbb{E} [\hat{Z}_{j,t+h} | F_t] \right]
\]

\[
= \sum_{i=1}^{N_R} a_i^2 \mathbb{E} [\hat{Z}_{i,t+h}^2 | F_t] - \sum_{i=1}^{N_R} a_i^2 \mathbb{E}^2 [\hat{Z}_{i,t+h} | F_t]
\]
\[
+ \sum_{j=1}^{N_R} \sum_{j \neq i}^{N_R} a_i a_j \mathbb{E} \left[ \hat{Z}_{i,t+h} \hat{Z}_{j,t+h} \mid \mathcal{F}_t \right] - \sum_{j=1}^{N_R} \sum_{j \neq i}^{N_R} a_i a_j \mathbb{E} \left[ \hat{Z}_{i,t+h} \mid \mathcal{F}_t \right] \mathbb{E} \left[ \hat{Z}_{j,t+h} \mid \mathcal{F}_t \right] + \\
+ \mathbb{E} \left[ \hat{\varepsilon}_{t+h}^2 \mid \mathcal{F}_t \right] - \mathbb{E}^2 \left[ \hat{\varepsilon}_{t+h} \mid \mathcal{F}_t \right]
\]

\[
+ 2 \sum_{j=1}^{N_R} a_i \mathbb{E} \left[ \hat{Z}_{i,t+h} \hat{\varepsilon}_{t+h} \mid \mathcal{F}_t \right] - 2 \sum_{j=1}^{N_R} a_i \mathbb{E} \left[ \hat{Z}_{i,t+h} \mid \mathcal{F}_t \right] \mathbb{E} \left[ \hat{\varepsilon}_{t+h} \mid \mathcal{F}_t \right]
\]

\[
= \sum_{i=1}^{N_R} a_i^2 \mathbb{E} \left[ \hat{Z}_{i,t+h}^2 \right] - \sum_{i=1}^{N_R} a_i^2 \mathbb{E}^2 \left[ \hat{Z}_{i,t+h} \right]
\]

\[
+ \sum_{j=1}^{N_R} \sum_{j \neq i}^{N_R} a_i a_j \mathbb{E} \left[ \hat{Z}_{i,t+h} \hat{Z}_{j,t+h} \mid \mathcal{F}_t \right] - \sum_{j=1}^{N_R} \sum_{j \neq i}^{N_R} a_i a_j \mathbb{E} \left[ \hat{Z}_{i,t+h} \mid \mathcal{F}_t \right] \mathbb{E} \left[ \hat{Z}_{j,t+h} \mid \mathcal{F}_t \right]
\]

\[
+ \mathbb{E} \left[ \hat{\varepsilon}_{t+h}^2 \right] - \mathbb{E}^2 \left[ \hat{\varepsilon}_{t+h} \right] + 2 \sum_{j=1}^{N_R} a_i \mathbb{E} \left[ \hat{Z}_{i,t+h} \hat{\varepsilon}_{t+h} \mid \mathcal{F}_t \right] - 2 \sum_{j=1}^{N_R} a_i \mathbb{E} \left[ \hat{Z}_{i,t+h} \mid \mathcal{F}_t \right] \mathbb{E} \left[ \hat{\varepsilon}_{t+h} \mid \mathcal{F}_t \right]
\]

\[
= \sum_{i=1}^{N_R} a_i^2 \cdot \text{var} \left[ \hat{Z}_{i,t+h} \right] + \sum_{j=1}^{N_R} \sum_{j \neq i}^{N_R} a_i a_j \text{cov} \left[ \hat{Z}_{i,t+h} \hat{Z}_{j,t+h} \right] + \text{var} \left[ \hat{\varepsilon}_{t+h} \right]
\]

\[
+ 2 \sum_{i=1}^{N_R} a_i \text{cov} \left[ \hat{Z}_{i,t+h} \hat{\varepsilon}_{t+h} \right].
\]

Denoting \( \varepsilon_{\sigma^2,t+h}^2 = \sum_{j=1}^{N_R} \sum_{j \neq i}^{N_R} a_i a_j \text{cov} \left[ \hat{Z}_{i,t+h} \hat{Z}_{j,t+h} \right] + \text{var} \left[ \hat{\varepsilon}_{t+h} \right] + 2 \sum_{i=1}^{N_R} a_i \text{cov} \left[ \hat{Z}_{i,t+h} \hat{\varepsilon}_{t+h} \right] \)
gives us equation (2). ■

Appendix 3. Proof of Proposition 2

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

\[
(39) \quad \gamma(aX) = \mathbb{E}[(aX - a\mathbb{E}X)^3] = \mathbb{E}[a^3X^3 - 3a^3X^2\mathbb{E}X + 3a^3X^2\mathbb{E}^2X - a^3\mathbb{E}^3X] = a^3[\mathbb{E}X^3 - 3\mathbb{E}X \cdot \text{var}(X) - \mathbb{E}^3X].
\]

Using the results of \( \text{var} \left( \hat{\gamma}_{t+h} \mid \mathcal{F}_t \right) \) from Appendix 2 and the linear assumption in (1), the conditional skewness of the ex-ante value of \( Y \) at period \( t \) and horizon \( h \) becomes:

\[
(40) \quad \gamma \left( \hat{\gamma}_{t+h} \mid \mathcal{F}_t \right) = \mathbb{E} \left[ \left( \hat{\gamma}_{t+h} - \mathbb{E}(\hat{\gamma}_{t+h}) \right)^3 \mid \mathcal{F}_t \right]
\]

\[
= \mathbb{E} \left[ \left( \hat{\varepsilon}_{t+h} + \sum_{i=1}^{N_R} a_i \hat{Z}_{i,t+h} \right)^3 \mid \mathcal{F}_t \right] \quad \text{section A}
\]

\[
- 3 \mathbb{E} \left[ \left( \hat{\varepsilon}_{t+h} + \sum_{i=1}^{N_R} a_i \hat{Z}_{i,t+h} \right) \mid \mathcal{F}_t \right] \cdot \text{var} \left( \left( \hat{\varepsilon}_{t+h} + \sum_{i=1}^{N_R} a_i \hat{Z}_{i,t+h} \right) \mid \mathcal{F}_t \right) \quad \text{section B}
\]

\[
- \mathbb{E}^3 \left[ \left( \hat{\varepsilon}_{t+h} + \sum_{i=1}^{N_R} a_i \hat{Z}_{i,t+h} \right) \mid \mathcal{F}_t \right]. \quad \text{section C}
\]
Expanding section $A$ of (40):

\[
\mathbb{E}
\left[
\left(\hat{\varepsilon}_{t+h} + \sum_{i=1}^{N_R} a_i \hat{Z}_{i,t+h}\right)^3 | \mathcal{F}_t\right] = \mathbb{E}[\hat{\varepsilon}_{t+h}^3 | \mathcal{F}_t] + \sum_{i=1}^{N_R} a_i^3 \mathbb{E}[\hat{Z}_{i,t+h}^3 | \mathcal{F}_t]
\]

\[
+ 3 \mathbb{E}[\hat{\varepsilon}_{t+h} | \mathcal{F}_t] \sum_{i=1}^{N_R} a_i^2 \mathbb{E}[\hat{Z}_{i,t+h}^2 | \mathcal{F}_t]
\]

\[
+ 3 \mathbb{E}[\hat{\varepsilon}_{t+h}^2 | \mathcal{F}_t] \sum_{i=1}^{N_R} a_i \mathbb{E}[\hat{Z}_{i,t+h} | \mathcal{F}_t]
\]

\[
= \mathbb{E}[\hat{\varepsilon}_{t+h}^3] + \sum_{i=1}^{N_R} a_i^3 \mathbb{E}[\hat{Z}_{i,t+h}^3]
\]

\[
+ 3 \sum_{i=1}^{N_R} a_i^2 \mathbb{E}[\hat{Z}_{i,t+h}^2] \cdot \mathbb{E}[\hat{\varepsilon}_{t+h}]
\]

\[
+ 3 \sum_{i=1}^{N_R} a_i \mathbb{E}[\hat{Z}_{i,t+h}] \cdot \mathbb{E}[\hat{\varepsilon}_{t+h}^2].
\]

Expanding section $B$ of (40):

\[
3 \cdot \mathbb{E}
\left[
\left(\hat{\varepsilon}_{t+h} + \sum_{i=1}^{N_R} a_i \hat{Z}_{i,t+h}\right) | \mathcal{F}_t\right] \cdot \text{var}
\left[
\left(\hat{\varepsilon}_{t+h} + \sum_{i=1}^{N_R} a_i \hat{Z}_{i,t+h}\right) | \mathcal{F}_t\right]
\]

\[
= 3 \cdot \mathbb{E}[\hat{\varepsilon}_{t+h} | \mathcal{F}_t] + \sum_{i=1}^{N_R} a_i \mathbb{E}[\hat{Z}_{i,t+h} | \mathcal{F}_t]
\]

\[
\cdot \left(\sum_{i=1}^{N_R} a_i^2 \cdot \text{var}[\hat{Z}_{i,t+h}] + \sum_{j=1}^{N_R} \sum_{i=1}^{N_R} a_i a_j \text{cov}[\hat{Z}_{i,t+h} \hat{Z}_{j,t+h}]
\]

\[
+ \text{var}[\hat{\varepsilon}_{t+h}] + 2 \sum_{i=1}^{N_R} a_i \text{cov}[\hat{Z}_{i,t+h} \hat{\varepsilon}_{t+h}]\right)
\]

\[
= 3 \cdot \mathbb{E}[\hat{\varepsilon}_{t+h} | \mathcal{F}_t] \sum_{i=1}^{N_R} a_i^2 \cdot \text{var}[\hat{Z}_{i,t+h}]
\]

\[
+ 3 \cdot \mathbb{E}[\hat{\varepsilon}_{t+h} | \mathcal{F}_t] \cdot \sum_{j=1}^{N_R} \sum_{i=1}^{N_R} a_i a_j \text{cov}[\hat{Z}_{i,t+h} \hat{Z}_{j,t+h}]
\]

\[
+ 3 \cdot \text{var}[\hat{\varepsilon}_{t+h}]
\]

\[
+ 3 \cdot 2 \cdot \mathbb{E}[\hat{\varepsilon}_{t+h} | \mathcal{F}_t] \sum_{i=1}^{N_R} a_i \text{cov}[\hat{Z}_{i,t+h} \hat{\varepsilon}_{t+h}]
\]

\[
+ 3 \cdot \sum_{i=1}^{N_R} a_i \mathbb{E}[\hat{Z}_{i,t+h} \mathcal{F}_t] \cdot \sum_{i=1}^{N_R} a_i^2 \cdot \text{var}[\hat{Z}_{i,t+h}]
\]

\[
+ 3 \cdot \sum_{i=1}^{N_R} a_i \mathbb{E}[\hat{Z}_{i,t+h} | \mathcal{F}_t] \cdot \sum_{j=1}^{N_R} a_i a_j \text{cov}[\hat{Z}_{i,t+h} \hat{Z}_{j,t+h}]
\]

\[
+ 3 \cdot \sum_{i=1}^{N_R} a_i \mathbb{E}[\hat{Z}_{i,t+h} | \mathcal{F}_t] \cdot \text{var}[\hat{\varepsilon}_{t+h}]
\]

\[
+ 3 \cdot 2 \cdot \sum_{i=1}^{N_R} a_i \mathbb{E}[\hat{Z}_{i,t+h} | \mathcal{F}_t] \cdot \sum_{i=1}^{N_R} a_i \text{cov}[\hat{Z}_{i,t+h} \hat{\varepsilon}_{t+h}]
\]

\[
= 3 \cdot \mathbb{E}[\hat{\varepsilon}_{t+h} | \mathcal{F}_t] \sum_{i=1}^{N_R} a_i^2 \cdot \text{var}[\hat{Z}_{i,t+h}]
\]

\[
+ 3 \cdot \mathbb{E}[\hat{\varepsilon}_{t+h} | \mathcal{F}_t] \cdot \sum_{j=1}^{N_R} \sum_{i=1}^{N_R} a_i a_j \text{cov}[\hat{Z}_{i,t+h} \hat{Z}_{j,t+h}]
\]
Expanding section $C$ of (40):

\[ + 3 \cdot \mathbb{E}[\hat{\epsilon}_{t+h}|\mathcal{F}_t] \cdot \text{var}[\hat{\epsilon}_{t+h}] + 6 \cdot \mathbb{E}[\hat{\epsilon}_{t+h}|\mathcal{F}_t] \sum_{i=1}^{N_R} a_i \text{cov}[\hat{Z}_{i,t+h} \hat{\epsilon}_{t+h}] \\
+ 3 \cdot \sum_{j=1}^{N_R} a_j \mathbb{E}[\hat{Z}_{j,t+h}|\mathcal{F}_t] \cdot a_j^2 \cdot \text{var}[\hat{Z}_{j,t+h}] + 3 \cdot \sum_{k=1}^{N_R} \sum_{j=1}^{N_R} a_i a_j a_k \cdot \text{cov}[\hat{Z}_{i,t+h} \hat{Z}_{j,t+h}] \cdot \mathbb{E}[\hat{Z}_{k,t+h}|\mathcal{F}_t] \cdot \\
+ 3 \cdot \sum_{i=1}^{N_R} a_i \mathbb{E}[\hat{Z}_{i,t+h}|\mathcal{F}_t] \cdot \text{var}[\hat{\epsilon}_{t+h}] + 6 \cdot \sum_{i=1}^{N_R} \sum_{j=1}^{N_R} a_i a_j \cdot \mathbb{E}[\hat{Z}_{i,t+h}|\mathcal{F}_t] \cdot \text{cov}[\hat{Z}_{j,t+h} \hat{\epsilon}_{t+h}] \\
= 3 \cdot \mathbb{E}[\hat{\epsilon}_{t+h}] \sum_{i=1}^{N_R} a_i^2 \cdot \text{var}[\hat{Z}_{i,t+h}] + 3 \cdot \mathbb{E}[\hat{\epsilon}_{t+h}] \sum_{j=1}^{N_R} \sum_{i=1}^{N_R} a_i a_j \text{cov}[\hat{Z}_{i,t+h} \hat{Z}_{j,t+h}] \\
+ 3 \cdot \mathbb{E}[\hat{\epsilon}_{t+h}] \cdot \text{var}[\hat{\epsilon}_{t+h}] + 6 \cdot \mathbb{E}[\hat{\epsilon}_{t+h}] \sum_{i=1}^{N_R} a_i \text{cov}[\hat{Z}_{i,t+h} \hat{\epsilon}_{t+h}] \\
+ 3 \cdot \sum_{j=1}^{N_R} \sum_{i=1}^{N_R} a_j \mathbb{E}[\hat{Z}_{j,t+h}|\mathcal{F}_t] \cdot a_j^2 \cdot \text{var}[\hat{Z}_{j,t+h}] + 3 \cdot \sum_{k=1}^{N_R} \sum_{j=1}^{N_R} a_i a_j a_k \cdot \text{cov}[\hat{Z}_{i,t+h} \hat{Z}_{j,t+h}] \cdot \mathbb{E}[\hat{Z}_{k,t+h}] \cdot \\
+ 3 \cdot \sum_{i=1}^{N_R} a_i \mathbb{E}[\hat{Z}_{i,t+h}|\mathcal{F}_t] \cdot \text{var}[\hat{\epsilon}_{t+h}] + 6 \cdot \sum_{i=1}^{N_R} \sum_{j=1}^{N_R} a_i a_j \cdot \mathbb{E}[\hat{Z}_{i,t+h}|\mathcal{F}_t] \cdot \text{cov}[\hat{Z}_{j,t+h} \hat{\epsilon}_{t+h}] \\
\]
\[
\mathbb{E}^3\left[ (\hat{\epsilon}_{t+h} + \sum_{i=1}^{NR} a_i \hat{Z}_{i,t+h}) | \mathcal{F}_t \right] = \left( \mathbb{E}[\hat{\epsilon}_{t+h} | \mathcal{F}_t] + \mathbb{E}\left[ \sum_{i=1}^{NR} a_i \hat{Z}_{i,t+h} | \mathcal{F}_t \right] \right)^3 \\
= \mathbb{E}^3[\hat{\epsilon}_{t+h} | \mathcal{F}_t] \\
+ \mathbb{E}^3\left[ \sum_{i=1}^{NR} a_i \hat{Z}_{i,t+h} | \mathcal{F}_t \right] \\
+ 3 \cdot \mathbb{E}[\hat{\epsilon}_{t+h} | \mathcal{F}_t] \cdot \mathbb{E}^2\left[ \sum_{i=1}^{NR} a_i \hat{Z}_{i,t+h} | \mathcal{F}_t \right] \\
+ 3 \cdot \mathbb{E}^2[\hat{\epsilon}_{t+h} | \mathcal{F}_t] \cdot \mathbb{E}\left[ \sum_{i=1}^{NR} a_i \hat{Z}_{i,t+h} | \mathcal{F}_t \right]. 
\]

Let us define the error term, \( \epsilon_{y,t+h}^3 \) in (44):

\[
\epsilon_{y,t+h}^3 = \gamma \left[ \hat{\epsilon}_{t+h} \right] - 3 \cdot \mathbb{E}[\hat{\epsilon}_{t+h}] \cdot \sum_{j=1}^{NR} a_i a_j \cdot \text{cov}\left[ \hat{Z}_{i,t+h}, \hat{Z}_{j,t+h} \right] \\
- 6 \cdot \mathbb{E}[\hat{\epsilon}_{t+h}] \cdot \sum_{i=1}^{NR} a_i \cdot \text{cov}\left[ \hat{Z}_{i,t+h}, \hat{\epsilon}_{t+h} \right] \\
- 3 \cdot \sum_{j=1}^{NR} \sum_{i=1}^{NR} a_i a_j \mathbb{E}[\hat{Z}_{j,t+h}] \cdot a_i^2 \cdot \text{var}\left[ \hat{Z}_{i,t+h} \right] \\
- 3 \cdot \sum_{k=1}^{NR} \sum_{j=1}^{NR} \sum_{i=1}^{NR} a_i a_j a_k \cdot \text{cov}\left[ \hat{Z}_{i,t+h}, \hat{Z}_{j,t+h}, \hat{Z}_{k,t+h} \right] \cdot \mathbb{E}[\hat{Z}_{k,t+h}] \\
- 6 \cdot \sum_{i=1}^{NR} \sum_{j=1}^{NR} a_i a_j \cdot \mathbb{E}[\hat{Z}_{i,t+h}] \cdot \text{cov}\left[ \hat{Z}_{j,t+h}, \hat{\epsilon}_{t+h} \right] 
\]

Inserting the results from (41), (42), and (43) into (40) and defining the error term \( \epsilon_{y,t+h}^3 \) as in (44) gives (3).

\section*{Appendix 4. Data transformations and log-normality assumptions}

The original series of absolute deviation of inflation targets and term spread are denoted via \( X_\pi \) and \( X_{TS} \), respectively. The original series are assumed to behave with means \( \mu_{X_\pi}, \mu_{X_{TS}} \), standard deviations \( \sigma_{X_\pi}, \sigma_{X_{TS}} \), and skewness (third central moment) \( \gamma_{X_\pi}, \gamma_{X_{TS}} \), respectively. For smoothing and variability comparison purposes, we construct the variables that enter equation (1) as \( Z_\pi = \frac{X_\pi}{\sigma_{X_\pi}} \), and \( Z_{TS} = \frac{X_{TS}}{\sigma_{X_{TS}}} \). The statistic \( \tilde{\sigma}_s \) 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.,

\[
r_{oil} = \ln \left( \frac{S_{oil,t}}{S_{oil,t_0}} \right) \sim N \left( \left( \mu_{oil} - \frac{\sigma_{oil}^2}{2} \right) \left( t_f - t_0 \right) \right) 
\]
and \( r_{SPX} = \ln \frac{S_{SPX,t_f}}{S_{SPX,t_0}} \sim N \left( \left( \mu_{SPX} - \frac{\sigma_{SPX}^2}{2} \right) (t_f - t_0), \sigma_{SPX}^2 (t_f - t_0) \right) \). Thus, 
\( \ln S_{k,t_f} \) is distributed as 
\[ N \left( \ln S_{k,t_0} + \left( \mu_{rf} - \frac{\sigma_{rf}^2}{2} \right) (t_f - t_0), \sigma_{rf}^2 (t_f - t_0) \right) \] for all \( k \in \{ oil, SPX \} \).

This price return distributional assumption implies that the stock prices can be approximated with log-normal distribution, 
\( S_{k,t_f} \sim \log N \left( \ln S_{k,t_0} + \left( \mu_{rf} - \frac{\sigma_{rf}^2}{2} \right) (t_f - t_0), \sigma_{rf}^2 (t_f - t_0) \right) \) for all \( k \in \{ oil, SPX \} \). The stock prices behave following the moments specified in (45)–(47).

The Black-Scholes-Merton model for option pricing assumes the log-normal distribution of the stock prices. This assumption allows to consistently proxy the ex-ante variance and skewness of the stock prices. The series of ex-ante variance and ex-ante skewness of the stock prices are recovered using equations (46) and (47). 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—\( \mu_{r,T+h} \), 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, the stock price series are transformed as in the case of the construction of the series \( Y, Z_{\pi}, Z_{TS} \). This transformation aims to smooth and benchmark the variation of the risk factors in equation (I) 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 equation (1) are generated using the transformation presented in (48).

\[ Z_k = \frac{S_k}{\sqrt{\text{var}[S_k]}}. \]
Where $\bar{v}ar[S_k]$ is the sample mean of the series $\{v\bar{a}r[S_{k,t}]\}^T_{t=t_0}$.

Appendix 5. Weights of the average scores in the empirical application

An example of the weighted penalization of negative oriented scores, like the CRPS, imaging the context of monthly-elicited forecasts and annual materialized events. 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 $h \in H$, such that $m = 1$ denotes January and $m = 12$ denotes December. To normalize the weights (penalization) system, it is assumed 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 (49). Following equation (49), Figure 3 presents the values of the weights for the sequence of months $\{m_1, \ldots, m_{12}\}$ and a three-year horizons set $H = \{0,1,2\}$.

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

\textbf{FIGURE 3. EXAMPLE OF WEIGHTS TO ASSESS MONTHLY RELEASED FORECASTS}

Notes: 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 month penalized if the forecaster does not elicit accurate annual predictions.
Appendix 6. Ex-ante global risks

Panel A. Ex-Ante Variance of Inflation Measure

Panel B. Ex-Ante Skewness of Inflation Measure

Panel C. Ex-Ante Variance of Term Spread

Panel D. Ex-Ante Skewness of Term Spread

Figure 4. Ex-ante variance and skewness of term spread and absolute deviations of a target inflation

Notes: Ex-ante variance and skewness of term spread and the absolute deviations from target inflation are estimated with country data from Consensus Forecasts. 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 biggest 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 are scale down using the maximum value of the corresponding original series.
Panel A. Ex-Ante Variance of Crude Oil Prices

Panel B. Ex-Ante Skewness of Crude Oil Prices

Panel C. Ex-Ante Variance of the S&P 500 Index Price

Panel D. Ex-Ante Skewness of the S&P 500 Index Price

**Figure 5. Ex-Ante Variance and Skewness of Crude Oil and S&P 500 Index Prices**

**Notes:** Ex-ante variance and skewness of crude oil and the S&P 500 index prices are estimated via implied volatilities of option contract prices. The raw data is taken from Bloomberg, which uses a variation of the option pricing model of Black-Scholes-Merton. The option contracts series are at the money. The variance series are scaled down using the sample standard deviation of the original series. The skewness series are scaled down using the maximum value of the corresponding original series.
### Table 2. World Real GDP Growth Regressions Using Ex-Post Information

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P 500 Index</td>
<td>0.42***</td>
<td>[3.63]</td>
<td>0.40***</td>
<td>[4.23]</td>
<td>0.34***</td>
<td>[3.95]</td>
<td>0.43***</td>
<td>[3.85]</td>
<td>0.39***</td>
<td>[4.24]</td>
<td>0.37***</td>
<td>[4.30]</td>
</tr>
<tr>
<td>Term Spread</td>
<td>0.066***</td>
<td>[3.30]</td>
<td>0.082***</td>
<td>[3.01]</td>
<td>0.12***</td>
<td>[5.39]</td>
<td>0.043**</td>
<td>[2.04]</td>
<td>0.069**</td>
<td>[2.42]</td>
<td>0.048**</td>
<td>[2.36]</td>
</tr>
<tr>
<td>Absolute Deviation from Inflation Target</td>
<td>-0.10**</td>
<td>[-2.28]</td>
<td>-0.0027</td>
<td>[-0.035]</td>
<td>-0.082</td>
<td>[-1.45]</td>
<td>-0.089***</td>
<td>[-1.94]</td>
<td>0.0031</td>
<td>[0.041]</td>
<td>-0.00087</td>
<td>[-0.017]</td>
</tr>
<tr>
<td>Crude Oil</td>
<td>-0.13***</td>
<td>[-9.91]</td>
<td>-0.0061</td>
<td>[-0.27]</td>
<td>-0.0060</td>
<td>[-0.33]</td>
<td>-0.12***</td>
<td>[-8.83]</td>
<td>-0.0026</td>
<td>[-0.12]</td>
<td>-0.0048</td>
<td>[-0.24]</td>
</tr>
<tr>
<td>S&amp;P 500 Index (t-1)</td>
<td>-0.47***</td>
<td>[-4.20]</td>
<td>-0.41***</td>
<td>[-4.03]</td>
<td>-0.32***</td>
<td>[-3.46]</td>
<td>-0.46***</td>
<td>[-4.36]</td>
<td>-0.42***</td>
<td>[-4.08]</td>
<td>-0.41***</td>
<td>[-4.36]</td>
</tr>
<tr>
<td>T81_89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T89_93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T93_01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T01_09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>World GDP Growth (t-1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.11***</td>
<td>[11.8]</td>
<td>0.56***</td>
<td>[3.77]</td>
<td>0.49***</td>
<td>[3.92]</td>
<td>1.11***</td>
<td>[11.8]</td>
<td>0.63***</td>
<td>[4.15]</td>
<td>0.74***</td>
<td>[6.18]</td>
</tr>
<tr>
<td>Observations</td>
<td>120</td>
<td></td>
<td>300</td>
<td></td>
<td>396</td>
<td></td>
<td>120</td>
<td></td>
<td>300</td>
<td></td>
<td>396</td>
<td></td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.654</td>
<td></td>
<td>0.300</td>
<td></td>
<td>0.286</td>
<td></td>
<td>0.685</td>
<td></td>
<td>0.308</td>
<td></td>
<td>0.414</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-Squared</td>
<td>0.64</td>
<td></td>
<td>0.28</td>
<td></td>
<td>0.27</td>
<td></td>
<td>0.67</td>
<td></td>
<td>0.29</td>
<td></td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Root Mean Squared Error</td>
<td>0.097</td>
<td></td>
<td>0.24</td>
<td></td>
<td>0.25</td>
<td></td>
<td>0.093</td>
<td></td>
<td>0.24</td>
<td></td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>Model Degrees of Freedom</td>
<td>6</td>
<td></td>
<td>8</td>
<td></td>
<td>9</td>
<td></td>
<td>7</td>
<td></td>
<td>9</td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Residual Degrees of Freedom</td>
<td>113</td>
<td></td>
<td>291</td>
<td></td>
<td>386</td>
<td></td>
<td>112</td>
<td></td>
<td>290</td>
<td></td>
<td>385</td>
<td></td>
</tr>
<tr>
<td>Model F-Test</td>
<td>43.1</td>
<td></td>
<td>23.5</td>
<td></td>
<td>29.5</td>
<td></td>
<td>46.6</td>
<td></td>
<td>19.4</td>
<td></td>
<td>46.1</td>
<td></td>
</tr>
<tr>
<td>P-Value of Model F-Test</td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Ordinary least squares estimates with robust statistics using ex-post monthly information. Robust t-statistics are denoted in brackets. *** p<0.01, ** p<0.05, * p<0.1. GDP growth and its covariates are in reduced form. The endogenous variable and its controls were divided by their own standard deviation. T81_89 is a binary variable taking the value of one over the period of February 1981–January 1989. T89_93 is a binary variable taking the value of one over the period of February 1990–January 1993. T93_01 is a binary variable taking the value of one over the period of February 1993–January 2001. T01_09 is a binary variable taking the value of one over the period of February 2001–January 2009.
Appendix 7. Bayesian entropy estimates

Panel A. March 2015 Model Predictions

Panel B. August 2015 Model Predictions

Notes: The monthly data is from October 2005 to August 2015. Balance of risks stands for the skewness decomposition. The section denoting “OTHERS” includes corresponding cross-moments of variance and skewness of the risk factors as well as other risk factors not considered in the exercise. Using equation (3), the contribution of the risk factor $Z_t$ to the global skewness is estimated as $\frac{\mu_t z_t}{\mu_t^2}$, Panel A and B use the predictions obtained via the Bayesian entropy problem—equations (16)-(29) —stated in the BEN procedure. Panel A reports the results from March 2015 while panel B reports the August 2015 results.
### TABLE 3 – POSTERIOR AND PRIOR MEAN RESPONSES OF WORLD GDP TO MOVEMENTS OF GLOBAL RISKS

<table>
<thead>
<tr>
<th>Alpha Coefficients</th>
<th>Horizon</th>
<th>World Term Spread</th>
<th>World (CPI) Inflation Absolute Deviations from Targets</th>
<th>World Energy (Crude Oil) Prices</th>
<th>S&amp;P 500 Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posterior</td>
<td>T</td>
<td>0.35</td>
<td>-0.29</td>
<td>-0.51</td>
<td>0.30</td>
</tr>
<tr>
<td>Posterior</td>
<td>T+1</td>
<td>0.43</td>
<td>-0.46</td>
<td>-0.50</td>
<td>0.47</td>
</tr>
<tr>
<td>Posterior</td>
<td>T+2</td>
<td>0.36</td>
<td>-0.36</td>
<td>-0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>Prior</td>
<td>T</td>
<td>0.51</td>
<td>-0.41</td>
<td>-0.69</td>
<td>0.24</td>
</tr>
<tr>
<td>Prior</td>
<td>T+1</td>
<td>0.51</td>
<td>-0.55</td>
<td>-0.63</td>
<td>0.62</td>
</tr>
<tr>
<td>Prior</td>
<td>T+2</td>
<td>0.54</td>
<td>-0.53</td>
<td>-0.66</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Notes: The posteriors are recovered via the Bayesian entropy problem stated in equations (16)-(29). These coefficients are used in equations (1), (2), and (3).

### FIGURE 7. POSTERIOR ESTIMATES OF EX-ANTE GLOBAL UNCERTAINTY-MEASURES OF WORLD GDP GROWTH

Note: The predictive variance and skewness series are recovered via the Bayesian entropy problem stated in equations (16)-(29).

### FIGURE 8. NAIVE-GENERATED DENSITY FORECASTS OF GDP GROWTH

Notes: The forecasts are symmetric and follow a normal distribution. Panel A density forecast varies with horizon-specific historical forecast error standard deviations. Panel B density forecast uses a single average historical forecast error standard deviation across horizons.
Table 4 – Historical Forecast Error of World GDP Growth Predictions

<table>
<thead>
<tr>
<th>Horizons:</th>
<th>Current Year (T)</th>
<th>T+1</th>
<th>T+2</th>
<th>Combined Horizons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.22</td>
<td>0.98</td>
<td>2.31</td>
<td>0.82</td>
</tr>
<tr>
<td>Median</td>
<td>0.16</td>
<td>0.73</td>
<td>1.49</td>
<td>0.51</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>1.20</td>
<td>1.85</td>
<td>1.78</td>
<td>1.66</td>
</tr>
<tr>
<td>Observations</td>
<td>13.00</td>
<td>11.00</td>
<td>4.00</td>
<td>28.00</td>
</tr>
</tbody>
</table>

Notes: World GDP growth forecast from World Bank semiannual predictions. Realizations of world GDP growth based on WDI. Historical forecast error defined as predicted minus realized value. A positive HFE value indicates overprediction.

Table 5 – Continuous Ranked Probability Score of Density Forecasts

<table>
<thead>
<tr>
<th>CRPS Average</th>
<th>CRPS Weighted Average</th>
<th>CRPS Average</th>
<th>CRPS Weighted Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPND Density Forecasts</td>
<td>TPND Density Forecasts</td>
<td>Normal Density Forecasts</td>
<td>Normal Density Forecasts</td>
</tr>
<tr>
<td>Bayesian Entropy Method</td>
<td>Bayesian Entropy Method</td>
<td>Varying HFE Standard Deviation Across Horizons</td>
<td>Varying HFE Standard Deviation Across Horizons</td>
</tr>
</tbody>
</table>

Panel A. Monthly Average Forecast Scores Across Period 2005 – 2015 and Annual Time-Horizons \(\{T, T+1, T+2\}\).

<table>
<thead>
<tr>
<th></th>
<th>CRPS Average</th>
<th>CRPS Weighted Average</th>
<th>CRPS Average</th>
<th>CRPS Weighted Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal Density Forecasts</td>
<td>Varying HFE Standard Deviation Across Horizons</td>
<td>Rule of Thumb</td>
<td>Rule of Thumb</td>
</tr>
<tr>
<td>January</td>
<td>1.6</td>
<td>0.53</td>
<td>1.79</td>
<td>0.62</td>
</tr>
<tr>
<td>February</td>
<td>1.55</td>
<td>0.58</td>
<td>1.79</td>
<td>0.67</td>
</tr>
<tr>
<td>March</td>
<td>1.51</td>
<td>0.61</td>
<td>1.79</td>
<td>0.72</td>
</tr>
<tr>
<td>April</td>
<td>1.55</td>
<td>0.66</td>
<td>1.79</td>
<td>0.77</td>
</tr>
<tr>
<td>May</td>
<td>1.51</td>
<td>0.69</td>
<td>1.79</td>
<td>0.82</td>
</tr>
<tr>
<td>June</td>
<td>1.45</td>
<td>0.7</td>
<td>1.76</td>
<td>0.85</td>
</tr>
<tr>
<td>July</td>
<td>1.44</td>
<td>0.73</td>
<td>1.76</td>
<td>0.89</td>
</tr>
<tr>
<td>August</td>
<td>1.47</td>
<td>0.79</td>
<td>1.76</td>
<td>0.94</td>
</tr>
<tr>
<td>September</td>
<td>1.46</td>
<td>0.82</td>
<td>1.79</td>
<td>1</td>
</tr>
<tr>
<td>October</td>
<td>1.46</td>
<td>0.86</td>
<td>1.76</td>
<td>1.03</td>
</tr>
<tr>
<td>November</td>
<td>1.47</td>
<td>0.9</td>
<td>1.76</td>
<td>1.08</td>
</tr>
<tr>
<td>December</td>
<td>1.45</td>
<td>0.93</td>
<td>1.76</td>
<td>1.12</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th>CRPS Average</th>
<th>CRPS Weighted Average</th>
<th>CRPS Average</th>
<th>CRPS Weighted Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal Density Forecasts</td>
<td>Varying HFE Standard Deviation Across Horizons</td>
<td>Rule of Thumb</td>
<td>Rule of Thumb</td>
</tr>
<tr>
<td>2005</td>
<td>1.04</td>
<td>0.62</td>
<td>1.5</td>
<td>0.9</td>
</tr>
<tr>
<td>2006</td>
<td>1.26</td>
<td>0.61</td>
<td>1.6</td>
<td>0.78</td>
</tr>
<tr>
<td>2007</td>
<td>2.66</td>
<td>0.92</td>
<td>2.65</td>
<td>0.99</td>
</tr>
<tr>
<td>2008</td>
<td>2.69</td>
<td>1.43</td>
<td>2.82</td>
<td>1.46</td>
</tr>
<tr>
<td>2009</td>
<td>1.16</td>
<td>1.03</td>
<td>1.97</td>
<td>1.15</td>
</tr>
<tr>
<td>2010</td>
<td>1.27</td>
<td>0.63</td>
<td>1.55</td>
<td>0.77</td>
</tr>
<tr>
<td>2011</td>
<td>1.23</td>
<td>0.56</td>
<td>1.47</td>
<td>0.68</td>
</tr>
<tr>
<td>2012</td>
<td>1.07</td>
<td>0.49</td>
<td>1.38</td>
<td>0.65</td>
</tr>
<tr>
<td>2013</td>
<td>0.89</td>
<td>0.4</td>
<td>1.32</td>
<td>0.62</td>
</tr>
<tr>
<td>2014</td>
<td>0.93</td>
<td>0.59</td>
<td>1.29</td>
<td>0.83</td>
</tr>
<tr>
<td>2015</td>
<td>0.72</td>
<td>0.57</td>
<td>0.97</td>
<td>0.77</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th>CRPS Average</th>
<th>CRPS Weighted Average</th>
<th>CRPS Average</th>
<th>CRPS Weighted Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal Density Forecasts</td>
<td>Varying HFE Standard Deviation Across Horizons</td>
<td>Rule of Thumb</td>
<td>Rule of Thumb</td>
</tr>
<tr>
<td>T</td>
<td>1.15</td>
<td>0.97</td>
<td>1.34</td>
<td>1.14</td>
</tr>
<tr>
<td>T+1</td>
<td>1.73</td>
<td>0.89</td>
<td>2.07</td>
<td>1.07</td>
</tr>
<tr>
<td>T+2</td>
<td>1.65</td>
<td>0.30</td>
<td>1.95</td>
<td>0.36</td>
</tr>
</tbody>
</table>


| Aggregated Score | 490.90 | 243.15 | 583.41 | 289.45 |

Notes: The weighted scores use weights derived following the rule presented in (49); the closer to the realization period the more penalized—weighted—the forecast. HFE stands for historical forecast error. The ex-post 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. Forecasts generated 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 horizons were estimated and not presented in this table. However, the scoring results remained the same: BEN density forecasts outperform naïve density predictions.