WP/20/78

IMF Working Paper

Predictive Density Aggregation: A Model for Global GDP Growth

by Francesca Caselli, Francesco Grigoli, Romain Lafarguette, Changchun Wang

INTERNATIONAL MONETARY FUND

WP/20/78

Predictive Density Aggregation: A Model for Global GDP Growth*

Francesca Caselli† Francesco Grigoli‡ Romain Lafarguette§ Changchun Wang¶

Abstract

In this paper we propose a novel approach to obtain the predictive density of global GDP growth. It hinges upon a bottom-up probabilistic model that estimates and combines single countries' predictive GDP growth densities, taking into account cross-country interdependencies. Specifically, we model non-parametrically the contemporaneous interdependencies across the United States, the euro area, and China via a conditional kernel density estimation of a joint distribution. Then, we characterize the potential amplification effects stemming from other large economies in each region—also with kernel density estimations—and the reaction of all other economies with parametric assumptions. Importantly, each economy's predictive density also depends on a set of observable country-specific factors. Finally, the use of sampling techniques allows us to aggregate individual countries' densities into a world aggregate while preserving the non-i.i.d. nature of the global GDP growth distribution. Out-of-sample metrics confirm the accuracy of our approach.

Keywords: Density aggregation, density evaluation, global GDP growth, predictive density.

JEL Codes: C12, E17, E37.

*The views expressed in this paper are those of the authors and do not necessarily represent those of the IMF or IMF policy. We thank, without implicating, Michal Andrle, John Bluedorn, Oya Celasun, Fei Han, Martina Hengge, Jaewoo Lee, Sheheryar Malik, Gian Maria Milesi-Ferretti, Malhar Nabar, Maury Obstfeld, Barbara Rossi, Damiano Sandri, and Mark Watson. We also thank participants to the seminars held at the IMF.

†International Monetary Fund, Research Department, fcaselli@imf.org.

‡International Monetary Fund, Research Department, fgrigoli@imf.org.

§International Monetary Fund, Monetary and Capital Markets Department, rlafarguette@imf.org.

¶International Monetary Fund, Information Technology Department, cwang2@imf.org.

Contents

List of Figures

List of Tables

1 Introduction

The use of predictive densities and fan charts has become common in many central banks and policy-making institutions. The Bank of England, for example, has been publishing a fan chart for inflation forecasts since 1996 (Holly et al., 2000), and has been emulated by other central banks ever since. The object of interest gradually broadened to other variables, notably GDP growth and unemployment. As of 2006, the International Monetary Fund (IMF) World Economic Outlook (WEO)'s global GDP growth projections have been accompanied by a fan chart summarizing the balance of risks (IMF, 2006).[1] Starting in 2017, the IMF Global Financial Stability Report (GFSR) has been publishing an assessment of unlikely, yet possible, global GDP growth collapses—thereby focusing on the left tail of the predictive density—employing a framework that links current economic and financial conditions to the probability distribution of future GDP growth (IMF, 2017).[2] These developments reflect the need to quantify—in probabilistic terms—uncertainties surrounding any point forecast along the forecasting horizon, as well as the need to pinpoint the probability of specific events of interest (e.g., the outcome associated with a five percent probability).

While the literature on predictive density generation is vast, aggregating predictive densities still poses a series of challenges.[3] Elliot and Timmermann (2016) and Aastveit et al. (2019) argue that—similar to the case of combinations of individual point forecasts outperforming a single point forecast—combinations of densities are also superior to individual models. The literature proposes different methods to combine predictive densities (see next section for a review), but these prove unsatisfactory when we need to aggregate densities of different objects (e.g., economic growth in different countries). In this case, in fact, the aggregation problem turns particularly severe because the data-generating process is different across countries and because country-specific densities are not identically and independently distributed (*i.i.d.*). In other words, country aggregates are subject to tail dependence: when one country suffers a negative shock, other countries may be more likely to suffer negative shocks as well. For these reasons, linear combinations of country-specific densities are not viable solutions as these marginal distributions do not account for their interdependences.

In this paper, we propose a novel approach to obtain the predictive density of global GDP growth, which relies on a bottom-up probabilistic model that estimates and combines single countries' predictive densities. To avoid the curse of dimensionality, we split the sample into three groups based on their share of world output and assume a reasonable structure of inter-dependencies: (i) *other economies* depend on large economies and their respective regional leaders, (ii) *regional leaders* depend on large economies, and (iii) regional leaders and other economies have no impact on *large economies*. Our approach consists of a four-step probabilistic model for global GDP growth that mixes parametric and non-parametric features. It first generates the joint predictive density via a multivariate conditional

[1]The IMF WEO's methodology was then revised in IMF (2009), incorporating information embedded in survey and market indicators to better anticipate changes in the distribution of risks (Kannan and Elekdag, 2009).

[2]See Prasad et al. (2019) for details about the methodology employed in the GFSR, which is largely inspired by Adrian et al. (2019b).

[3]See Tay and Wallis, 2000 for a review of the methods to generate density forecasts and Elliot and Timmermann (2016) and Aastveit et al. (2019) for a summary on density forecast combination.

kernel density estimation for a set of large economies (see Li and Racine 2007).[4,5] This means that a certain probability is assigned to all the possible combinations of future GDP growth rates in the three large economies (the US, the euro area, and China), conditional on the current values of country-specific factors and the other two large economies' contemporaneous GDP growth rates. Second, our approach estimates the regional leaders' kernel density for each conditioning triplet of large economies' GDP growth rates obtained in the previous step and samples from it to construct the predictive densities. Third, for each tuple of large economies' and regional leaders' GDP growth rates, we estimate the other economies' predictive density drawing from a normal distribution with historical mean and standard deviation.[6] And fourth, it aggregates each vector of GDP growth rates obtained in the previous steps with a simple PPP-weighted average. If judgement needs to be introduced—for example to re-center the mode around a given forecast—we propose a rejection sampling algorithm or by fitting the parametric distribution that minimizes the Kolmogorov-Smirnov distance between the empirical distribution and the parametric one.

In Figure 1, we report the evolution of the four-quarter ahead predictive density of global GDP growth obtained with the proposed methodology during 2005–2019. Despite a relatively short sample length, some interesting facts become apparent. First, the central tendency of the distribution shifts significantly over time, tracking relatively well the realization of actual global GDP growth. Second, the distribution is often close to be symmetric. This is not surprising given that, for the world and in 'normal' times, upsides risks in some economies are generally counterbalanced by downside risks in other economies, resulting in a distribution close to a normal one, consistent with the central limit theorem. Third, in correspondence to the global financial crisis, the collapse in GDP growth led to an upward skewed predictive density, signalling that in the aftermath of the crisis the probability of growth above the mode was higher than the probability of a further worsening, as in standard mean-reverting processes. Fourth, during the Euro area crisis, the density becomes significantly flatter and shows a predominant left tail, indicating large downside risks. Density evaluation exercises based on out-of-sample metrics confirm the accuracy of our approach.

The proposed approach has several advantages. First, at the cost of a few assumptions, it characterizes the probability of each state of the world allowing for inter-dependencies across large economies, potential amplification effects stemming from regional dynamics, and the endogenous reaction of other economies. Second, it circumvents the problem of density aggregation by employing sampling techniques, which allow to preserve the consistency needed to perform ex-post aggregation. Third, by estimating country-specific specifications instead of treating the world as an aggregate, it avoids the averaging of the conditioning factors across countries, which might dilute the signals they are expected to provide. And fourth, compared with quantile regression based approaches, kernel density estimators impose less constraints and does not require linearity assumptions, often resulting in accuracy gains.[7]

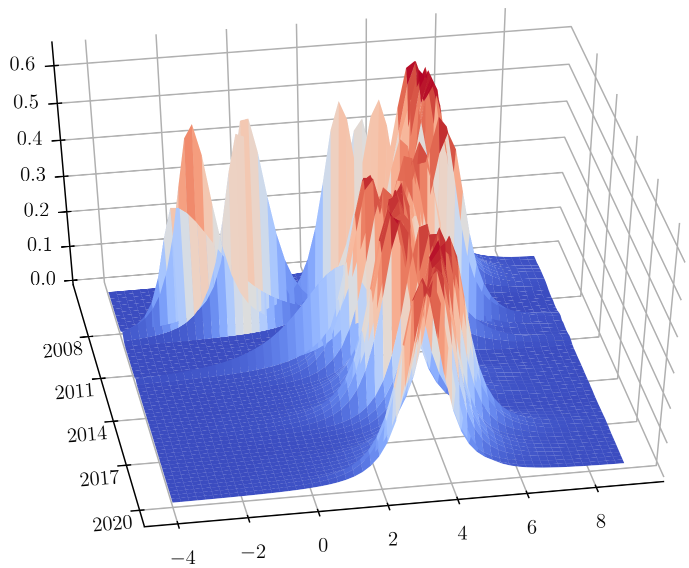
[4] Ideally, one would want to generate the joint density for all countries, however the curse of dimensionality would make the procedure computationally prohibitive.

[5] In a recent paper, Adrian et al. (2019a) use the same estimation approach to predict the density of US GDP growth based on past economic and financial conditions.

[6] This assumption conveniently speeds up the estimation procedure without much inaccuracy, as we only do it for the group of other economies.

[7] Quantile regressions approximate each quantile linearly, while kernel density estimators do not impose a functional form at the quantile level. This is particularly relevant when forecasting higher moments of the distribution, which relies on multiple quantile estimations that compound approximation errors.

Figure 1: Predictive Density of Global GDP Growth
(Four-quarter ahead forecast, percent)



Notes: The figure shows the four-quarter ahead distribution of global GDP growth. Blue colors indicate low probability, red colors indicate high probability.

The rest of the paper is organized as follows. Section 2 provides a brief review of the literature on density aggregation and summarizes its shortcomings when the object of interest is global GDP growth. Section 3 outlines the empirical strategy. Section 4 presents the results of the estimation and the evidence from density evaluation tests. Section 5 concludes.

2 Background

Similar to individual point forecasts, predictive density combinations tend to outperform individual predictive densities (Elliot and Timmermann, 2016). Rossi (2013), for instance, shows that the relative performance of different models varies over time and therefore suggests that forecast combination is a natural solution to eliminate the uncertainty about the best model.[8] Kascha and Ravazzolo (2010) show that combining densities is better than pre-selecting a particular model, as combinations insure against selecting improper models. However, compared to individual point forecasts, predictive density combination poses more challenges, as convex combinations could result in shapes completely different from the individual components, thereby delivering different forecasts. Most applications of predictive density combinations—generally across forecasters or forecasting models—aim at aggregating subjective probability distributions to create an opinion pool (Genest and Zidek, 1986). The commonly

[8] See Aastveit et al. (2019) for a comprehensive review of the evolution of the literature on forecasts density combination.

adopted combination strategy in the literature is the "linear opinion pool", which can be formalized as follows:

$$\hat{f}_{lin}(y) = \sum_{i=1}^{n} \omega_i f_i(y) \tag{1}$$

where $f_i(y)$ is the predictive density for the variable of interest y from forecaster/mode i, and ω is the weight such that $\sum_{i=1}^{n} \omega_i = 1$ (Stone, 1961). A notorious variation of the linear opinion pool is the "logarithmic opinion pool".[9]

These approaches, however, came under scrutiny in recent years. Hora (2004) shows that any non-trivial convex combination of two calibrated density forecasts is uncalibrated. Ranjan and Gneiting (2010) argue that linear combination formulas with strictly positive coefficients may not be coherent and demonstrated potential improvement under nonlinear aggregation. Busetti (2017) notes that the logarithmic opinion pool gives probability of zero to events that have zero probability under any of the individual distributions. He proposes a quantile aggregation strategy based on quantile functions of the individual forecast distributions and shows that the properties of this strategy are between those of the linear and logarithmic pool.

Rossi and Sekhposyan (2019) argue that equal weights to predicting densities often result in forecast improvements. However, the drawback of such weighting scheme is that the wealth of information about the individuals' predicting ability remains neglected. Hall and Mitchell (2007), Geweke and Amisano (2011), Waggoner and Zha (2012), Billio et al. (2013), Conflitti et al. (2015), Kapetanios et al. (2015), Pettenuzzo and Ravazzolo (2016), and Del Negro et al. (2016) propose various schemes to optimize weights for density combinations, including time-varying weights reflecting the past performance of different models. They show that they can lead to improvements in the forecasting performance.

When the object of interest is global GDP growth, the density combination problem becomes particularly severe. To make this point, let the world consist of two countries $i = 1, 2$, so that global GDP growth is defined as $y = \sum_{i=1}^{2} \omega_i y_i$, with ω being the country weights. If $x_i = \omega_i y_i$ with distribution $g_i(x_i)$ and if countries' GDP growth were independent from each other, their joint distribution would be:

$$f(y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} j(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{+\infty} j(x_1, y - x_1) dx_1 = g_1(x_1) g_2(x_2) \tag{2}$$

which is different from the weighted average of the distributions $\hat{f}_{lin}(y)$ as defined in equation (1).

An additional and critical layer of complexity is given by the tail dependence inherent to countries' GDP growth. That is, the aggregation scheme used to obtain $\hat{f}_{lin}(y)$ is based on the assumption that each element is unrelated to the others. However, $\hat{f}_i(y)$ describes the marginal distribution of economic growth of each country, and not its dependence on other countries.[10] This represents a major limita-

[9] Applications to predictive densities and generalizations of the linear opinion pool are in Mitchell and Hall (2005), Wallis (2005), Hall and Mitchell (2007), Jore et al. (2010), Kascha and Ravazzolo (2010), Garratt et al. (2011), Geweke and Amisano (2011), Billio et al. (2013), and Bassetti et al. (2020), among others.

[10] To see that, consider a world of two countries with equal weight ($\omega_1 = \omega_2 = 0.5$) and three possible GDP growth outcomes, say, 2, 4, and 6 percent. The distribution of global GDP growth resulting would feature 2 percent

tion if the aim is to construct world aggregates, as risks in some countries may lead to risks in other countries, either in the same direction or the opposite one.

How to then construct the predictive density for global GDP growth taking into account inter-dependencies across countries? The simplest workaround is to estimate directly the predictive density using aggregate data. This is the *status quo* in the IMF WEO and the IMF GFSR. The revised methodology employed in the IMF WEO since 2009 assumes that both global GDP growth and the key risk factors are drawn from a two-piece normal distribution function.[11] Survey or options price data for the term spread, S&P500 returns, oil prices, and inflation are used to construct one-year-ahead probability distributions for these variables. The variance and skew of these distributions are then mapped to the confidence intervals around the projections for global GDP growth using weights obtained with OLS regressions, thereby imposing linearity. Ohnsorge et al. (2016) proposed a similar approach based on vector autoregression models with time-varying weights that depend on the horizon of interest. In this (still linear) framework, the weights for the dispersion are obtained from the variance decomposition (i.e., share of global GDP growth forecast error variance explained by each risk factor) and the weights for the skewness are obtained from the impulse responses. These weights are then used to inform the parameters of the two-piece normal distribution function. Following the methodology proposed by Adrian et al. (2019b), since 2017 the IMF GFSR has been constructing the predictive density of global GDP growth by first estimating quantile regressions for the 10, 25, 50, 75 and 90 precentiles, conditioning on current economic and financial conditions (proxied by current global GDP growth and a financial condition index). Then, a skew-t distribution is fitted through the predicted conditional quantiles for the horizon of interest.

While these approaches based on aggregate data take into account inter-dependencies by construction, they suffer from other shortcomings. For example, they often rely on shorter series, possibly neglecting periods in which extreme events took place. In the case of the quantile regression-based approach, this may lead to characterizing the tails of the predictive density with very few data points. In the case of the techniques relying on survey data, the results may turn out to be unstable because of the relatively small sample of analysts submitting the forecasts. Also, averaging out country-level information may end up diluting the signals that it is expected to provide.

3 A Model of Predictive Density for Global GDP Growth

In this section, we describe our approach to aggregate the predictive densities of single countries into the predictive density of global GDP growth. We first layout the model setup and assumptions and then we present the steps of the model estimation.

3.1 Setup

Global GDP growth is defined as the PPP-weighted average of N individual economies' real GDP growth:

$(0.5 \times 2 + 0.5 \times 2)$ with probability 1/9, 3 percent with probability 2/9, 4 percent with probability 3/9, 5 percent with probability 2/9, and 6 percent with probability 1/9. This is different from the distribution implied by equation (1), that would feature 2 percent with probability 1/3, 4 percent with probability 1/3, and 6 percent with probability 1/3.

[11] The two-piece normal distribution has the benefit of a simple-to-compute density function and the ability to directly incorporate asymmetries.

$$y_t^W = \sum_{i=1}^{N} \omega_{t-1}^i \; y_t^i \quad \text{with} \quad y_t^i \;, \; \forall i \in \{1, \cdots, N\} \tag{3}$$

where $\{y_t^i\}_{i \,\in\, 1, \cdots, N}$ is a collection of non-$i.i.d$ GDP growth rates; and the time-varying weights, ω_{t-1}^i, are determined by the relative size of each economy's GDP in the world economy as of the previous period $\omega_{t-1} = Y_{t-1}^i / \sum_{j=1}^n Y_{t-1}^j$, with Y_t^j being the real GDP of economy j.

The non-$i.i.d$ nature of the sample rules out the possibility of estimating each economy's GDP separately. In this paper, we approach the problem by estimating the *joint* distribution of countries' GDP growth, $f(y_t^1, \cdots, y_t^n)$, to take into account cross-country inter-dependencies. The estimation of a joint density with a large number of countries, however, is computationally intensive because the dimensionality curse quickly kicks in as the number of dependent variables increases. To make the problem tractable, we implement an empirical strategy that reduces dimensionality by partitioning countries in three groups and running a sequential estimation.

After selecting a sample of countries based on data availability of quarterly series, we group them in three categories based on their share of world output as of 2018:[12]

1. Large economies: the United States (US), the euro area, and China. At the end of 2018, the combined GDP of these three countries represented 45 percent of the world GDP, based on the IMF WEO PPP weights.

2. Regional leaders: South Africa in Africa, India and Japan in Asia, Russia and the United Kingdom in Europe, Brazil and Mexico in the Western Hemisphere.[13] These represent the two largest economies in each region, with the exception of Africa for which we only have one country in the sample due to quarterly data limitations. At the end of 2018, the share of world GDP for these countries was 22 percent.

3. Other economies: Argentina, Australia, Belarus, Canada, Chile, Colombia, Croatia, Czech Republic, Denmark, Hong Kong SAR China, Hungary, Indonesia, Israel, Jordan, Kazakhstan, Korea, Malaysia, Moldova, Norway, Peru, Philippines, Poland, Romania, Serbia, Singapore, Sweden, Switzerland, Taiwan Province of China, Thailand, Turkey, Ukraine, Venezuela, and Vietnam. At the end of 2018, this group accounted for 19.5 percent of world GDP.

Then, we decompose the joint density of all economies into the conditional densities of the three groups via the density chain-rule:

$$f(\boldsymbol{y^L}, \boldsymbol{y^R}, \boldsymbol{y^O}) \;=\; \underbrace{f(\boldsymbol{y^O}|\boldsymbol{y^R}, \boldsymbol{Y^L})}_{\text{Third step}} \times \underbrace{f(\boldsymbol{y^R}|\boldsymbol{y^L})}_{\text{Second step}} \times \underbrace{f(\boldsymbol{y^L})}_{\text{First step}} \tag{4}$$

where $\boldsymbol{y^L}$ denotes the vector of the large economies, $\boldsymbol{y^R}$ is the vector of regional leaders, and $\boldsymbol{y^O}$ is the vector of the other economies.

Finally, we assume that:

[12]See Appendix A for the list of countries with the relative share of world GDP in PPP terms.
[13]We rely on the IMF WEO regional classification.

H1: Other economies' predictive densities depend on large economies' and their respective regional leaders', as well as country-specific factors. Hence, the conditional density for an economy of this group can be written as:

$$f(y^o \mid \mathbf{X^o},\ y^{US}, y^{EA}, y^{CHN}, \{y^i\}_{i\ \in\ 1,\cdots,R}, \{y^i\}_{i\ \in\ 1,\cdots,o-1,o+1,\cdots,O}) = \\ f(y^o \mid \mathbf{X^o},\ y^{US}, y^{EA}, y^{CHN}, \{y^i\}_{i\ \in\ 1,\cdots,R})$$

where $\mathbf{X^o}$ is a vector of country-specific conditioning factors.[14] This assumption allows to drastically reduce dimensionality, as there is no need to estimate the joint distribution of all other economies.

H2: Regional leaders' predictive densities is a function of large economies' GDP growth rates and country-specific factors $\mathbf{X^r}$. The conditional density for each regional leader y^r can then be expressed as:

$$(y^r \mid \mathbf{X^r},\ y^{US}, y^{EA}, y^{CHN}, \{y^i\}_{i\ \in\ 1,\cdots,r-1,r+1,\cdots,R}, \{y^i\}_{i\ \in\ 1,\cdots,O}) = \\ f(y^r \mid \mathbf{X^r}, y^{US}, y^{CHN}, y^{EA})$$

As in the case of other economies, this assumption simplifies the model by ruling out inter-dependencies among regional leaders. Instead, the model only considers the spillovers regional leaders might have from the US, the Euro Area, and China.

H3: Regional leaders and other economies have no impact on large economies, so that the predictive densities of large economies only depend contemporaneously on the other large economies' ones. In the case of the US, for example, the conditional density can be expressed as:

$$f(y^{US} \mid \mathbf{X^{US}}, y^{EA}, y^{CHN}, \{y^i\}_{i\ \in\ 1,\cdots,O\cup R}) = \\ f(y^{US} \mid \mathbf{X^{US}},\ y^{EA}, y^{CHN})$$

where $\mathbf{X^{US}}$ is a vector of US-specific conditioning factors.

Based on these three simplifying assumptions, we design the empirical strategy sequentially. First, the estimation of the joint density of the three large economies. Second, the estimation of the density of regional leaders, conditional to the joint density estimated in the first step. Third, the estimation of the other economies' conditional densities. And finally, the computation of the world density.

[14]For instance, Argentina's future GDP growth can be a function of Argentina's current GDP growth, financial conditions, industrial production, etc.

3.2 A Sequential Probabilistic Approach

We now go over the four steps of the estimation strategy. The first estimates the predictive densities of large economies, the second estimates the regional leaders' predictive density, the third estimates the other economies' predictive density, and the fourth aggregates predictive densities to obtain the predictive density for global GDP growth. To deal with the reduced number of observations, we use a rolling window that excludes the 12 quarters following the quarter being forecast.

3.2.1 Large Economies

The first and core step of the empirical strategy is to estimate the joint density of large economies' GDP growth at a given horizon h, conditional on a set of observed country-specific macroeconomic variables. This can be written as:

$$f(y_{t+h}^{US},\ y_{t+h}^{EA},\ y_{t+h}^{CHN} \mid \boldsymbol{X}_t^{US}, \boldsymbol{X}_t^{EA}, \boldsymbol{X}_t^{CHN}) \tag{5}$$

Several techniques are available to estimate joint densities, including the Gibbs sampler, copula functions, kernel density estimators, among others. In our approach, we rely on kernel density estimations—as presented in Li and Racine (2007) and applied by Adrian et al. (2019a)—since they require less observations than copulas while avoiding the problems of compatibility of the conditional distributions associated with the Gibbs sampler.[15] Moreover, as noted in Adrian et al. (ibid.), kernel density estimations allow to "remain agnostic about the nature of dynamic interactions between variables of interest, allowing the data to directly inform us instead."

The probability kernel density of a multivariate process $\boldsymbol{x}$ is given by:

$$\hat{f}_H(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{x} - \boldsymbol{x}_i) \tag{6}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^T$ and $\boldsymbol{x_j} = (x_{i1}, x_{i2}, \ldots, x_{id})^T$, $\forall\ i \in \{1, 2, \ldots, n\}$ are d-vectors; $\boldsymbol{H}$ is the bandwidth $d \times d$ matrix which is symmetric and positive definite; and K is the kernel function, a symmetric multivariate density, where $K_{\boldsymbol{H}}(\boldsymbol{x}) = |\boldsymbol{H}|^{-\frac{1}{2}} K(\boldsymbol{H}^{\frac{1}{2}}\boldsymbol{x})$.[16]

The conditional density $f(Y|X)$ can be estimated directly using the Bayes' theorem:

$$f(Y|X) = \frac{f(Y, X)}{f(X)}$$

where both the joint and the marginal multivariate distributions are estimated with the kernel presented in equation (6), with appropriate (potentially different) bandwidths. Estimating directly the joint density has the appealing feature of conditioning the joint distribution of the three countries on all the country-specific variables. In other words, US-specific variables not only affect the joint density through the US GDP growth, but have also a direct impact on the joint GDP growth of China and the euro area, hence capturing spillovers not necessarily mediated by GDP growth (or not mediated within

[15] See Chen and Ip (2015) for a discussion about Gibbs sampler behavior with incompatible conditional distributions.

[16] We use a Gaussian distribution for the kernel function as Gaussian kernels have convenient computational and asymptotic properties. However, we analyze the sensitivity of the results to the use of other distributions.

the forecast horizon).

One crucial aspect of the kernel density estimation is the choice of the bandwidth matrix $\boldsymbol{H}$, as it controls the amount and orientation of the smoothing induced. In our case, we follow Silverman's rule of thumb (Silverman, 1986):

$$\sqrt{\hat{\boldsymbol{H}}_{\boldsymbol{ij}}} = \left(\frac{4}{d+2}\right)^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \sigma_i$$

where d is the dimension or number of variables, n is number of points, and σ_i is the standard deviation of variable i, and $\hat{\boldsymbol{H}}_{\boldsymbol{ij}} = 0,\ \forall i \neq j$.[17] We estimate the optimal bandwidth matrix using historical data for $y^{US}_{t+h},\ y^{EA}_{t+h},\ y^{CHN}_{t+h},\ X^{US}_t,\ X^{EA}_t,\ X^{CHN}_t\ \forall t \in [1, \ldots, T]$.

To estimate the joint conditional future density in each conditioning period t_0, the state-space is discretized over an equally spaced three-dimensional support:

$$\{y^{US}_{t0+h,g},\ y^{EA}_{t0+h,g},\ y^{CHN}_{t0+h,g}\}_{g \in G} \in \mathcal{I}_{US} \times \mathcal{I}_{EA} \times \mathcal{I}_{CHN} \iff \mathcal{C}^G$$

where $\mathcal{I}_j$, $\forall\ j \in \{US, EA, CHN\}$ is the support for each economy's GDP growth, G is the number of points on the grid, and $\mathcal{C}^G$ the cubic grid of dimension G. In our baseline application, we set the number of points on the grid to be 50.[18] The minimum and the maximum of the supports have to be carefully chosen as, if they are too small, they may artificially truncate the tails of the joint distribution. Thus, we apply a rule for which each economy's specific support is bounded between the country's historical minimum GDP growth minus 10 percent and the historical maximum GDP growth plus 10 percent. Then, we implement an iterative procedure that (i) computes the probability of 100 equally distant points on the boundary surface of each country pair of the joint distribution; (ii) if the probability of the points at the extremes is larger than 0.1 percent, it extends the support by 0.5 percentage points in the direction of those points; and (iii) repeat (i) and (ii) until the points at the extremes have probability below 0.1 percent. This ensures that the tails of the joint distribution are defined up to a very low probability.

The estimation of the joint probability density function, conditional on a given set $\boldsymbol{X}^{US}_{t0},\ \boldsymbol{X}^{EA}_{t0},\ \boldsymbol{X}^{CHN}_{t0}$ can be then written as:

$$\hat{f}_{\hat{\boldsymbol{H}}}(y^{US}_{t0+h,g},\ y^{EA}_{t0+h,g},\ y^{CHN}_{t0+h,g} \mid \boldsymbol{X}^{US}_{t0},\ \boldsymbol{X}^{EA}_{t0},\ \boldsymbol{X}^{CHN}_{t0}) = \frac{\frac{1}{n}\sum_{i=1}^{n} K_{\hat{\boldsymbol{H}}_{\boldsymbol{y},\boldsymbol{x}}}(\boldsymbol{[y_l, x_{t0}]' - [y_i, x_i]'})}{\frac{1}{n}\sum_{i=1}^{n} K_{\hat{\boldsymbol{H}}_{\boldsymbol{x}}}(\boldsymbol{x_{t0}} - \boldsymbol{x}_i)} \quad \forall\ g \in \mathcal{C}^G \tag{7}$$

where the density is estimated over the 3-d grid. In other words, equation (7) estimates the probability distribution function for each point in the cube $\mathcal{C}^G$, conditional on country-specific variables.

[17] Alternative approaches include least squares cross-validation and biased cross-validation (Li et al., 2013), which return broadly comparable results. Adrian et al. (2019a) take a different approach, using the conditional standard deviation of each variable, up to a constant determined by out-of-sample performance.

[18] While it is desirable to have more points, the size of the cubic grid increases exponentially with G limiting its practical application: using a 50-point grid per variables gives a cube of 125,000 points, while using a 100-point grid would generate a cube of one million points, slowing considerably the procedure.

At the end of the first step of this procedure, the joint density of large economies $\hat{f}(y^{US}, y^{EA}, y^{CHN})$ is stored as a collection of jointly estimated 3-tuple of GDP growth rates $(y^{\cdot,US}_{t0+h}, y^{\cdot,EA}_{t0+h}, y^{\cdot,CHN}_{t0+h})$ for a given support:

$$\int_{\mathcal{G}^{US}} \int_{\mathcal{G}^{EA}} \int_{\mathcal{G}^{CHN}} \hat{f}(y^{US}, y^{EA}, y^{CHN}) \, \mathrm{d}y^{US} \, \mathrm{d}y^{EA} \, \mathrm{d}y^{CHN} = 1 \tag{8}$$

where $\mathcal{G}^{US}, \mathcal{G}^{EA}$, and $\mathcal{G}^{CHN}$ are the different discrete supports for real GDP growth in the US, the euro area, and China, respectively.[19]

3.2.2 Regional Leaders

We now turn to the estimation of the predictive density of regional leaders. As explained Section 3.1, future GDP growth of regional leaders is assumed to be conditional on the contemporaneous GDP growth of large economies as well as on past realizations of country-specific variables. Hence, as regional leaders' GDP growth rates are assumed to be independent of each other, we can estimate regional leaders' future GDP growth separately.

The estimation approach is similar to the one used for large economies since it consists of a kernel density estimation. However, the object of interest is now a univariate density rather than a multivariate one. Thus, the conditional kernel density estimation is given by:

$$\hat{f}_{\hat{\boldsymbol{H}}}(y^{r}_{t0+h,g}, \mid y^{US}_{t0+h,g}, y^{EA}_{t0+h,g}, \ y^{CHN}_{t0+h,g}, \ \boldsymbol{X}^{r}_{t0}) = \frac{\frac{1}{T}\sum_{i=1}^{T} K_{\hat{\boldsymbol{H}}_{\boldsymbol{y},\boldsymbol{x}}}(\boldsymbol{[y_r, x_{t0}]' - [y_i, x_i]'})}{\frac{1}{T}\sum_{i=1}^{T} K_{\hat{\boldsymbol{H}}_{\boldsymbol{x}}}(\boldsymbol{x_{t0}} - \boldsymbol{x}_i)} \tag{9}$$

where T is the length of the sample, which can vary across regional leaders.

In other words, for any three-tuple of large economies' GDP growth rates, $(\tilde{y}^{US}_{t0+h}, \ \tilde{y}^{EA}_{t0+h}, \ \tilde{y}^{CHN}_{t0+h})$, drawn from the joint support, $\mathcal{G}^{US} \times \mathcal{G}^{EA} \times \mathcal{G}^{CHN}$, and the current observed value for $\boldsymbol{X}^{r}_{t0}$, we obtain the predictive density of $\tilde{y}^{r}_{t0+h}$. Then, we simply draw one point from it. The process is repeated for all regional leaders, and, at the end of the second step of the procedure, we store a four-tuple for each region.

3.2.3 Other Economies

Spillovers from GDP growth of other economies to global GDP growth—either directly or via regional leaders and large economies—is likely to be marginal. This is because their country-specific share in global GDP is relatively small and because their impact on their neighbors is dominated by the spillovers stemming from large economies and regional leaders.[20] Hence, the main focus when estimating the predictive density for these economies resides on their conditional first moment—expected GDP growth conditional on the performance of larger economies, as well as country-specific factors—rather than on higher moments.

Even though the kernel density approach used in the two previous steps could in principle be ap-

[19]The joint density informs about the probability of any tuple drawn from the joint support: $\mathcal{G}^{US} \times \mathcal{G}^{EA} \times \mathcal{G}^{CHN} = \mathcal{C}^{G} : P\left[y^{US} = \tilde{y}^{US} \ , \ y^{EA} = \tilde{y}^{EA} \ , \ y^{CHN} = \tilde{y}^{CHN}\right] = \tilde{p}$.

[20]See, for instance, Rey (2015) for the importance of shocks in the US in shaping global financial conditions for the rest of the world and Huidrom et al. (2017) for evidence on spillovers from large emerging markets.

plied to other economies, this would make the procedure computationally challenging. Thus, with the objective of reducing the computation time, we adopt a parametric approach. Specifically, we impose that the predictive density of other economies follows a Gaussian distribution. We obtain the first moment of the Gaussian distribution via ordinary least squares: each economy's future GDP growth is conditioned on the future GDP growth of the regional leader of reference (for instance, Peru is associated with Brazil), the future GDP growth of the three large economies, as well as the economy's own set of regressors as follows:

$$y^o_{t+h} = \beta^o \boldsymbol{X}^o_t \ + \ \beta^r y^r_{t+h} \ + \ \beta^{US} y^{US}_{t+h} \ + \ \beta^{EA} y^{EA}_{t+h} \ + \ \beta^{CHN} y^{CHN}_{t+h} \ + \ \epsilon^o \quad \forall o \in O \tag{10}$$

where $\boldsymbol{X}^o_t$ is a vector of country-specific variables observable at time t, and O represents the set of other economies. Then, we compute the second moment of the Gaussian distribution using the economy's historical GDP growth variance estimator $\hat{\sigma}^o$.

Hence, for each four-tuple of future GDP growth rates for the three largest economies and one regional center, the density of the GDP growth a given economy o is given by:

$$y^o_{t+h} \ \sim \ \mathcal{N}(\hat{y}^o_{t+h}, \hat{\sigma}^o) \tag{11}$$

where $\hat{y}^o_{t+h}$ is the expected fitted value from estimating equation (10).

3.2.4 The World

Using the chain-rule formula in equation (4), the derivation of the predictive density via direct sampling is straightforward. Any given tuple $(\tilde{y}^{US}_{t+h}, \tilde{y}^{EA}_{t+h}, \tilde{y}^{CHN}_{t+h}, \{\tilde{y}^j_{t+h}\}_{j \in 1,\cdots,R}), \{\tilde{y}^j_{t+h}\}_{j \in 1,\cdots,O})$ generates a possible future state of the world economy by applying the aggregation formula in equation (3). Thus, drawing a large enough number of tuples from the joint support generates the distribution of interest $f(y^W_t)$.

4 Results

In this section, we first present the estimation results and then we report perform some density evaluation tests.

4.1 Estimation Results

As discussed in Section 3, our sample is composed by 43 countries and one country group of 19 countries (i.e., the euro area), representing 87 percent of the world output. Each country's predictive density depends on other countries' predictive densities as well as a set of observable country-specific factors, as described in Section 3.1. We group these factors in five categories: financial conditions (proxied by the financial conditions index),[21] leading indicators of economic activity (industrial production and

[21] See IMF (2018) for details about the construction of financial condition indexes for advanced economies and emerging markets; for the US, see Brave and Butters (2011).

purchasing managers' index), external variables (import demand, real effective exchange rate, and commodity terms of trade),[22] proxies of uncertainty (business confidence, economic policy uncertainty, and geopolitical risk),[23] and demographics factors (population growth).[24]

The estimation period varies by country depending on data availability. In the case of large economies, however, the joint distribution has to be estimated on the same sample period for the three countries. This is limited by the GDP growth data for China, which start in 1995q1.[25] Another relevant aspect of the specification is that, if covariates are available at monthly frequency, we include the lead of the average of monthly observations over a quarter.[26] The choice of variables for each country's specification is based on data availability as well as our understanding of what factors are likely to have an impact on GDP growth vulnerability for each specific country. To the cost of introducing some arbitrariness in the specification, we prefer this more subjective approach to one based on information criteria or regression analysis as some variables may become important in shaping GDP growth risks in the future but were not clearly doing so in the past. Beyond the other larger economies' contemporaneous GDP growth, the specification of the US includes past GDP growth, the financial condition index, industrial production, and business confidence; for the euro area, the set of regressors includes past GDP growth, industrial production, and the real effective exchange rate; and for China the specification features past GDP growth, the real effective exchange rate, import demand, commodity terms of trade, and population growth. However, as the object of interest is a joint predictive density, any of the variables above also plays a role for the other large economies.[27]

We first present the results of the estimation of the joint predictive density of the large economies. We show a contour plot between three pairs of countries. In Figure 2, we plot the joint predictive distribution between the US (on the x-axis) and the euro area (on the y-axis), where brighter colors indicate a larger probability. We display the four-quarter ahead joint distribution for selected horizons. The first horizon is 2006q1 (panel [a]), for which the probability is concentrated around relatively high growth rates for both economies: about 3 percent of the US and 2 percent for the euro area. As the global financial crisis approaches in 2008q4 (panel [b]), the mode of the joint distribution shifts towards lower growth rates for both economies. In the case of the United states, there is a relatively high probability of negative growth, but the outcome becomes more uncertain as shown by the relatively longer tails. The predictive joint distribution in 2009q1 (panel [c]) displays two distinguishable modes: a good one, characterized by higher growth rates in both economies, and a bad one, with lower growth rates. As of 2009q2 (panel [d]), the predicted distribution clearly assigns a higher probability to the bad outcome compared to the good one. In 2009q3 (panel [e]), bi-modality disappears and the probability of a

[22]See Gruss and Kebhaj (2019) for details about the construction of the commodity terms of trade.

[23]See Baker et al. (2016) for details on economic policy uncertainty and Caldara and Iacoviello (2018) on geopolitical uncertainty.

[24]See Appendix A for a description of the data sources.

[25]Given the prominent role of this group of countries in our approach, in a robustness test we extend the quarterly GDP growth series of China and the euro area with interpolated annual GDP growth rates back to 1981q1. The main results presented in the paper hold.

[26]The average is computed over the available monthly observations. For example, if no monthly observations are available, data up to 2018q4 is used to predict the GDP growth of 2019q4; however, if one monthly observation for 2018q2 is available, that observation is also used to predict 2019q4.

[27]Due to the limited sample, we use a parsimonious specification for the joint distribution. In our baseline we exclude the financial conditions index for the euro area and China because the indexes are quite correlated with the financial conditions index of the US. Including them do not significantly change the results.

deep recession in both economies increases considerably. Bi-modality was also observed in correspondence of the peak Euro-zone debt crisis's peak. In 2010 and 2011, the fear of excessive sovereign debt led to higher interest rates, which made it harder for countries with large deficits to finance their budgets. The predictive joint distribution for 2012q3 (panel [f]) indicates that, while the prevalent mode was still one featuring good growth outcomes, a certain probability could have been attached to lower growth rates.

To get an overview of the country-specific predictive densities from the joint distribution of large economies, Figure 3 reports the estimated 5–95 and 25–75 percentile range of the forecast, the mode of the forecast density, and the realized GDP growth since 2005 for the three large economies. In all panels, actual GDP growth falls within the 25–75 percentile range most of the time, providing some indication that the model is well specified to capture changes in GDP growth. This is true even in the case of China (panel [c]), which shows a negative trend in GDP growth since 2007. If anything, the 50 percent interval appears too wide. However, this assessment is complicated by the fact that we are estimating a joint distribution and not a country-specific predictive density. Consistent with the results in Figure 2, panels (a) and (b) for the US and euro area, respectively, clearly show a larger probability mass to the left of the mode in correspondence to the global financial crisis and the Euro-zone debt crisis.

We now present the results for the regional leaders.[28] Figure 4 shows the four-quarter ahead predictive densities of Brazil (panel [a]), India (panel [b]), Japan (panel [c]), Mexico (panel [d]), Russia (panel [e]), the United Kingdom (panel [f]), and South Africa (panel [g]). In each panel we plot the kernel fit of the sampled predictive densities at two different points in time. In particular, to appreciate the shifts in the predictive density, we take a quarter representative of normal times (2006q1), and a quarter when the country experienced some deceleration.[29] Unsurprisingly, predicting normal times is easier and the mode of the predictive density is not very far from the GDP growth realization. However, even at times of large swings in GDP growth and despite the relatively long forecast horizon (i.e., 4 quarters ahead), the results show consistent movements in the predictive density. In some cases and periods—for example the United Kingdom in Figure 4—some parts of the densities could appear jagged, as the bandwidth obtained with the Silverman's rule of thumb turns out too small for the grid. While the bandwidth can be changed, this may lead to some discretionality in the treatment of different countries or periods so we refrain from that.

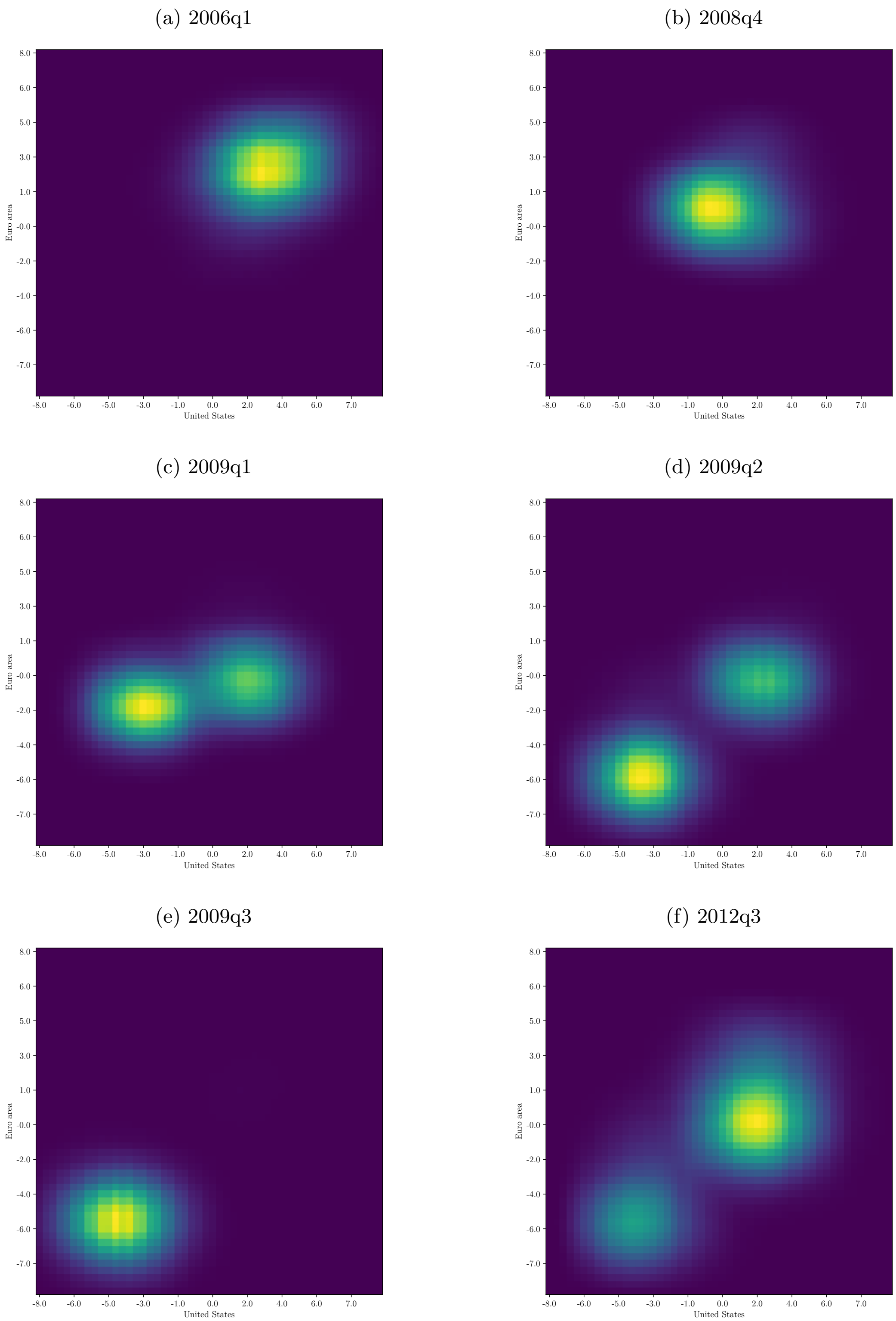
As an illustrative example for the other economies' densities, we report results for one advanced economy and one emerging market, namely Switzerland (panel [a]) and Turkey (panel [b]).[30] Figure 5 shows that also in the case of other economies the predictive densities move during bad times and that the mode of the density is relatively close to the actual realization. The results show that the assumption of Gaussianity for other economies does not significantly affect the accuracy of the forecasts.

[28]The sample period generally starts in the early 1990s, except in the case of India, for which data are available from 1997q1.

[29]For Brazil, we take 2016q; for India, 2019q3; for Japan 2019q2; for Russia 2015q4; and for Mexico and the United Kingdom we take a quarter in the middle of the global financial crisis, 2009q3.

[30]For Turkey we compare 2006q1 to 2018q2 and for Switzerland we report 2006q1 and 2009q3. The full set of other economies' densities are available upon request.

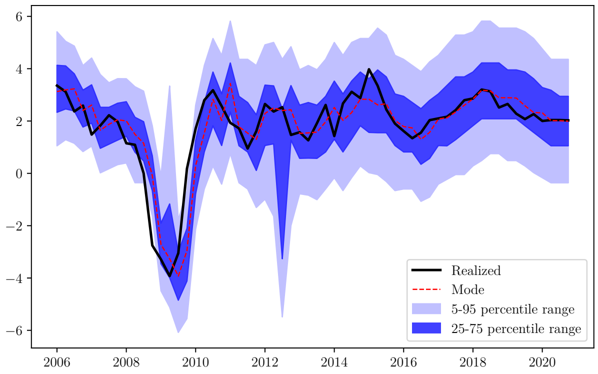
Figure 2: Predictive Density of the United States and the Euro Area from the Joint Distribution
(Four-quarter ahead forecast, percent)



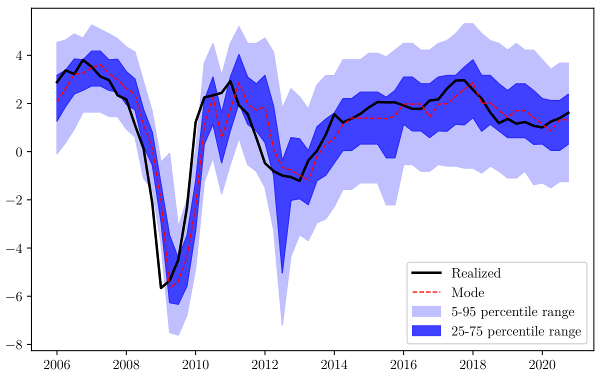
Notes: Brighter colors indicate higher probability.

Figure 3: Interval Forecast Evaluation for Large Economies
(Four-quarter ahead forecast, percent)

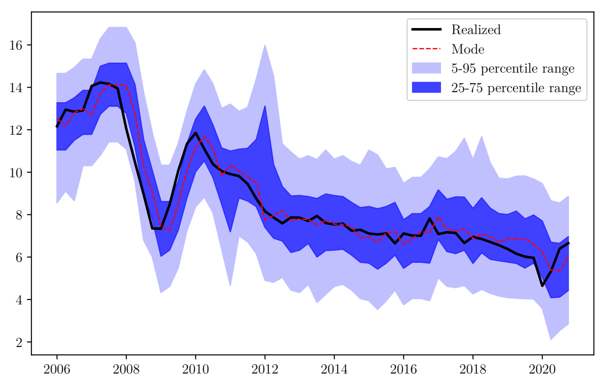
(a) United States



(b) Euro area

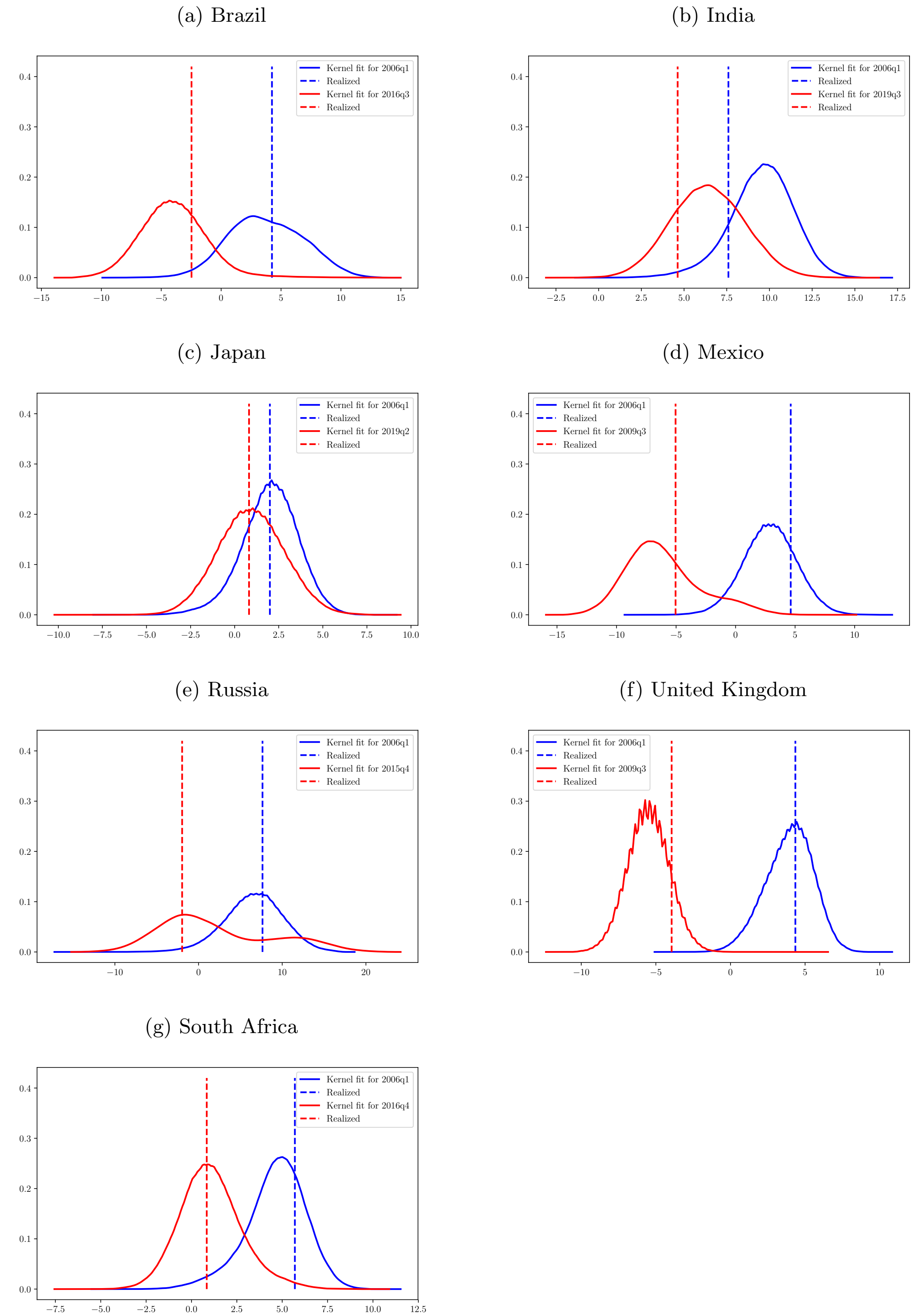


(c) China



Notes: The light blue and dark blue shades denote the estimated 5–95 and 25–75 percentile range of the predictive densities, the red line denotes the mode of the predictive density, and the black line denotes actual GDP growth.

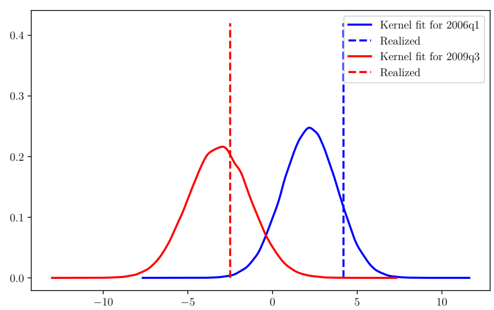
Figure 4: Predictive Density of Regional Leaders
(Four-quarter ahead forecast, percent)



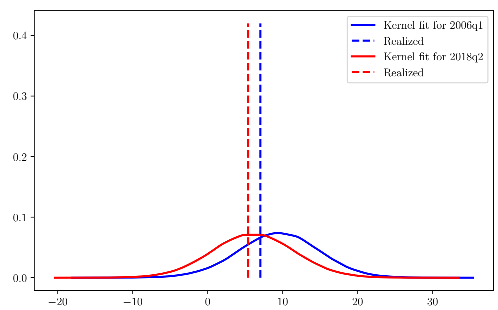
Notes: The blue lines denote the predictive densities in a quarter of 'normal times' (i.e. 2006q1). The red lines denote the predictive densities during GDP growth slowdowns or recessions (the quarter in this case varies by country). The blue and red vertical dashed lines denote the GDP growth realizations in the corresponding quarters.

Figure 5: Predictive Density of Other Countries
(Four-quarter ahead forecast, percent)

(a) Switzerland



(b) Turkey



Notes: The blue lines denote the predictive densities in a quarter of 'normal times' (i.e. 2006q1). The red lines denote the predictive densities during GDP growth slowdowns or recessions (the quarter in this case varies by country). The blue and red vertical dashed lines denote the GDP growth realizations in the corresponding quarters.

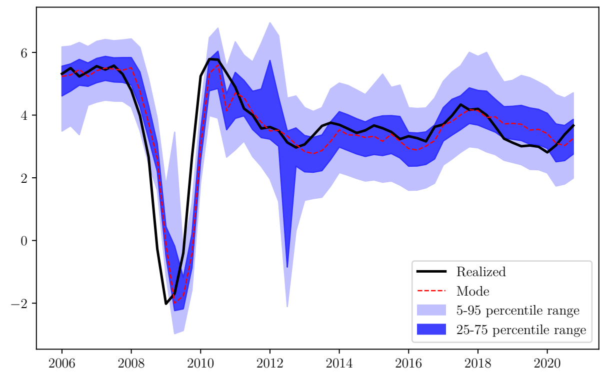
We now turn to the results for the world. Figure 7 shows the interval forecast of the predictive density for global GDP growth. This reveals that growth fell within the 25–75 percentile range of the predictive density most of the time. Since 2006 the 5–95 percentile range ranges between 1.8 percent and 6.4 percent, with the exception of the global financial crisis and the euro-zone crisis when the 5 percentile of the distribution reached -2.6 percent and -2 percent, respectively. As the global financial crisis is a unique event in the dataset, our data-driven approach struggles in predicting it and only does so with a lag.[31] However, as the technique leverages on past data, the interval forecast shows a significant downward skewness in correspondence of the euro-zone crisis.

The results of the global GDP growth density can be used to compute the probability that global GDP growth slows below a given threshold. To set a meaningful threshold, we take the global GDP growth rate corresponding to the 25th percentile of the historical distribution, which is about 3 percent. In Figure 6, we plot the probability associated with a GDP growth outcome below 3 percent over the entire sample period. The probability of a significant slowdown was unsurprisingly high during the global financial crisis. It is interesting that the left tail was already wide enough to predict a high probability of a slowdown at the end of 2007. The probability spiked again to about 60 percent in 2011 and 2012, and then again in 2015. Since 2017, it started to crawl upwards and reached about 43 percent in 2020.

The user of our approach might be interested in forcing the mode of the derived predictive density to an alternative central tendency. In the case of the IMF, for example, the mode should correspond to the baseline IMF WEO projection for global GDP growth. This is calculated as a PPP-weighted average of the individual country projections based on different models and, importantly, on informed judgement, which are meant to convey the desk estimate of the most likely GDP growth outcome. To do that, there are at least two options. A first option is to fit a parametric distribution with a constraint on the mode, which requires choosing across a wide family of parametric distributions. To limit

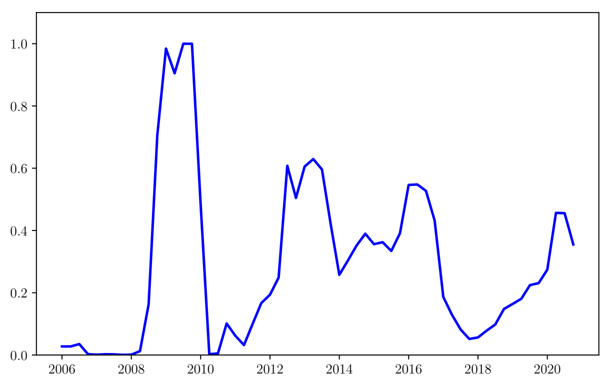
[31] This is not uncommon in the literature. Adrian et al. (2019b) also show that their approach based on quantile regressions fails to predict the collapse in output in the US during the global financial crisis.

Figure 6: Interval Forecast Evaluation for the World
(Four-quarter ahead forecast, percent)



Notes: The light blue and dark blue shades denote the estimated 5–95 and 25–75 percentile range of the predictive densities, the red line denotes the mode of the predictive density, and the black line denotes actual GDP growth.

Figure 7: Probability of a global GDP Growth Slowdown
(Four-quarter ahead forecast, percent)



Notes: The blue line denotes the probability associated with global GDP growth falling below the 25th percentile of the historical distribution (i.e., 3 percent).

arbitrariness, the minimum Kolmogorov-Smirnov distance between the empirical distribution and the parametric one can be used as a criterion.[32] A second option is to use a rejection sampling algorithm. This consists of a procedure in which a new sample is generated from the original one, so that the new sample exhibits particular features of interest. Specifically, the idea is to draw points from the original sample, and reject some points with a certain probability if they do not respect a given criteria. In our case, the probability of rejection is proportional to the distance of the drawn points to the mode of the

[32]One could also pre-select a parametric distribution. A particularly attractive option is the skewed-t distribution, which has a flexible probability density function with four parameters. Thus, all parameters but the location would be optimized to reflect the empirical distribution.

new sample. This approach ensures that the new sample distribution not only has the desired mode, but also that the moments of the distribution are as consistent as possible with the original data generating process, because points are randomly drawn from the original unconstrained sample.

In figure 8, we present an illustrative example of these two approaches based on the global predictive density for 2019q4, estimated with data up to 2018q4. Specifically, in panel (a) we plot the sampled density with a histogram, as well as the kernel fit and some alternative parametric fits: the skewed-t distribution, used in Adrian et al. (2019a) and in the IMF GFSR, the asymmetric-t distribution; and the best parametric fit among the beta distribution, the skew normal distribution, and the Weibull distribution. The sampled density of the global GDP growth distribution for 2019q4 varies between one and six percent, it is broadly symmetric and presents a mode of 3.2 percent. It is evident that parametric distributions do a better job than others at tracking the empirical distribution. In this case, for example, the Kolmogorov-Smirnov distance is the smallest for the beta distribution, which overlaps with the kernel fit. However, in some other cases, it is the skewed-t that produces the smallest Kolmogorov-Smirnov distance. Thus, given the flexibility of our approach, the preferred parametric form can be chosen on a case-by-case basis. In panel (b), we force the mean to the IMF WEO projection of 2.9 percent. Forcing the mode affects the tails and asymmetry of the distribution, highlighting the importance of the mode accuracy prediction. In this example, this leads to a modest change in the symmetry of the sample density obtained via rejection sampling, which becomes slightly skewed to the right. We re-fit all the parametric distributions and observe that, again, the beta distribution is the one better capturing the underlying sampled density.

4.2 Evaluating the Accuracy of Density Forecasts

The literature on density evaluation proposes two main approaches to assess the quality of a density forecast. The first approach includes methods that assess the *absolute* performance, that is how accurate is the specification of the conditional predictive density. These methods include the probability integral transform (PIT) (see Diebold et al., 1999 and Rossi and Sekhposyan, 2019, among others) and entropy measures. The second approach evaluates the *relative* performance, comparing competing predictive densities. Examples of this second approach include testing the equal predictive ability based on logarithmic scoring rules and expected shortfall and longrise.[33]. This section presents the tests and the results based on both approaches.

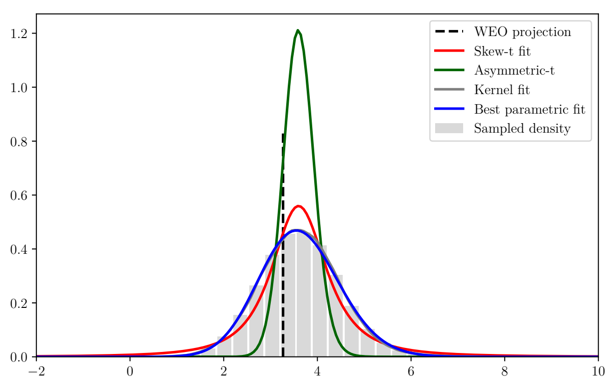
4.2.1 Probability Integral Transform

As explained in Rossi (2014), "a PIT is the cumulative probability evaluated at the actual, realized value of the target variable. It measures the likelihood of observing a value less than the actual realized value, where the probability is measured by the density forecast." That is, very small outcomes corresponding to 1 percent and 5 percent of the left tail of their probability forecasts would be assigned PIT values of 0.01 and 0.05, respectively. Similarly, an outcome in the middle of the probability distribution would be assigned a PIT value of 0.5, and large outcomes that are expected to occur with little

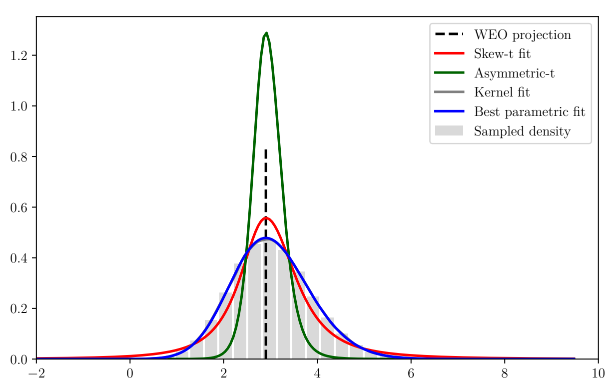
[33] The relative performance can also be tested using the Kullback-Leibler Information Criterion (KLIC). See Amisano and Giacomini (2007) for a comparison between scoring rules and the KLIC

Figure 8: Forcing the Mode of the Predictive Density of Global GDP Growth
(Four-quarter ahead forecast, percent)

(a) Unconstrained fitting



(b) Constrained fitting



Notes: The panels present the predictive density for 2019q4. The red line denotes the fitting of the skewed-t distribution; the green line denote the fitting of the asymmetric t distribution; the grey line denotes the fitting of the kernel; the blue line denotes the fitting of the best parametric fit among the beta distribution, the skewed-normal distribution, and the Weibull distribution; the grey bars denote the sampled density; and the vertical dashed line denotes the IMF, WEO projection for 2019q4.

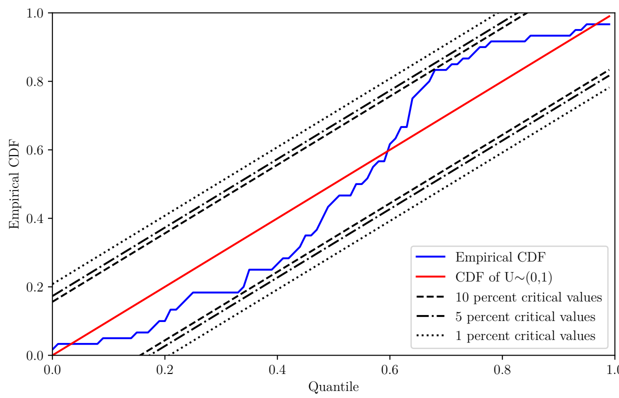
probability would be assigned PIT values close to one.

Diebold et al. (1999) show that, for correctly specified probability forecasts, the PIT values should be uniform, independent, and identically distributed. The uniformity property can then be used to argue that if, for example, we observe 30 out of 100 probability forecasts in the 5 percent probability region with PIT values between 0.01 and 0.05, the specification for the predictive density is misspecified. This is because such outcomes would be highly unlikely if the probability forecasts were correctly specified. In fact, we should expect to find only 5 observations with PIT values below 0.05. In other words, in this case the predictive density overestimates the likelihood of large negative outcomes. Thus, a test of the correct specification of density forecasts simply involves plotting the empirical distribution of the

PIT.[34] If the PITs were uniform, their cumulative distribution function would be close to a 45-degree line.

We report the results from the PIT test in Figure 9. Specifically, we show the empirical cumulative distribution function of the PIT of world GDP growth four quarters ahead, as well as the theoretical PIT (i.e., the 45-degree line) and the confidence bands around it.[35] The results indicate that the empirical distribution of the PIT is within the confidence band for all quantiles, even for the most conservative confidence band based on 10 percent critical values. This suggests that the approach followed in this paper has a satisfactory out-of-sample accuracy and that it generates robust predictive distributions that capture well upside and downside risks.

Figure 9: Probability Integral Transform Test
(Four-quarter ahead forecast)



Notes: The blue line denotes the empirical cumulative distribution function of the PIT, the red line denotes the theoretical cumulative distribution function of the PIT, and the black dash, dash-dotted, and dotted lines denote the confidence bands constructed off alternative critical values from Rossi and Sekhposyan (2019).

4.2.2 Entropy Measures

Another way of assessing the out-of-sample accuracy is to compute downside and upside entropy metrics. These compare the probability assigned to extreme growth outcomes by the conditional density to the probability assigned by the unconditional density. In other words, entropy corresponds to the extra probability mass that the predicted density assigns to the extreme right and left tail outcomes relative to the probability of these outcomes under the unconditional density. Following Adrian et al. (2019b), we define the upside entropy, $\mathcal{L}_t^U$, and the downside entropy, $\mathcal{L}_t^D$, of $\hat{g}_{y_{t+h}}(y)$ relative to $\hat{f}_{yt+h|x_t}(y|x_t)$ as:

[34] Testing for independence of the PIT can be done via a correlogram of the PIT, as explained in Diebold et al. (1999).

[35] Rossi and Sekhposyan (2019) derive the critical values for probability forecasts computed using a rolling scheme, assuming uniformity and independence of the PIT.

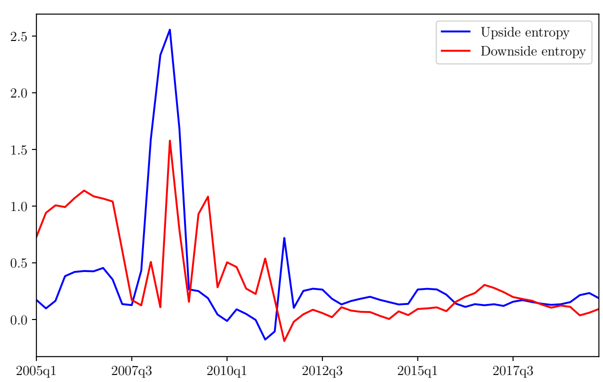
$$\mathcal{L}_t^D(\hat{f}_{y_{t+h}|x_t};\hat{g}_{y_{t+h}}) = -\int_{-\infty}^{\hat{F}^{-1}_{y_{t+h|x_t}}(0.5|x_t)} \left(log\hat{g}_{t+h}(y) - log\hat{f}_{y_{t+h|x_t}}(y|x_t) \right) \hat{f}_{yt+h|x_t}(y|x_t)dy \tag{12}$$

$$\mathcal{L}_t^U(\hat{f}_{y_{t+h}|x_t};\hat{g}_{y_{t+h}}) = -\int_{\hat{F}^{-1}_{y_{t+h|x_t}}(0.5|x_t)}^{\infty} \left(log\hat{g}_{t+h}(y) - log\hat{f}_{y_{t+h|x_t}}(y|x_t) \right) \hat{f}_{yt+h|x_t}(y|x_t)dy \tag{13}$$

where $\hat{F}_{y_{t+h}|x_t}(y|x_t)$ is the cumulative distribution associated with $\hat{f}_{y_{t+h}|x_t}(y|x_t)$ and $\hat{F}^{-1}_{y_{t+h|x_t}}(0.5|x_t)$ is the conditional median.

Figure 10 shows the evolution of the upside and downside entropy measures for the four-quarter ahead global GDP growth forecast. To perform this test, we fit the parametric distribution that minimizes the Kolmogorov-Smirnov distance between the empirical distribution and the parametric one. Compared to the unconditional density, the conditioning factors contribute to a higher probability of worse global GDP growth outcomes in the run-up to the global financial crisis. However, when the global financial crisis erupted, the conditional density shows fatter tails, indicating a significant increase in uncertainty. In fact, as the mode quickly shifted to worse outcomes, the conditional density shows an increased probability of a return to pre-crisis growth as well as an increase probability of a further worsening. During the euro area crisis, downside entropy remained high, suggesting that our conditioning factors contributed to predict higher downside risks compared to the unconditional density. Finally, while upside and downside entropy comove to some extent, downside entropy is more volatile than upside entropy.

Figure 10: Growth Entropy
(Four-quarter ahead forecast, value of upside and downside entropy)



Notes: The blue and red lines denote the upside and downside entropy, respectively, calculated relative to the unconditional distribution.

4.2.3 Expected Shortfall and Longrise

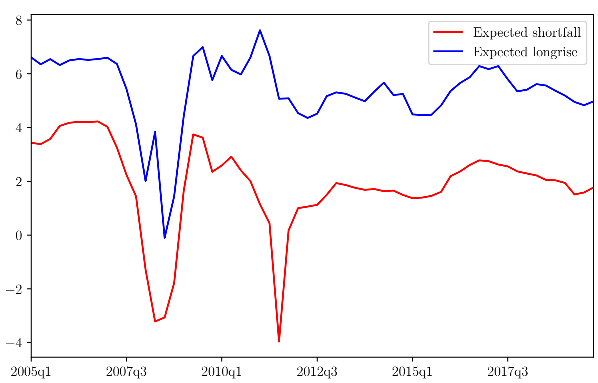
A complementary piece of information is given by the expected shortfall and longrise. These convey information about the tails of the predictive density. Differently from the entropy measures, the expected shorfall and expected longrise are absolute measures. Hence, if both the conditional and unconditional distribution have large left tails, the downside entropy will be low and the expected shortfall will be high. We compute the expected shortfall,SF, and longrise, LR, for a certain probability π as:

$$SF_{t+h} = \frac{1}{\pi}\int_0^{\pi} \hat{F}^{-1}_{y_{t+h|x_t}}(\tau|x_t)d\tau \tag{14}$$

$$LR_{t+h} = \frac{1}{\pi}\int_{1-\pi}^{1} \hat{F}^{-1}_{y_{t+h|x_t}}(\tau|x_t)d\tau \tag{15}$$

Figure 11 report the expected shortfall and longrise, for the 5th and 95th percentiles respectively, over time for the four-quarter ahead global GDP growth forecast. In line with the results obtained with the entropy metrics, when the downside entropy spikes up at the time of the global financial crisis the expected shortfall also moves quite significantly. This also happens in correspondence of the euro area crisis. Such similarity between the downside and upside entropy measures and the expected shortfall and longrise reveals non-Gaussian behavior of the global GDP growth predictive density.

Figure 11: Expected Shortfall and Longrise
(Four-quarter ahead forecast, value of expected shortfall and longrise)



Notes: The blue and red lines denote the expected longrise and expected shortfall computed at the 95th and 5th percentiles, respectively.

4.2.4 Log Scores

Log scores allow to test the predictive density against a benchmark performance (see Corradi and Swanson, 2006 and Amisano and Giacomini, 2007). In particular, they compare ex-post what was the probability associated ex-ante with the realization $\hat{f}_t(y_{t+h})$. In other words, they evaluate the accuracy of a forecast distribution, given that a certain outcome was observed.

Consider a stochastic process y_{t+h} and its density forecast in period t, $\hat{f}_t$. The likelihood score function S is defined as:

$$S(\hat{f}_t, y_{t+h}) \;= \log \hat{f}_t(y_{t+h}) \tag{16}$$

The null hypothesis of equal performance between two density forecasts $\hat{f}_t(y_{t+1})$ and $\hat{g}_t(y_{t+1})$ can be formulated using the score differential $d_{t+1}^{ls} = S(\hat{f}_t, y_{t+1}) - S(\hat{g}_t, y_{t+1})$. Let $\bar{d}_{m,n}$ be the average of the score differences between quarters m and $n = T - m$. To test the null hypothesis $H_0 : E[\bar{d}_{m,n}] = 0$ against the alternative $H_a : E[\bar{d}_{m,n}] \neq 0$ (or < 0 or > 0), the following Diebold and Mariano (1995) test statistic can be used:

$$t_{m,n} = \frac{\bar{d}_{m,n}}{\sqrt{\hat{\sigma}_{m,n}^2 / n}}$$

where $\hat{\sigma}_{m,n}^2$ is a heteroskedasticity and autocorrelation-consistent variance estimator of $\sigma_{m,n}^2 = \mathrm{Var}(\sqrt{n}\bar{d}_{m,n})$, which satisfies $\hat{\sigma}_{m,n}^2 - \sigma_{m,n}^2 \xrightarrow{p} 0$.

We test for equal performance between the conditional predictive density and the unconditional one. The results from the test indicate that the average logscore difference between the two predictive densities is 0.7 and that null hypothesis of equal performance is rejected at the 95 percent confidence interval. This confirms the better performance, in relative terms, of the conditional predictive density.

5 Conclusions

In the past few decades, the use of predictive densities and the associated fan charts became common in many central banks and policy-making institutions. Predicting the global GDP growth density poses a series of challenges as countries' GDP growth dependent on each other. That is, country aggregates are subject to tail dependence: when one country suffers a negative shock, other countries are more likely to suffer negative shocks as well. Besides, the low frequency of GDP data and the relatively short-time frame of GDP series for some large economies further complicate the estimation of predictive densities.

In this paper, we propose a novel approach to obtain the predictive density of global GDP growth, which relies on a bottom-up probabilistic model that estimates and combines single countries' predictive densities. At the cost of a few assumptions, our proposed method characterizes the probability of each state of the world allowing for inter-dependencies across large economies, potential amplification effects stemming from regional dynamics, and the endogenous reaction of other economies at different points of the joint distribution. Importantly, it circumvents the problem of density aggregation by employing sampling techniques, which allow to preserve the consistency needed to perform ex-post aggregation. Also, our approach incorporates country-specific information, while remaining agnostic about the data generating process at the country level and on the cross-country dependency.

Our results suggest that the obtained density provides generally accurate forecasts of the central tendency of global GDP growth and describes well the risks around it. Moreover, density evaluation exercises point to a satisfactory out-of-sample performance.

References

Aastveit, Knut Are, James Mitchell, Francesco Ravazzolo, and Herman K Van Dijk (2019). *The Evolution of Forecast Density Combinations in Economics*. Oxford Research Encyclopedia of Economics and Finance.

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone (2019a). "Multimodality in Macro-Financial Dynamics". Federal Reserve Bank of New York Staff Report 903.

— (2019b). "Vulnerable Growth". *American Economic Review* 109.4, pp. 1263–89.

Amisano, Gianni and Raffaella Giacomini (2007). "Comparing density forecasts via weighted likelihood ratio tests". *Journal of Business & Economic Statistics* 25.2, pp. 177–190.

Baker, Scott R, Nicholas Bloom, and Steven J Davis (2016). "Measuring Economic Policy Uncertainty". *The Quarterly Journal of Economics* 131.4, pp. 1593–1636.

Bassetti, Federico, Roberto Casarin, and Francesco Ravazzolo (2020). *Density Forecasting*. Springer, pp. 465–494.

Billio, Monica, Roberto Casarin, Francesco Ravazzolo, and Herman K Van Dijk (2013). "Time-Varying Combinations of Predictive Densities using Nonlinear Filtering". *Journal of Econometrics* 177.2, pp. 213–232.

Brave, Scott A and R Butters (2011). "Monitoring Financial Stability: A Financial Conditions Index Approach". *Economic Perspectives* 35.1, p. 22.

Busetti, Fabio (2017). "Quantile Aggregation of Density Forecasts". *Oxford Bulletin of Economics and Statistics* 79.4, pp. 495–512.

Caldara, Dario and Matteo Iacoviello (2018). "Measuring Geopolitical Risk". *Federal Reserve Bank International Finance Discussion Paper* 1222.

Chen, Shyh-Huei and Edward H Ip (2015). "Behaviour of the Gibbs Sampler when Conditional Distributions are Potentially Incompatible". *Journal of Statistical Computation and Simulation* 85.16, pp. 3266–3275.

Conflitti, Cristina, Christine De Mol, and Domenico Giannone (2015). "Optimal Combination of Survey Forecasts". *International Journal of Forecasting* 31.4, pp. 1096–1103.

Corradi, Valentina and Norman R Swanson (2006). "Predictive Density Evaluation". *Handbook of Economic Forecasting* 1, pp. 197–284.

Del Negro, Marco, Raiden B Hasegawa, and Frank Schorfheide (2016). "Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance". *Journal of Econometrics* 192.2, pp. 391–405.

Diebold, Francis X, Jinyong Hahn, and Anthony S Tay (1999). "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange". *Review of Economics and Statistics* 81.4, pp. 661–673.

Diebold, Francis X and Roberto S Mariano (1995). "Comparing Predictive Accuracy". *Journal of Business and Economic Statistics* 13.3, pp. 253–263.

Elliot, Graham and Allan Timmermann (2016). *Economic Forecasting*. Princeton University Press.

Garratt, Anthony, James Mitchell, Shaun P Vahey, and Elizabeth C Wakerly (2011). "Real-Time Inflation Forecast Densities from Ensemble Phillips Curves". *The North American Journal of Economics and Finance* 22.1, pp. 77–87.

Genest, Christian and James V Zidek (1986). "Combining Probability Distributions: A Critique and an Annotated Bibliography". *Statistical Science* 1.1, pp. 114–135.

Geweke, John and Gianni Amisano (2011). "Optimal Prediction Pools". *Journal of Econometrics* 164.1, pp. 130–141.

Gruss, Bertrand and Suhaib Kebhaj (2019). "Commodity Terms of Trade: A New Database". International Monetary Fund Working Paper 19/21.

Hall, Stephen G and James Mitchell (2007). "Combining Density Forecasts". *International Journal of Forecasting* 23.1, pp. 1–13.

Holly, Sean, Martin Weale, and Brian Corby (2000). *Econometric modelling: Techniques and applications*. Vol. 41. Cambridge University Press.

Hora, Stephen C (2004). "Probability Judgments for Continuous Quantities: Linear Combinations and Calibration". *Management Science* 50.5, pp. 597–604.

Huidrom, Raju, M Ayhan Kose, and Franziska L Ohnsorge (2017). "How Important Are Spillovers from Major Emerging Markets?" World Bank Policy Research Working Paper 8093.

IMF (2006). "Economic Prospects and Policy Issues". World Economic Outlook, Chapter 1, April, International Monetary Fund.

— (2009). "Global Prospects and Policy". World Economic Outlook, Chapter 1, April, International Monetary Fund.

— (2017). "Are Countries Losing Control of Domestic Financial Conditions?" World Economic Outlook, Chapter 3, October, International Monetary Fund.

— (2018). "A Decade after the Global Financial Crisis: Are We Safer?" World Economic Outlook, Annex 1.1, October, International Monetary Fund.

Jore, Anne Sofie, James Mitchell, and Shaun P Vahey (2010). "Combining Forecast Densities from VARs with Uncertain Instabilities". *Journal of Applied Econometrics* 25.4, pp. 621–634.

Kannan, Prakash and Selim Elekdag (2009). "Incorporating Market Information into the Construction of the Fan Chart". International Monetary Fund Working Paper 09/178.

Kapetanios, G, James Mitchell, Simon Price, and Nicholas Fawcett (2015). "Generalised Density Forecast Combinations". *Journal of Econometrics* 188.1, pp. 150–165.

Kascha, Christian and Francesco Ravazzolo (2010). "Combining Inflation Density Forecasts". *Journal of Forecasting* 29.1-2, pp. 231–250.

Li, Qi, Juan Lin, and Jeffrey S Racine (2013). "Optimal Bandwidth Selection for Nonparametric Conditional Distribution and Quantile Functions". *Journal of Business & Economic Statistics* 31.1, pp. 57–65.

Li, Qi and Jeffrey Scott Racine (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

Mitchell, James and Stephen G Hall (2005). "Evaluating, Comparing and Combining Density Forecasts using the KLIC with an Application to the Bank of England and NIESR Fan Charts of Inflation". *Oxford Bulletin of Economics and Statistics* 67, pp. 995–1033.

Ohnsorge, Franziska L, Marc Stocker, and Modeste Y Some (2016). "Quantifying Uncertainties in Global Growth Forecasts". World Bank Policy Research Working Paper 7770.

Pettenuzzo, Davide and Francesco Ravazzolo (2016). "Optimal Portfolio Choice Under Decision-Based Model Combinations". *Journal of Applied Econometrics* 31.7, pp. 1312–1332.

Prasad, Ananthakrishnan, Selim Elekdag, Phakawa Jeasakul, Romain Lafarguette, Adrian Alter, and Changchun Wang (2019). "Growth at Risk: Concept and Application in IMF Country Surveillance". International Monetary Fund Working Paper 19/36.

Ranjan, Roopesh and Tilmann Gneiting (2010). "Combining Probability Forecasts". *Journal of the Royal Statistical Society* 72.1, pp. 71–91.

Rey, Hélène (2015). "Dilemma not Trilemma: the Global Financial Cycle and Monetary Policy Independence". National Bureau of Economic Research Working Paper No. 21162.

Rossi, Barbara (2013). "Advances in Forecasting under Instability". *Handbook of Economic Forecasting.* Vol. 2. Elsevier, pp. 1203–1324.

— (2014). "Density Forecasts in Economics, Forecasting and Policymaking". Citeseer.

Rossi, Barbara and Tatevik Sekhposyan (2019). "Alternative Tests for Correct Specification of Conditional Predictive Densities". *Journal of Econometrics* 208.2, pp. 638–657.

Silverman, Bernard W (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Stone, Mervyn (1961). "The Opinion Pool". *The Annals of Mathematical Statistics*, pp. 1339–1342.

Tay, Anthony S and Kenneth F Wallis (2000). "Density Forecasting: A Survey". *Journal of forecasting* 19.4, pp. 235–254.

Waggoner, Daniel F and Tao Zha (2012). "Confronting Model Misspecification in Macroeconomics". *Journal of Econometrics* 171.2, pp. 167–184.

Wallis, Kenneth F (2005). "Combining Density and Interval Forecasts: A Modest Proposal". *Oxford Bulletin of Economics and Statistics* 67, pp. 983–994.

Appendix A. Sample and Data Sources

Table 1 lists the economies in the sample along with each economy's share of world output as of 2018q4, and Table 2 lists the data sources used in the paper.

Table 1: Sample

Economy	PPP share in 2018q4
Large economies	45.25
China	18.66
Euro area	11.40
United States	15.19
Regional leaders	22.22
Brazil	2.50
India	7.74
Japan	4.12
Mexico	1.90
Russia	3.12
South Africa	0.58
United Kingdom	2.26
Smaller economies	19.50
Argentina	0.68
Australia	0.97
Belarus	0.14
Canada	1.36
Chile	0.36
Colombia	0.55
Croatia	0.08
Czech Republic	0.29
Denmark	0.22
Hong Kong SAR China	0.35
Hungary	0.23
Indonesia	2.58
Israel	0.25
Jordan	0.07
Kazakhstan	0.38
Korea	1.65
Malaysia	0.75
Moldova	0.02
Norway	0.29
Peru	0.34
Philippines	0.70
Poland	0.90
Romania	0.38
Serbia	0.09
Singapore	0.42
Sweden	0.41
Switzerland	0.41
Taiwan Province of China	0.92
Thailand	0.98
Turkey	1.70
Ukraine	0.29
Venezuela	0.23
Vietnam	0.52

Table 2: Data Sources

Variable	Source
Business confidence	Federal Reserve Economic Data for the US; and Haver Analytics for other countries
Consumer confidence	Federal Reserve Economic Data for the US; and Haver Analytics for other countries
Commodity terms of trade	Gruss and Kebhaj (2019)
Economic policy uncertainty	Baker et al. (2016)
Financial Condition Index	Federal Reserve Economic Data for the US; and IMF, Money and Capital Market Department database for other countries
Geopolitical uncertainty	Caldara and Iacoviello (2018)
Import demand	IMF, Global Assumption Dataset
Industrial production	Fred for the US; and Haver Analytics for other countries
Population	IMF, World Economic Outlook
PPP share	IMF, World Economic Outlook
Purchasing managers' index	Haver Analytics
Real effective exchange rate	IMF staff calculations
Real GDP	IMF, World Economic Outlook