

Empirically-transformed Linear Opinion Pools*

Anthony Garratt
(University of Warwick)

Timo Henckel
(ANU and CAMA)

Shaun P. Vahey
(University of Warwick and CAMA)

September 30, 2021

Abstract

The Linear Opinion Pool (LOP) produces potentially non-Gaussian combination forecast densities. In this paper, we propose a computationally-convenient transformation for the LOP to mirror the non-Gaussianity exhibited by the target variable. Our methodology involves a Smirnov transform to reshape the LOP combination forecasts using the empirical cumulative distribution function. We illustrate our Empirically-transformed Opinion Pool (EtLOP) approach with an application examining quarterly real-time forecasts for U.S. inflation evaluated on a sample from 1990:1 to 2020:2. EtLOP improves performance by approximately 10% to 30% in terms of the continuous ranked probability score across forecasting horizons.

JEL codes: C32; C53; E37

Keywords: Density forecast combination; linear opinion pool; Smirnov transform; inflation.

*We thank Todd Clark, Domenico Giannone, Dean Croushore, Kevin Lee, Craig Thamotheram, Liz Wakerly, Mike McCracken, Yunyi Zhang, Anastasia Allayioti, two referees and participants at the National Bank of Belgium Real-time Economics Conference 2019 for helpful comments and suggestions.

1 Introduction

The literature on opinion pooling has examined extensively the accuracy of Linear Opinion Pool (LOP) forecast densities for macroeconomic variables; see the discussion in Rossi (2019). The LOP ensures that the shape of the combination has the scope to be more flexible than the individual forecast densities being combined. Via a variant of the “wisdom of the crowds”, non-Gaussian distributional features of the sample can be approximated, even if the individual experts utilise linear and (approximately) Gaussian reduced-form models, such as Vector Autoregressions (VARs) of macroeconomic variables.

In this paper, we propose a new methodology to improve the matching of the LOP to the marginal distribution of the target variable. Our approach involves applying a (modified) Smirnov transform to reshape the LOP combination forecasts using the empirical cumulative distribution function.

We illustrate our methodology with an example for U.S. inflation. Since we aim to study the scope for performance gains from opinion pools in the presence of misspecification, we consider a VAR model space, with misspecified elliptical errors. Each expert uses a unique VAR to produce “real-time” multi step ahead approximately Gaussian forecast densities for U.S. inflation. Jore et al. (2010), Garratt et al. (2011) and Rossi and Sekhposyan (2014) consider closely related density forecasting exercises with many misspecified VAR models.

We compare the combination forecast densities from both the Empirically-transformed Opinion Pool (EtLOP) and the benchmark LOP using a quarterly evaluation sample from 1990:1 to 2020:2. Relative to the conventional LOP, the EtLOP improves forecast performance by around 10% to 33% in terms of the Continuous Ranked Probability Score (CRPS). For longer horizons, the performance gains are somewhat larger than for the one step ahead forecasts. Furthermore, the EtLOP forecast densities exhibit greater asymmetry and heteroskedasticity than the benchmark LOP and provide

more plausible probabilistic assessments of U.S. inflation events, particularly prior to the Great Recession. Hence, our applied work demonstrates that our methodology improves the opinion pool’s forecasting performance by mirroring the non-Gaussian characteristics of inflation.

Economically important professional forecasters and policymakers use non-Gaussian forecast densities to communicate risks. For example, individual experts in the Survey of Professional Forecasters report non-Gaussian predictive densities and the Bank of England has published “fan charts” since 1997 for various macroeconomic variables. Cogley, Morozov and Sargent (2005) show that individual VAR specifications produce approximately symmetric forecast densities with stochastic volatility. Arguably then, the mid-2000s default reduced-form macroeconomic forecasting methodology is hard to reconcile with the Bank of England’s published forecasts from the perspective of an individual model.¹ An important implication of our study is that a forecaster estimating many misspecified VARs, or running a research team considering many VARs, could match the non-Gaussianity of the target variable by using EtLOP, rather than LOP, in the aggregation step. In effect, through this aggregation the “decision maker” introduces a form of distributional judgment—somewhat at odds with the common but incorrect distributional assumption of the experts embodied in the model space.

A number of academic studies have explored the scope for individual nonlinear and non-Gaussian time series models to improve forecast performance. Recent contributions include the copula modelling approach of Smith and Vahey (2016) and the quantile regression based methodology developed by Adrian et al. (2019). In contrast, Carriero, Clark and Marcellino (2020) argue that stochastic volatility models accommodate sufficient asymmetries for effective density forecasting in practice. In this study, we remain agnostic on the debate about the best single model to forecast inflation (and other key variables), and instead focus on the scope to improve the accuracy of

¹Galvao et al. (2021) and Allayioti (2020) find that survey information improves the density forecasting accuracy for a single macroeconomic model and small ensembles of models.

forecast densities produced from opinion pools, where the experts utilise misspecified linear and (approximately) Gaussian models.

Turning to the extant opinion pooling literature, a number of studies have noted calibration issues can arise with LOP forecast densities. Even if the individual experts have correctly calibrated exactly Gaussian forecast densities, the LOP combination will not generally be correctly calibrated; see, Ranjan and Gneiting (2010) and Gneiting and Ranjan (2013). In practice, the LOP tends to add diffusion to the combination density.² Focusing on the second moment of the conditional densities from LOP, Ranjan and Gneiting (2010) and Gneiting and Ranjan (2013) propose Beta transforms to reduce the spread. Extensions are explored by Bassetti et al. (2018) and Ganics (2017).

In terms of methodology, our EtLOP approach builds on empirical copula papers by, among others, Deheuvals (1979), Deheuvals (1981), Velásquez-Giraldo et al. (2018) and Coe and Vahey (2020) by fitting marginal distributions with non-parametric methods. Recent macroeconomic applications with semi-parametric copulas utilising non-parametrically fitted Empirical Cumulative Distribution Functions (ECDFs) include Smith and Vahey (2016), Karagedikli et al. (2019), Amengual et al. (2020) and Loaiza-Maya and Smith (2020). In contrast, Odendahl (2018) uses a parametric copula to model the multivariate dependence in the aggregate SPF.

Even though we adapt copula methods in our study, the EtLOP approach does not involve fitting the dependence in the combination. This is a natural approach given that the LOP ignores the dependence between experts, with computational advantages with a large number of experts.

The remainder of this paper is structured as follows. In Section 2, we set out our methodology for empirically-transformed opinion pools. In Section 3, we apply our methodology to a simulated example and in our application forecasting U.S. inflation. In the final section, we draw some conclusions.

²There exists a special case where the LOP is appropriately weighted to select a single “correct” expert. Arguably these conditions never arise in applied macroeconomic applications with experts using misspecified models subject to “uncertain instabilities”.

2 A Framework for Opinion Pooling

In this section, we present the details of our proposal to empirically transform the predictive densities from the LOP. We describe briefly conventional opinion pooling and contrast with our own approach, before discussing some practical considerations.

2.1 Conventional Opinion Pooling

In the opinion pooling framework, aggregation by a “decision maker” ignores how the individual experts make predictions. In effect, the decision maker combines out-of-sample forecasts (only) for the target variable from the experts. For example, for a one step ahead forecast, the LOP aggregation gives:

$$p^{LOP}(\pi_\tau) = \sum_{j=1}^J w_{j,\tau} g(\pi_\tau | I_{j,\tau}), \quad \tau = \underline{\tau}, \dots, \bar{\tau}, \quad (1)$$

where $g(\pi_\tau | I_{j,\tau})$ are one step ahead forecast densities from expert j , $j = 1, \dots, J$, for the target variable π_τ (inflation in our application), conditional on the information set $I_{j,\tau}$. The publication delay in the production of real-time macroeconomic data ensures that this information set contains lagged variables, here assumed to be dated $\tau - 1$ and earlier. The non-negative weights, $w_{j,\tau}$, in this finite mixture sum to unity and potentially change through time in the evaluation period $\tau = \underline{\tau}, \dots, \bar{\tau}$; see the discussion in, for example, Garratt et al. (2014). Multi-step forecasting (by either direct or iterative methods) and various weight schemes (including time varying weights) have been extensively explored in the literature.

2.1.1 Illustrative example

An example provided by Kascha and Ravazzolo (2010) helps illustrate visually the capacity for the LOP to introduce non-Gaussianity into the combination even with Gaussian experts. The example considers a single observation of the target variable with two experts, where LOP utilises equal

weight aggregation, to mimic the approach of FRB Philadelphia for the Survey of Professional Forecasters (SPF). Both panels of Figure 1 plot the experts' forecasts, where the prediction of Expert 1 (blue, dash and dot line) has mean -2.0 and standard deviation 1 and that of Expert 2 (red, dash line) has mean 2.0 and standard deviation 2.0. The top panel also plots the LOP density (black, solid line) which is bimodal, with a slightly higher peak associated with the forecast mean of Expert 1.

This simple example illustrates several relevant features of conventional opinion pooling with Gaussian experts. First, although the experts' forecast densities are individually Gaussian, the combined LOP density is non-Gaussian. Second, the LOP tends to preserve disagreement across experts about the location of the central probability mass. Hence, the LOP does not inherit the Gaussian distributional beliefs displayed in each expert's forecast. Moreover, this form of the LOP, used in practice for the SPF, introduces no information from the history of the target macroeconomic variable (other than as captured in the experts' forecasts).

2.1.2 Discussion

Even though the SPF aggregation by FRB Philadelphia uses equal weights, many studies have examined whether "recursive" and "optimal" weighting improves the LOP's predictive accuracy and calibration for macroeconomic data. For example, Geweke and Amisano (2011) argue for optimal combinations based on maximising the Kullback-Leiber Information Criterion (KLIC), and Jore et al. (2010) report stronger forecasting performance from LOP with recursively updated weights based on the logarithmic score of each expert. In both cases, the history of the target variable influences the weights.

Nevertheless, the FRB Philadelphia's equal weight approach has considerable empirical support. In the short samples typical of most macroeconomic forecasting applications, many studies have noted that equal weights perform as well as more complex weighted schemes. See, for example, the

discussions in Timmermann (2006) and Diebold and Shin (2018) on the difficulty of estimating weights in short samples.

Even the recursive weights based on relative forecast performance used by (among others) Jore et al. (2010) tend to give little variation across experts in very short samples. As a result, the LOP forecasts with a large number of experts and relatively little disagreement across experts often exhibit approximately Gaussian features.³

On the other hand, in long samples of, say, high-frequency financial time series data, recursive weights can result in expert selection, where one expert dominates. In these circumstances, the functional form of the conventional LOP densities tend to mirror those of the dominant expert. As noted by Geweke and Amisano (2011), optimisation based on the KLIC offers scope to weight experts more evenly, typically generating more complex forecast densities from the LOP as a result. Aastveit et al. (2019) and Rossi (2019) provide discussions of various weighting schemes aimed at improving forecast performance.

Ranjan and Gneiting (2010) and Gneiting and Ranjan (2013) demonstrate that despite “optimal” weights across experts—as constructed by Amisano and Geweke (2011)—and individually well-calibrated experts, the LOP will not typically be correctly calibrated in terms of the Probability Integral Transforms (PITs) being uniformly distributed.⁴ The LOP aggregation approach tends to amplify the spread of the forecast densities in practice.⁵ Gneiting and Ranjan’s proposal for the Beta Opinion Pool (BLOP) suggests a correction to the second moment of the conditional densities from the conventional LOP.

We stress that our EtLOP methodology is aimed at improving the higher-order moments of the conditional forecast densities. (Nevertheless, by matching the non-Gaussianity in the marginal

³Knüppel and Krüger (2021) propose improving the LOP by removing the disagreement between experts.

⁴And, for one step ahead forecasts, having *i.i.d* forecast errors.

⁵Given that parsimonious macro-economic models tend to produce under-dispersed forecast densities, and therefore over-confident experts, this feature perhaps contributes to the forecasting performance of the SPF with an equal weight LOP.

distribution for the target variable, EtLOP also influences the spread of the forecast densities, as we discuss in Section 2.2.) Similarly, BLOP offers scope to accommodate simple forms of non-Gaussianity because the Beta distribution has two parameters to shape the forecast densities.

Accordingly, we experimented with the approach for our application but did not find the methodology to be effective in practice, either in terms of calibration or relative density forecasting performance. This perhaps reflects the more complex form of the departures from Gaussianity displayed in our sample for U.S. inflation. We describe our BLOP simulations and results for our application in an online appendix.

2.2 Empirically-transformed Pooling

Recall that our EtLOP methodology involves reshaping the combination forecast densities from the LOP using a fitted marginal distribution for the target variable. The ECDF provides a convenient choice for the marginal distribution, being a step function that represents the entire history of the observations for the target. By construction, the ECDF is marginally calibrated. But, this does not necessarily mean that the PITS of the conditional forecast densities from EtLOP will be appropriately distributed. Among others, Rosenblatt (1952), Diebold et al. (1998), Galbraith and van Norden (2012) and Rossi and Sekhposyan (2019) discuss (what is usually known as “probabilistic”) calibration and the relationship to the properties of the PITS.

Given our choice for the marginal distribution of the target variable, $F(\pi_t)$, our aim is to reshape the LOP combination density forecast by adapting methods developed for pseudo-random number generation. The Smirnov transform allows a researcher to generate a conditional density forecast with the same distribution as a (stationary) target variable via the inverse ECDF, $F^{-1}(\pi_t)$. The approach is often used to generate pseudo-random numbers from a known but non-parametric distribution. In the empirical copula literature, the transform provides a computationally convenient route

for prediction from a non-parametric copula density with non-parametric marginal distributions.⁶ Because the inverse of the ECDF is used, the methodology is sometimes referred to as “inverse transform sampling”.

The idea behind EtLOP is to use the Smirnov transform, and the target variable’s ECDF, to generate a conditional forecast density with the same distribution as the target variable from the LOP. However, the Smirnov transform step requires that the LOP forecast densities first be transformed to the (0,1) interval; and, the various aggregation issues discussed above imply that the distribution of the LOP is unknown but non-Gaussian. We proxy this unknown distribution by using the entire history of LOP forecasts.

For expositional ease, we describe our algorithm for a one step ahead forecast case, considering a single candidate LOP combination forecast density, $p^{LOP}(\pi_t)$, for one observation of the target variable, π_t , given the history for the target variable, $\pi_1 \dots \pi_{t-1}$, and the history of the LOP forecast densities, $p^{LOP}(\pi_1), \dots, p^{LOP}(\pi_{t-1})$.⁷

We break our EtLOP algorithm into four steps.

1. Construct the proxy empirical CDF, $\phi(\cdot)$, for the LOP forecast density from the history of LOP forecast densities. Computationally, this involves pooling (equal weighted) draws (iterates) from the extant historical ensemble of LOP forecasts.
2. Convert the candidate LOP forecast density, $p^{LOP}(\pi_t)$, to the unit interval using $\phi(\cdot)$. To achieve this in a computationally convenient manner, we rank draws from $p^{LOP}(\pi_t)$ such that $r_t = R_t/(N + 1)$, where R_t denotes the rank of each draw within the historical distribution, ϕ , and N is the total number of draws from that distribution.⁸

⁶Random number generation for a known parametric distribution utilises a parametric CDF instead of the ECDF.

⁷Our forecasting U.S. inflation application that follows extends consideration to multiple forecast origins and horizons.

⁸The denominator avoids boundary issues in the subsequent Smirnov transform.

3. Fit the ECDF for the target variable, $F(\pi_1, \dots, \pi_{t-1})$. In practice, there are a number of ways to do this, as we discuss below, but non-parametric methods are a pragmatic choice given the unknown distribution of the target variable.
4. Convert the candidate LOP forecast defined on the unit interval, r_t , to the observed scale using the inverse ECDF, $F^{-1}(\cdot)$, for the target variable. This Smirnov transform involves mapping the ranked draws onto the observed scale of the forecast target variable.

2.2.1 Illustrative example

We illustrate the impact of our EtLOP algorithm by re-considering the Kascha-Ravazzolo example using two Gaussian experts. Recall, both panels of Figure 1 plot the experts' forecasts as densities, where the prediction of Expert 1 (blue, dash and dot line) has mean -2.0 and standard deviation 1 and that of Expert 2 (red, dash line) has mean 2.0 and standard deviation 2.0. The upper panel also plots the LOP density (black, solid line); whereas the lower panel plots the EtLOP density (black, solid line). As noted in the algorithm description, EtLOP requires the extant histories of the target variable and the LOP forecasts. For illustrative purposes, we used the end-sample objects from our inflation forecasting example to produce the EtLOP density plotted in the lower panel of Figure 1.⁹

The EtLOP density forecast preserves the unimodality of the individual (Gaussian) densities, with the peak relatively close to the forecast mean of Expert 1, with visible asymmetry and a long right tail. This contrasts with the conventional LOP forecast density, plotted in the upper panel, which is bimodal.

2.2.2 Discussion

Considering the Kascha-Ravazzolo example illustrates the contrast between the EtLOP and LOP approaches to opinion pooling. Although the experts' forecast densities are individually Gaussian,

⁹We describe the non-parametric methods used to fit the ECDF, $F(\cdot)$, below.

the LOP forecast density is non-Gaussian, and the EtLOP suggests that the data do not support bimodality, but do support a degree of asymmetry. Put differently, the LOP preserves the disagreement between experts about the central probability mass, whereas the EtLOP consolidates, with greater probability mass between the two experts’ densities. The example also demonstrates the scope for EtLOP to produce less diffuse forecast densities than LOP. This reflects the marginal calibration of the ECDF for the target variable. Although this does not guarantee probabilistic calibration (in terms of the distribution of the PITS), it often helps in practice.

Throughout our applied work, and in the Kascha-Ravazzolo example above, we use a non-parametric method to fit the ECDF. The non-parametric approach is a pragmatic modelling choice given the unknown distribution of the target variable in practice. We fit the ECDF for the target variable with the SSV locally adaptive kernel density estimator by Shimazaki and Shinomoto (2010).¹⁰ Figure 2 plots the density for inflation in our U.S. sample based on the full sample, which displays considerable asymmetry and a relatively long right tail. As a rough guide to non-Gaussianity, the Shapiro-Wilk test rejects the null of normality with a p-value of zero. We emphasise that in our subsequent application, we follow the standard approach in the “real time” macroeconomic literature, fitting all models and the non-parametric margin to data vintages in the public domain at the forecast origin. As a result, the actual fitted ECDF evolves with the expanding window in the analysis. Since the non-parametrically fitted ECDF is (typically) far from Gaussian for any given window of observations, the EtLOP algorithm adds a limited form of heteroskedasticity to the aggregate forecast density.

Of course, there is scope to restrict the marginal distribution to be Gaussian distributed. We explored this variant in our applied work and discovered that while the resulting forecast densities were better calibrated than the benchmark LOP, the non-parametric approach was strongly preferred

¹⁰Non-parametric kernel fitting used the MATLAB function `ksdensity` in an earlier draft, which gave similar forecast performance.

on our inflation sample.¹¹

We emphasise that, as with the LOP, our EtLOP approach does not estimate the dependence structure between experts. Under the information assumptions of conventional linear opinion pooling, the decision maker assumes that the experts' information sets are conditionally independent; see, for example, the discussion in DeGroot and Mortera (1991).

3 Simulation and Application

We now illustrate our approach by exploring a simulation, before turning to our application considering EtLOP forecasts for U.S. inflation.

3.1 Simulation Example

In this simulation, we adapt an example of Gneiting and Ranjan (2013, section 4.1) to consider a non-Gaussian distributed target variable, matching the historical features of U.S. inflation considered in our application. We begin by summarising briefly the baseline Gneiting-Ranjan experiment and then discuss our variation, together with our results for EtLOP and LOP.

3.1.1 Baseline experiment

The Gaussian Data Generating Process (DGP) considered by Gneiting and Ranjan (2013) is:

$$Y = X_0 + a_1X_1 + a_2X_2 + a_3X_3 + \epsilon \tag{2}$$

where $X_i, i = 1, \dots, 3$ denote the random and independent variables, a_i denote the respective parameters and the disturbance term ϵ is *i.i.d.* standard normal, $\epsilon \sim \mathcal{N}(0, 1)$. The independent

¹¹The EtLOP should perhaps be known as the GLOP in this case.

variables are also *i.i.d.* standard normal.

The three individual experts observe some but not all variables. For example, Expert f_1 observes X_0 and X_1 , but does not observe X_2 and X_3 so that the forecast densities are:

$$f_1 = \mathcal{N}(X_0 + a_1X_1, 1 + a_2^2 + a_3^2). \tag{3}$$

Similarly, for the remaining experts:

$$f_2 = \mathcal{N}(X_0 + a_2X_2, 1 + a_1^2 + a_3^2), \tag{4}$$

$$f_3 = \mathcal{N}(X_0 + a_3X_3, 1 + a_1^2 + a_2^2), \tag{5}$$

where variable X_0 is observed by all experts.

Hence, by definition, all three experts produce well-calibrated forecast densities. Based on these experts' forecasts, it is straightforward to compute the forecast densities from the LOP and to apply the EtLOP algorithm. Following Gneiting and Ranjan (2013), we considered 500 observations in the experiment based on the parameter values $a_1 = a_2 = 1$ and $a_3 = 1.1$, for both a “train” and a “test” sample.

We generated similar test parameter values and PIT histograms to those reported by Gneiting and Ranjan (2013) for the experts and the LOP combination. Specifically, the experts' PITS appeared uniformly distributed consistent with probabilistic calibration, whereas the LOP's PITS displayed a hump-shaped distribution, indicating overly diffuse forecast densities (under-confidence). In contrast, with the Gaussian distributed target variable, the EtLOP fitted to the train sample generated approximately uniformly distributed PITS—like those of the individual experts.

As an additional check, we investigated the scope for optimised weights to improve the performance of the equal weight LOP. On the simulation test data, optimisation gave weight values close to those

reported by Gneiting and Ranjan (2013), with a small improvement over equal weights in terms of CRPS and the distributional characteristics of the PITS. We also fitted the BLOP proposed by Gneiting and Ranjan (2013), and obtained approximately similar parameter values and forecast performance gains over LOP with optimised weights

3.1.2 Experiment with an asymmetrically distributed target variable

To illustrate the performance of EtLOP for a variable that is not Gaussian distributed, we transformed the target variable from the Gneiting-Ranjan example as follows.

First, we fitted non-parametrically the ECDF, $F(\pi_t)$, to our inflation data exactly as described in Section 2.2.2; see the density plotted in Figure 2. Recall that the non-parametrically fitted empirical distribution is considerably more peaked than the Gaussian, and asymmetric with a long right tail. Second, we ranked the simulated 500 target variable observations, Y , from the baseline (Gaussian) experiment and divided by 501 (to avoid boundaries). And third, we used the inverse of the empirical CDF, $F^{-1}(\cdot)$ so that the transformed target variable, denoted \tilde{Y} , matched the distribution of the inflation sample—via the Smirnov transform.

We then repeated the Gneiting-Ranjan experiment, with the same experts’ parameters, but using the asymmetrically distributed target variable, \tilde{Y} , rather than Y , again considering both a “train” and a “test” sample, with the marginal distribution fitted to the train sample only to limit overfitting.

Figure 3 plots the (kernel smoothed) densities of the EtLOP’s sample averaged CRPS, based on the 2000 test samples, for simulation sample lengths 100, 150, 200, 250 and 500.¹² The x-axis displays the CRPS ratio, measuring relative forecast performance for EtLOP, where values of less

¹²We compute the CRPS using the Gneiting and Raftery (2004) representation, defined as:

$$CRPS(F, x) = E[|X - x|] - 0.5E[|X - X^*|]$$

where X and X^* are the iterates from the inflation forecast densities and a random permutation of these iterates respectively, and x is the inflation target variable, π_τ , we use to evaluate our forecasts.

than one indicates an improvement on the LOP with optimised weights.

There are two striking features from our simulations. First, regardless of which sample length we consider, the EtLOP is never inferior to the LOP benchmark forecasting model in terms of the CRPS. Second, the relative CRPS plots indicate that forecast performance is robust to sample size, with little variation in central location of the distribution across sample lengths. However, for smaller samples, the performance gain varies more across simulated samples, but where the central mass indicates considerable gain of around 10% to 15%.¹³

As noted in Section 2.1.2, Gneiting and Ranjan (2013) propose the BLOP to correct the second moment of the LOP forecast densities but this methodology also has some potential to accommodate departures from Gaussianity. Accordingly, we replicated our simulation with BLOP and compared the CRPS results with EtLOP and the LOP with optimised weights. The BLOP failed to match the forecast performance of EtLOP and, for small samples, had approximately similar performance to the LOP with optimised weights, with the forecast densities often poorly calibrated. We provide CRPS analysis for the BLOP in the online appendix.

3.2 Application: Forecasting U.S. Inflation

With the simulation as background, we apply the algorithm to conventional LOP densities for quarterly U.S. inflation, using the target dates from 1990:1 to 2020:2 as the evaluation sample. Throughout this section, we report results for the EtLOP with the equal weight LOP as a benchmark, to ensure consistency with the approach used by FRB Philadelphia for the SPF. As further aids to gauge the forecasting performance of the EtLOP, we report results for BLOP in the appendix, together with those from a univariate Unobserved Components Stochastic Volatility (UCSV) model

¹³We repeated the analysis using the sample averaged logarithmic score and found similar performance gains for EtLOP across our range of sample sizes. We also used the Rossi and Sekhposyan (2019) calibration test which confirmed that the EtLOP is (typically) well calibrated but the LOP with optimised weights is poorly calibrated for the majority of simulations.

for inflation, estimated by Bayesian methods.¹⁴

3.2.1 Experts' models

Each expert utilises a (unique) bivariate VAR model space for inflation, π_t , and the output gap (the deviation of real output from potential), y_t . The standard theory that aggregate demand, captured by the output gap, influences the movements in inflation (with unknown time lags), provides some foundation for the empirical specification.

Since we aim to study the scope for performance gains from opinion pools in the presence of misspecification, each expert's VAR is misspecified with elliptical errors.

The j^{th} VAR model takes the form:

$$\begin{bmatrix} \pi_t \\ y_t^j \end{bmatrix} = \begin{bmatrix} a_{\pi\pi}^j & a_{\pi y}^j \\ a_{y\pi}^j & a_{yy}^j \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ y_{t-1}^j \end{bmatrix} + \begin{bmatrix} \epsilon_{\pi t}^j \\ \epsilon_{y t}^j \end{bmatrix}, \quad t = 1, \dots, T, \quad (6)$$

where $[\epsilon_{\pi t}^j, \epsilon_{y t}^j]' \sim i.i.d. N(\mathbf{0}, \Sigma^j)$. That is, we consider a baseline VAR specification in which the output gap measure has been varied to give J linear and Gaussian VAR models, indexed $j = 1, \dots, J$. For expositional ease, we ignore the intercept and restrict the lag order of the J VARs to one. Following Garratt et al. (2011), our VAR model space uses seven output gap measures derived from the set of univariate off-model filters considered by Orphanides and van Norden (2002, 2005).

We define the output gap as the difference between observed output and unobserved potential (or the trend component of) output. We denote the (logarithm of) real output in t as q_t , and let μ_t^j be its trend using definition j , where $j = 1, \dots, J$. The output gap, y_t^j , is therefore defined as the difference between actual output and its j^{th} trend measure at time t . We assume the following linear trend-cycle decomposition:

$$q_t = \mu_t^j + y_t^j. \quad (7)$$

¹⁴The appendix describes the UCSV model.

The seven methods of univariate trend extraction in our VAR model space are: quadratic, Hodrick-Prescott (HP), forecast-augmented HP, Christiano and Fitzgerald, Baxter-King, Beveridge-Nelson and Unobserved Components. We describe these seven well-known univariate filters in Appendix 1.

In our application, we vary a single auxiliary assumption to generate the expert (model) space. Specifically, we vary the lag length in the VAR.¹⁵ If we have J output gap measures, and for any given y_t^j we have L variants defined by different values of the maximum lag length, then in total we have $J \times L$ models, each with a corresponding forecast of inflation (and the output gap) from the VAR model space. We restrict L to a maximum of four and therefore we consider 7×4 models—28 forecasts from the experts to be combined.

Although the motivation for deploying these models stems from their common usage by central banks around the world, Orphanides and van Norden (2005) note indifferent real-time out of sample forecasting performance for individual VAR models. In contrast, Garratt et al. (2014) note that LOPs of VARs provide a useful benchmark, albeit not as accurate as the univariate UCSV model for U.S. inflation.

3.2.2 Data considerations

Orphanides and van Norden (2002, 2005) stress that output gap measures are subject to considerable data revisions. Failing to account for this by using heavily-revised data can mask real-time predictive content. Since we are interested in real-time prediction, parameter estimation is recursive for all specifications. Each recursion uses a different vintage of data, where a vintage of data is the vector of time series observations available from a data agency at the forecast origin.

The quarterly real-time real gross domestic product (GDP) U.S. dataset has 124 vintages, with the first vintage dated 1990:1 and the last 2020:4. The raw data for GDP (in practice, Gross National

¹⁵For ease of exposition, we fixed this at one in equation (6).

Product, GNP, for some vintages) are taken from the Federal Reserve Bank of Philadelphia’s Real-Time Data Set for Macroeconomists. The data comprise successive vintages from the National Income and Product Accounts, with each vintage reflecting the information available around the middle of the respective quarter. Croushore and Stark (2001) provide a description of the real-time GDP database. The GDP deflator price series used to measure inflation is constructed analogously. We define inflation (output growth) as the first difference in the logarithm of the GDP deflator (GDP) multiplied by 400.

Figure 4 displays inflation (upper panel) and real output growth (lower panel) from 1970:1 to 2020:2 based on the final vintage of data.¹⁶ For the observations displayed during the Great Moderation inflation typically exhibits lower volatility and lower conditional mean than during the 1970s and early 1980s. However, during the run up to the Great Recession, between 2003 and 2006, there are several realisations of high inflation. The upward spikes apparent during this period are often regarded as (the response to) relative price movements, and, in particular, commodity prices. See, for example, the analysis of Garratt and Petrella (2021). A striking feature of the Great Recession and its aftermath is the increased threat of low inflation, and an apparent increase in volatility. The recent pandemic resulted in a downward spike for both inflation and output growth in early 2020.

3.2.3 Forecast combination and empirical transformation

The decision maker recursively combines the forecast densities from the experts. Each expert uses an expanding window for parameter estimation. For the first recursion, the estimation sample is 1970:1 to 1989:4 (window size 80 observations) and the last 1970:1 to 2020:1 (window size 201 observations).

As our U.S. GDP deflator data are released with a one quarter lag, the first vintage, dated 1990:1, contains time series observations from 1970:1 to 1989:4, and the last vintage, dated 2020:4, has

¹⁶The empirical analysis which follows uses a time sequence of vintages.

data from 1970:1 to 2020:3. Following Clark and McCracken (2010) and others, we use the second estimate as the target “final” data. For example, when evaluating the $h = 1$ forecast (nowcast) for 2020:2, we use the 2020:4 vintage observation of inflation for 2020:2.

Each VAR (expert) produces forecast densities for inflation (and the (j^{th}) output gap) through our evaluation period: $\tau = \underline{\tau}, \dots, \bar{\tau}$ where $\underline{\tau} = 1990:1$ and $\bar{\tau} = 2020:2$ (122 quarterly observations).

To deploy EtLOP, the decision maker must fit a marginal distribution for inflation, π_t . Since the decision maker uses recursive fitting based on expanding windows of data, the fitted distributions for inflation vary by forecast origin in practice but all appear non-Gaussian. The null hypothesis of normality is rejected at the 1% significance level for all vintages using the Shapiro-Wilk test.

We emphasise that in this application, we follow the standard approach in the “real time” macroeconomic literature, fitting all models and the non-parametric margin to data vintages in the public domain at the forecast origin. We do not fit any parameters, or the ECDF of the target variable, on ex post data.

In the following section, we compare and contrast the forecast performance of EtLOP with LOP. The performance metrics used to gauge forecast accuracy include RMSFE and the sample-averaged CRPS, together with tail-weighted CRPS metrics, and the probabilistic calibration test of Rossi and Sekhposyan (2019).

3.2.4 Results

In this section, we report results for the one through four ahead forecasts. All results reported here use the equal weight LOP as a benchmark.¹⁷

The second column of Table 1 reports the RMSFE and the third column reports the time-averaged CRPS over the evaluation sample. Columns 4 to 6 give the tail-weighted, right tail-weighted, and

¹⁷Recursive weighted combinations based on the sample-averaged logarithmic score and sample averaged CRPS gave similar results.

left tail-weighted CRPS, respectively. The values displayed in columns 2 to 6 are computed as ratios to the LOP benchmark. Ratios less than one, for both RMSFE and CRPS, indicate an improvement in forecast performance, relative to the LOP benchmark model.

The RMSFEs reported in the second column indicate a gain for the EtLOP of approximately 13% over the LOP benchmark for horizon $h = 1$, displayed in row 2. As the forecast horizon lengthens, rows 3 through 5, the performance gain from EtLOP increases monotonically to 34%. Similarly, the CRPS values reported in column 3 indicate a comparable performance gain from EtLOP, monotonically increasing from 14% at $h = 1$, row 2, to 33% at $h = 4$, row 5.

Turning to the tail-weighted CRPS performance, columns 4 through 6, EtLOP performance is fairly consistent when considering both tails or just the right tail, columns 4 and 5, whereas the left tail-weighted CRPS results in a slightly smaller performance gain, column 6.

Figure 5 provides a visual confirmation of the $h = 1$ performance differentials in terms of the levels of RMSFE, displayed in the upper panel, and CRPS, displayed in the lower panel, computed recursively through the evaluation sample. Regardless of the forecast performance metric, the EtLOP (red, solid line) dominates the LOP (blue, dash line) from 1992 onwards. Plots for the other forecast horizons display similar patterns. The online appendix provides analysis for $h = 4$.

Figure 6 reports the (end-sample) $h = 1$ PITs distributions for EtLOP, upper panel, and LOP, lower panel. These are close to uniformly distributed for EtLOP, consistent with good probabilistic calibration, whereas the PITs for LOP look consistent with poor probabilistic calibration, with too many realisations in the left tail of the forecast density, and too few in the right tail. The calibration tests of Rossi and Sekhposyan (2019), for $h = 1$ and $h = 4$, confirm that EtLOP is well calibrated, with test statistics of 0.546 and 1.23 respectively, which do not reject the null of good calibration. The same tests confirm LOP's poor calibration, with values of 2.085 and 3.803 respectively, both of which reject the null of good calibration.¹⁸

¹⁸The non-rejection and rejection of the null of good calibration, for EtLOP and LOP respectively, is also true at

To summarise the results so far, EtLOP outperforms the LOP benchmark. To give further context, using the same metrics, EtLOP also outperformed the Bayesian estimated UCSV model and the BLOP at all horizons. The analysis is described in the online appendix.

To provide additional insight into EtLOP’s forecast performance, focusing on the $h = 1$ case, the upper and lower panels of Figure 7 display the 5% and 95% bands of the forecast densities, together with the conditional mean forecasts and observed inflation, for the EtLOP and the LOP, respectively. EtLOP’s bands are narrower than the LOP’s 90% for nearly all of the evaluation sample but there is variation throughout the evaluations.

Figure 8 complements Figure 7, with four panels, displaying (from upper to lower panel) for both $h = 1$ combination densities: the differences between the 5% and 95% percentiles of the forecast density, a measure of skew, the p-value for skew, and the probability of inflation being less than 2.6%, respectively. The top panel supports the assessment of Figure 7 that the EtLOP densities (red, solid line) are less diffuse than the LOP densities (blue, dash line) and that the EtLOP’s densities display greater time variation, especially prior to 2000. The second panel in Figure 8, which plots skew, reveals EtLOP to have positive skew for much of the evaluation, whereas LOP has near zero skew throughout. The p-values displayed in the third panel confirm the significance of EtLOP’s skew from 1990 to 2005, but thereafter is often not significantly different from zero. The fourth panel reveals that the EtLOP assigns a higher probability than the LOP to the event of inflation being less than the unconditional mean, 2.6%, throughout the evaluation. The EtLOP assesses higher risk of below mean inflation events.

Considering Figures 7 and 8 reveals that the EtLOP $h = 1$ forecast densities have a degree of skew, whereas the LOP forecast densities are approximately symmetric. Furthermore, a careful comparison reveals that the EtLOP forecast densities are less diffuse, with more variation in the diffusion through time. Both features are less apparent after the Great Recession.

forecast horizons $h = 2$ and $h = 3$.

Finally, Figure 9 displays the forecast densities for EtLOP (red, solid line) and LOP (blue dash line) for the target observations of 2009:1 through to 2009:4, when inflation was unusually low. The EtLOP densities have less probability mass on high inflation and are somewhat less diffuse than their conventional counterparts. The EtLOP displays some asymmetry in the four panels, but the skew is not strong—according with the analysis of Figure 8.

4 Conclusions

In this paper, we have proposed a methodology to improve the accuracy of the LOP. Our approach involves transforming the conventional combination forecast densities using an ECDF to match the distribution of the sample data. In our U.S. inflation application, we combined forecast densities from a system of VAR models. We demonstrated that the Empirically-transformed LOP considerably improved forecasting performance relative to the more conventional opinion pool.

Figure 1: Opinion Pools and Expert's Density Forecasts

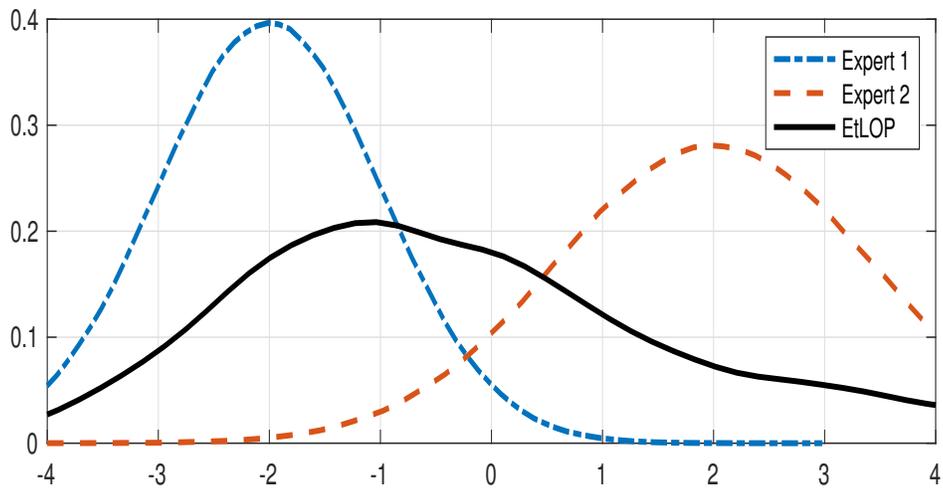
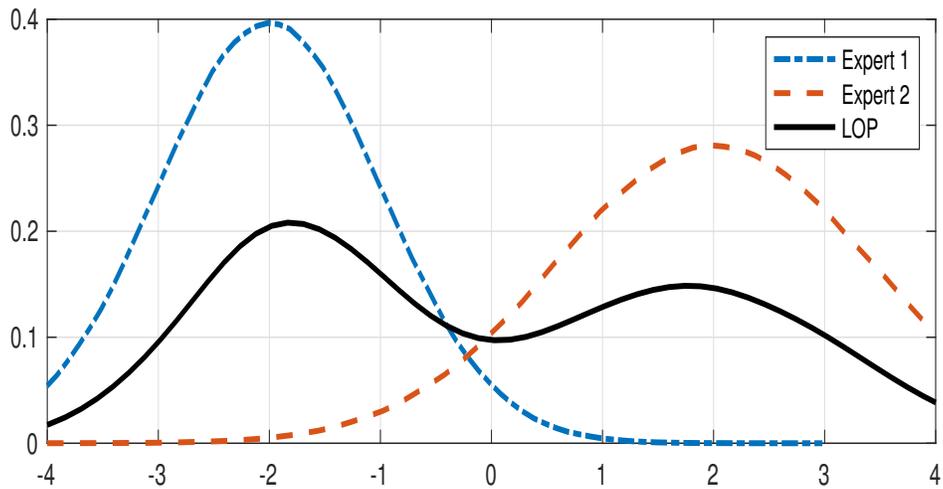


Figure 2: Marginal Density for U.S. Inflation, 1970:1 to 2020:2

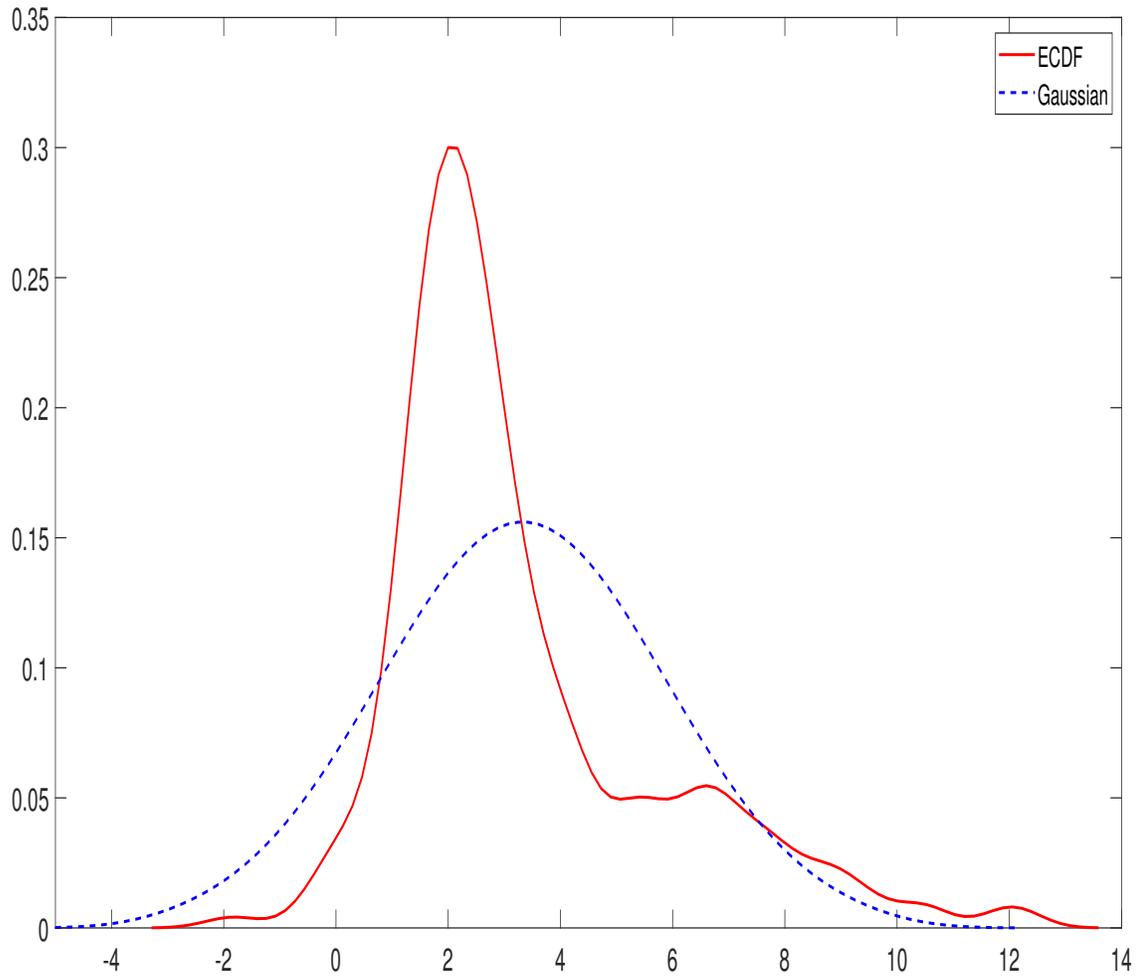


Figure 3: EtLOP's Relative CRPS Performance, Simulation, By Sample Size

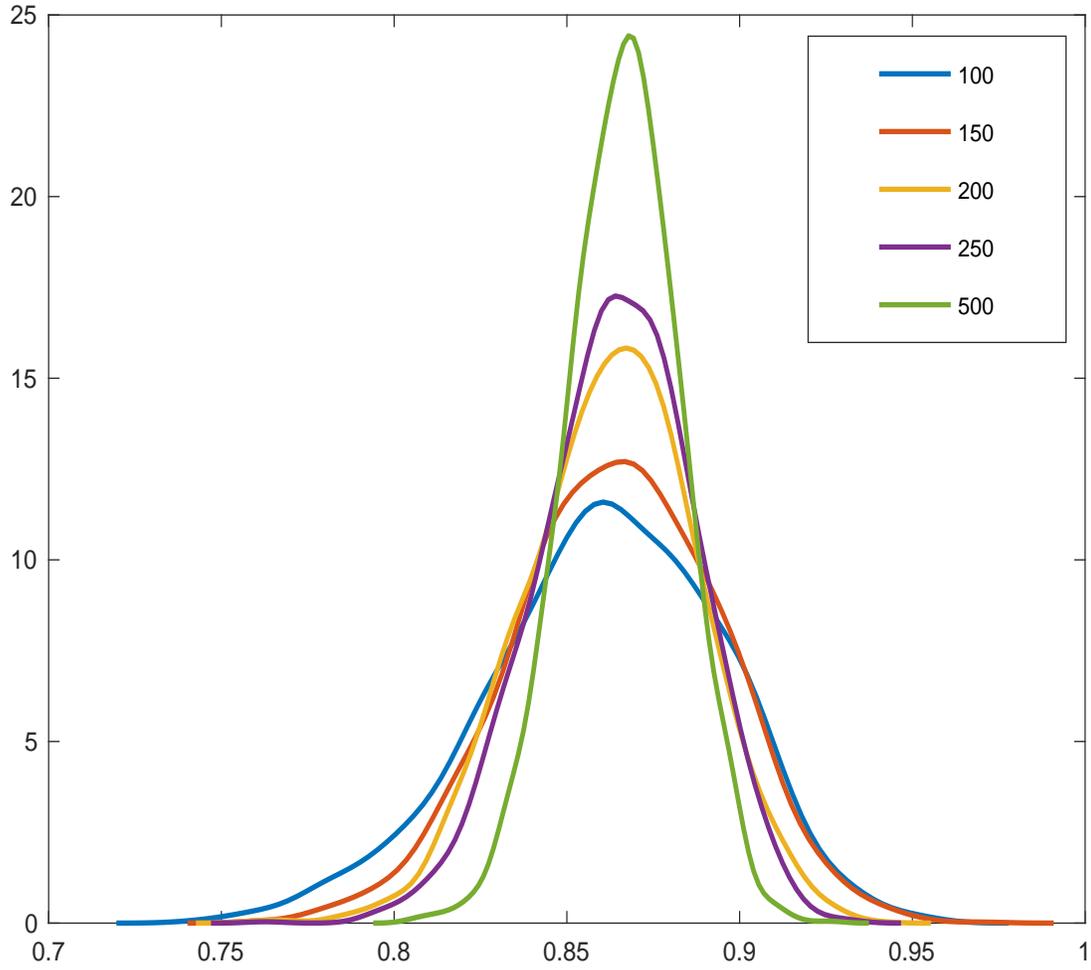


Figure 4: U.S. Inflation and Real Output Growth, 1970:1 to 2020:2

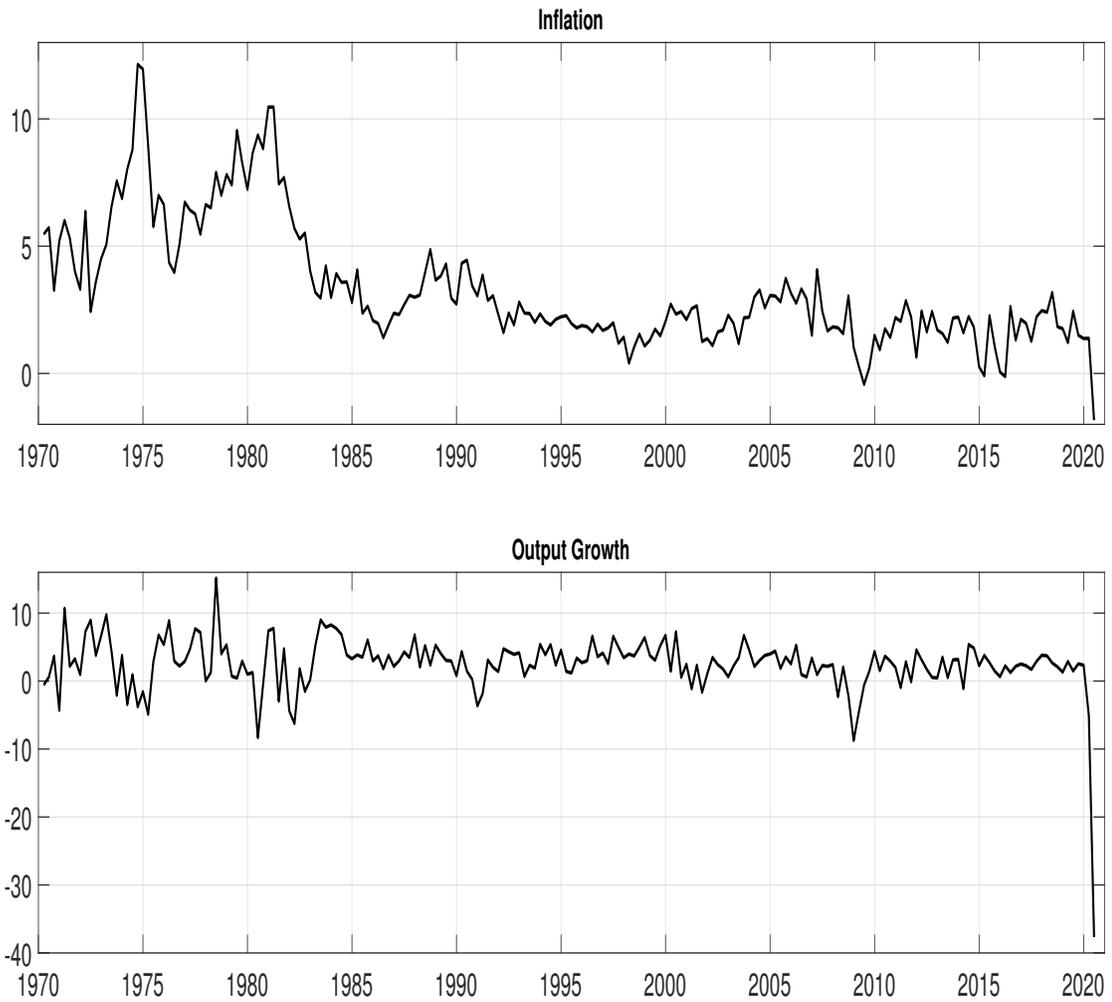


Figure 5: Recursive Forecast Performance, $h=1$, 1990:1 to 2020:2

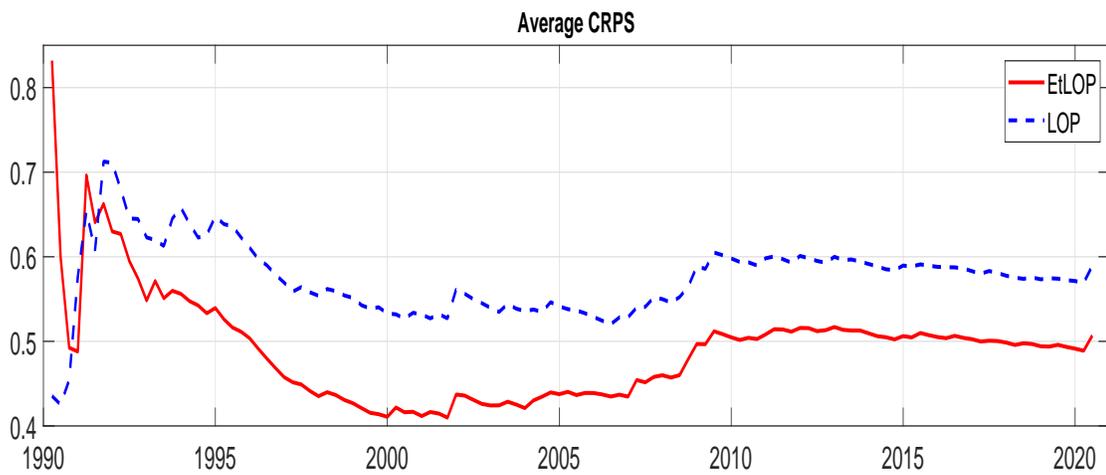
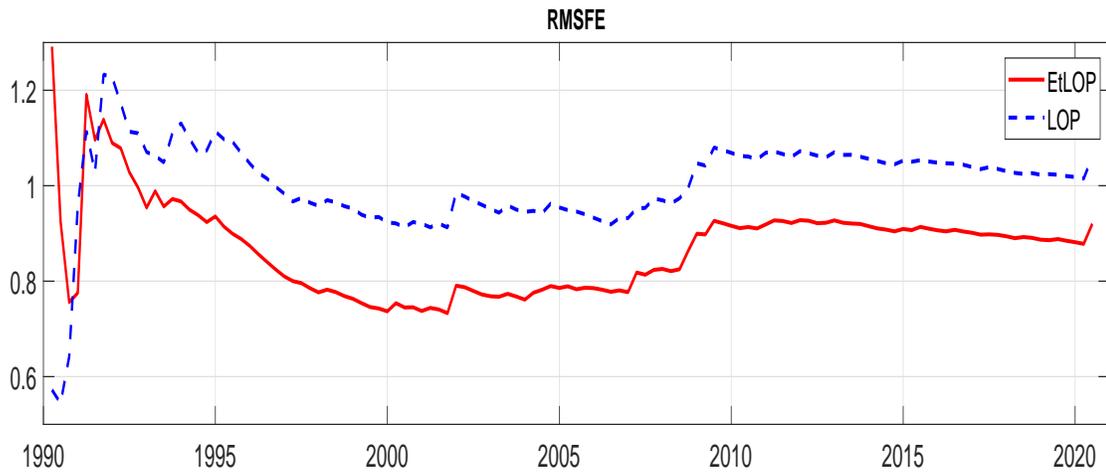
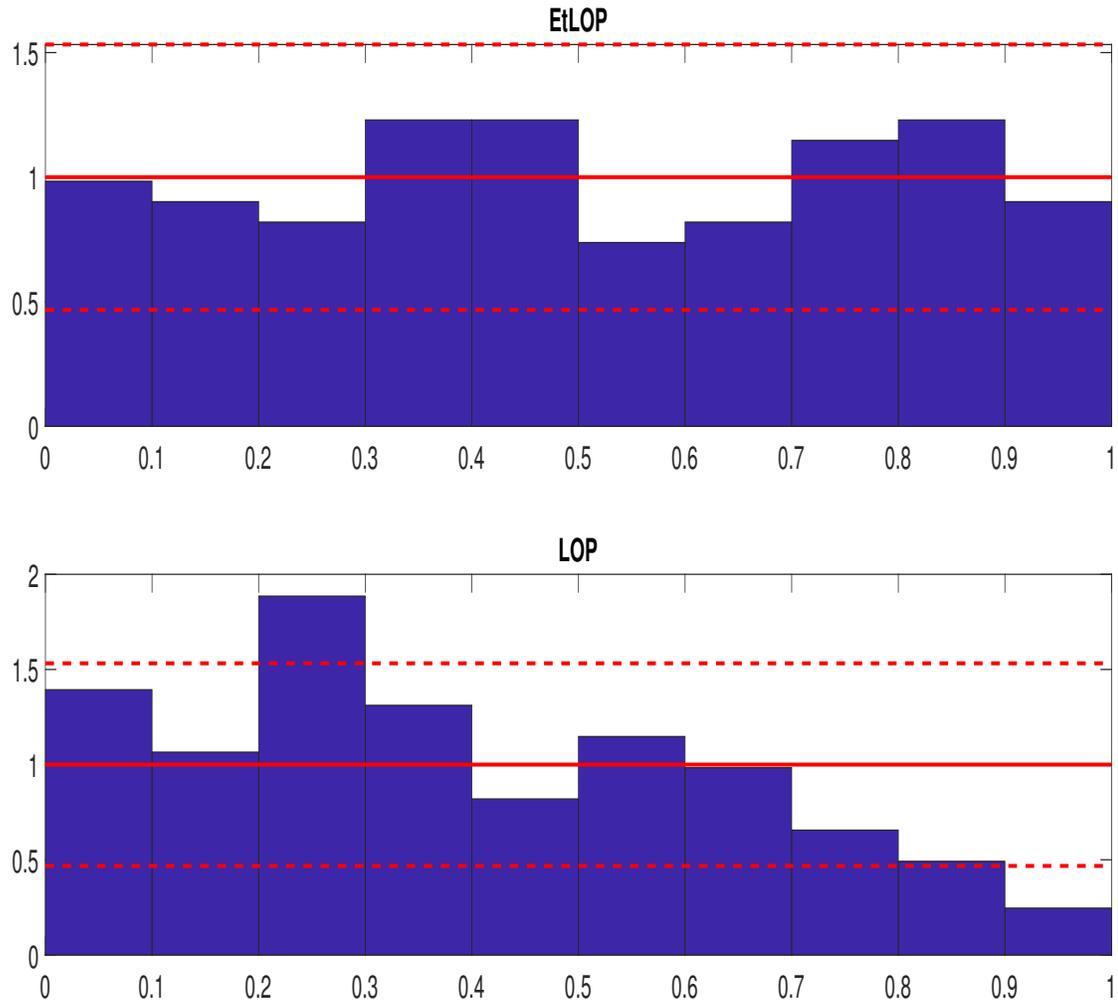


Figure 6: PITS, $h=1$, 1990:1 to 2020:2



Notes: The histograms depict the empirical distributions of the PITs. Solid lines represent the frequency of draws that are expected to be in each bin under a $U(0, 1)$ distribution. Dashed lines represent the 95% confidence intervals under the normal approximation of a binomial distribution.

Figure 7: EtLOP and LOP Forecasts, $h=1$, 1990:1 to 2020:2

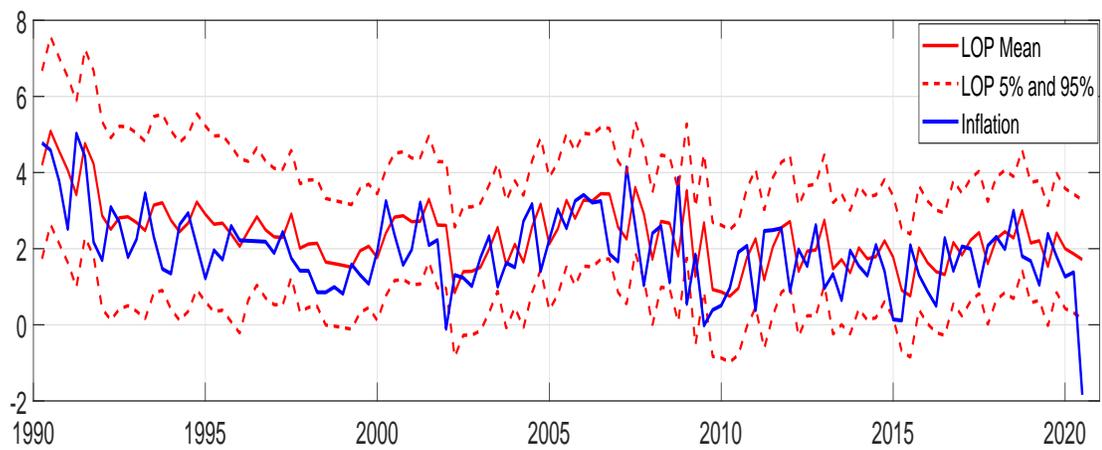
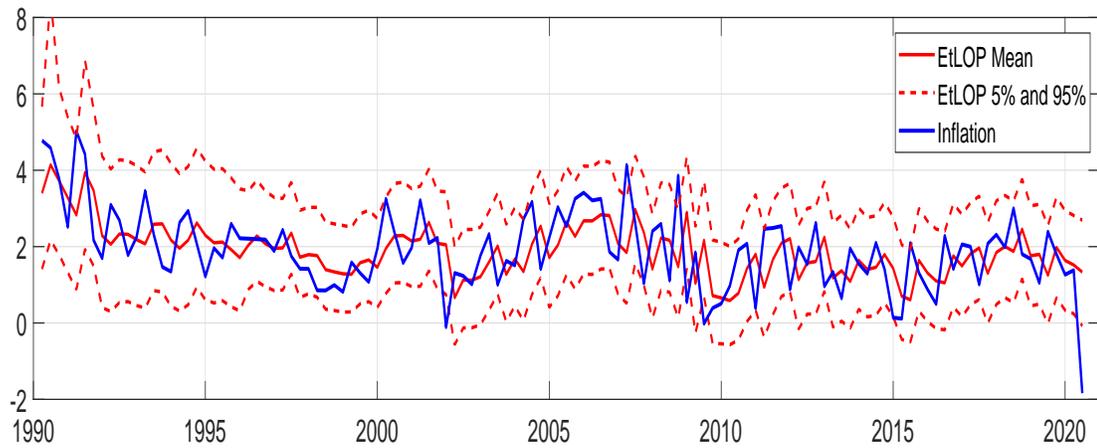


Figure 8: Uncertainty Range, Skew and Event Probability, $h=1$, 1990:1 to 2020:2

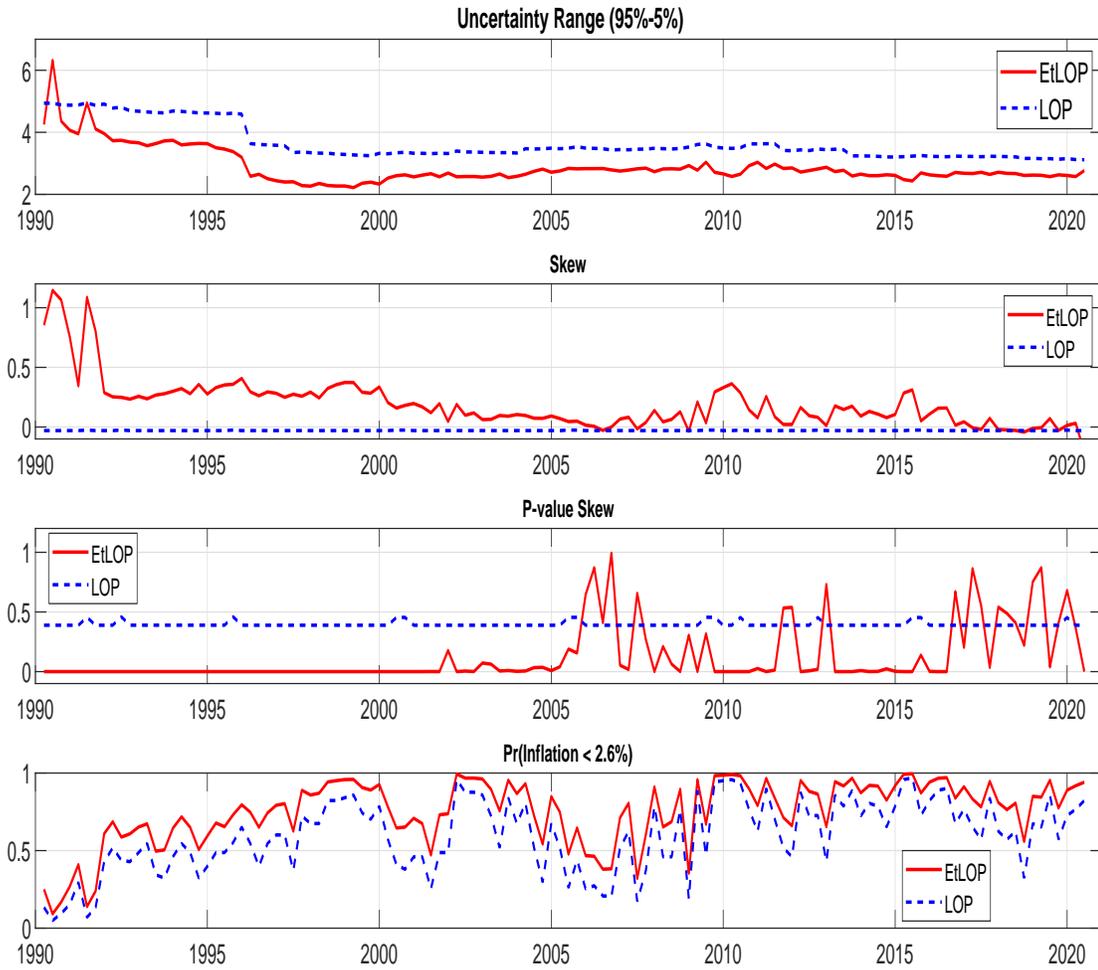


Figure 9: Forecast Densities for 2009, h=1

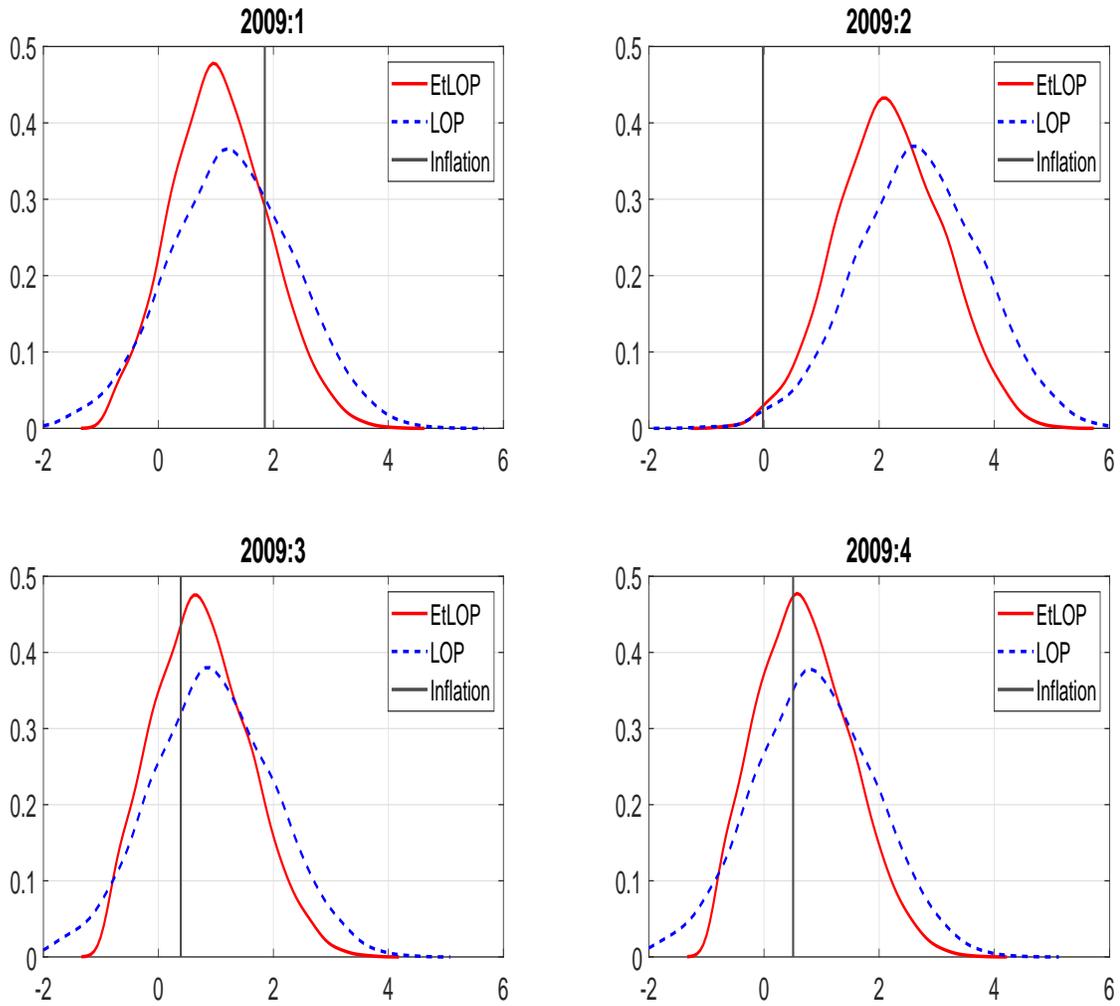


Table 1: Forecast Evaluation for EtLOP

Horizon	RMSFE	CRPS	TW	RTW	LTW
$h = 1$	0.867**	0.860**	0.850**	0.839**	0.877*
$h = 2$	0.802**	0.789**	0.774**	0.754**	0.828*
$h = 3$	0.734**	0.727**	0.726**	0.657**	0.814*
$h = 4$	0.660**	0.670**	0.696**	0.594**	0.751*

Notes: Columns 2 to 6 report ratios of EtLOP relative to LOP benchmark, for RMSFE, the CRPS and the tail-weighted (TW), right-tail weighted (RTW) and left-tail weighted (LTW) CRPS statistics, respectively. Ratios less than one indicate an improvement in forecast performance relative to the benchmark LOP. Improvements, using the two-sided test of Giacomini and White (2006), at the 1% and 5% significance levels are denoted ** and *, respectively.

References

- [1] Aastveit, K. A., J. Mitchell, F. Ravazzolo, and H.K. van Dijk (2019), “The Evolution of Forecast Density Combinations in Economics”, Oxford Research Encyclopedia of Economics and Finance, Oxford University Press.
- [2] Adrian, T., N. Boyarchenko and D. Giannone (2019), “Vulnerable Growth”, *American Economic Review*, 109, 4, 1263-1289.
- [3] Allayioti, A. (2020), “Bayesian Entropic Tilting for Macroeconomic Variables with Reshaped Survey Information”, mimeo, Warwick Business School, University of Warwick.
- [4] Amengual, D., E. Sentana and Z. Tian (2020), “Gaussian Rank Correlation and Regression”, CEPR Discussion Papers 14914, June.
- [5] Bassetti, F., R. Casarin and F. Ravazzolo (2018), “Bayesian Nonparametric Calibration and Combination of Predictive Distributions”, *Journal of the American Statistical Association*, 113, 552, 675-685.
- [6] Baxter, M, and R.G. King (1999), “Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series”, *Review of Economics and Statistics*, 81, 594-607.
- [7] Beveridge, S., and C.R. Nelson (1981), “A New Approach to Decomposition of Time Series into Permanent and Transitory Components with Particular Attention to Measurements of the Business Cycle”, *Journal of Monetary Economics*, 7, 151-174.
- [8] Chan, J.C.C., and Y. Song (2018), “Measuring Inflation Expectations Uncertainty Using High-Frequency Data”, *Journal of Money Credit and Banking*, 50, 1139-1166.

- [9] Carriero, A., T.E. Clark and M. Marcellino (2020) “Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions”, Federal Reserve Bank of Cleveland, Working Paper No. 20-02, January.
- [10] Christiano, L. and T.J. Fitzgerald (2003), “The Band Pass Filter”, *International Economic Review*, 44, 2, 435-465.
- [11] Clark, T.E. and M.W. McCracken (2010), “Averaging Forecasts from VARs with Uncertain Instabilities”, *Journal of Applied Econometrics*, January-February, 5-29.
- [12] Coe, P.J. and S.P. Vahey (2020), “Financial Conditions and the Risks to Economic Growth in the United States Since 1875”, CAMA Working Paper No. 36/2020, Australian National University, April.
- [13] Croushore, D. and T. Stark (2001), “A Real-time Data Set for Macroeconomists”, *Journal of Econometrics*, 105, 111-130.
- [14] DeGroot M.H. and J. Mortera (1991), “Optimal Linear Opinion Pools”, *Management Science*, 37, 5, 546-558.
- [15] Deheuvels, P. (1979), “La Fonction de Dépendance Empirique et Ses Propriétés. Un Test non Paramétrique d’Indépendance”, *Bulletin Royal Belge de l’Académie des Sciences*, 65, 274-292.
- [16] Deheuvels, P. (1981), “An Asymptotic Decomposition for Multivariate Distribution-free Tests of Independence”, *Journal of Multivariate Analysis*, 11, 1, 102-113.
- [17] Diebold, F.X., T.A. Gunther and A.S. Tay (1998), “Evaluating forecast densities; with Applications to Financial Risk Management”, *International Economic Review*, 39, 863-83.
- [18] Diebold, F.X. and R.S. Mariano (1995), “Comparing Predictive Accuracy”, *Journal of Business and Economic Statistics*, 13, 253-263.

- [19] Diebold, F.X. and M. Shin (2019) “Machine Learning for Regularised Survey Forecast Combination: Partially-Egalitarian Lasso and its Derivatives”, *International Journal of Forecasting*, 35, 4, 1679-1691.
- [20] Galbraith, J.W. and S. van Norden (2012), “Assessing Gross Domestic Product and Inflation Probability Forecasts Derived from Bank of England Fan Charts”, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 175, 3, 713-727.
- [21] Galvao, A.B., A. Garratt and J. Mitchell (2021), “Does Judgement Improve Macroeconomic forecast densities?”, *International Journal of Forecasting*, 37, 3, 1247-1260.
- [22] Ganics, G. (2017), “Optimal Density Forecast Combinations”, Bank of Spain Working Paper No. 1751.
- [23] Garratt, A., K. Lee, E. Mise and K. Shields (2008), “Real-Time Representations of the Output Gap”, *Review of Economics and Statistics*, 90, 4, 792-804.
- [24] Garratt, A., J. Mitchell, S.P. Vahey and E. Wakerly (2011), “Real-time Inflation Forecast Densities from Ensemble Phillips Curves”, *North American Journal of Economics and Finance*, 22, 77-87.
- [25] Garratt, A., J. Mitchell, and S.P. Vahey (2014), “Measuring Output Gap Nowcast Uncertainty”, *International Journal of Forecasting*, 30, 2, 268-279.
- [26] Garratt, A., and I. Petrella (2021), “Commodity Prices and Inflation Risk”, *Journal of Applied Econometrics*, forthcoming.
- [27] Giacomini, R. and H. White (2006), “Tests of Conditional Predictive Ability”, *Econometrica*, 74, 1545-1578.

- [28] Geweke, J. and G. Amisano (2011), “Optimal Prediction Pools”, *Journal of Econometrics*, 164, 130-141.
- [29] Gneiting, T. and A.E. Raftery (2007), “Strictly Proper Scoring Rules, Prediction, and Estimation”, *Journal of the American Statistical Association*, 102, 359-378.
- [30] Gneiting, T. and R. Ranjan (2011), “Comparing forecast densities Using Threshold- and Quantile-Weighted Scoring Rules”, *Journal of Business and Economic Statistics*, 29, 3, 411-422.
- [31] Gneiting, T. and R. Ranjan (2013), “Combining Predictive Distributions”, *Electronic Journal of Statistics*, 7, 7471782.
- [32] Harvey, D., S. Leybourne and P. Newbold (1997), “Testing the Equality of Prediction Mean Squared Errors”, *International Journal of Forecasting*, 13, 281-291.
- [33] Hodrick, R. and E. Prescott (1997), “Post-War U.S. Business Cycles: An Empirical Investigation”, *Journal of Money, Banking and Credit*, 29, 1-16.
- [34] Jore, A.S., J. Mitchell and S.P. Vahey (2010), “Combining Forecast Densities from VARs with Uncertain Instabilities”, *Journal of Applied Econometrics*, 25, 621-634.
- [35] Karagedikli, O., S.P. Vahey and E.C. Wakerly (2019), “Improved Methods for Combining Point Forecasts for an Asymmetrically Distributed Variable”, CAMA Working Paper No. 15/2019, Australian National University, February.
- [36] Kascha, C. and F. Ravazzolo (2010), “Combining Inflation Forecast Densities”, *Journal of Forecasting*, 29, 231-250.
- [37] Knüppel, M. and F., Krüger (2021), “Forecast Uncertainty, Disagreement, and the Linear Pool”, *Journal of Applied Econometrics*, forthcoming.

- [38] Loaiza-Maya, R. and M.S. Smith (2020), “Real-Time Macroeconomic Forecasting with a Heteroskedastic Inversion Copula”, *Journal of Business and Economic Statistics*, 38, 2, 470-486.
- [39] Mise, E., T-H. Kim and P. Newbold (2005), “On the Sub-Optimality of the Hodrick-Prescott Filter”, *Journal of Macroeconomics*, 27, 1, 53-67.
- [40] Odendahl, F. (2018), “Survey-Based Joint forecast densities”, mimeo, UPF, October.
- [41] Orphanides, A. and S. van Norden (2002), “The Unreliability of Output-Gap Estimates in Real Time”, *Review of Economics and Statistics*, 84, 4, 569-583.
- [42] Orphanides, A. and S. van Norden (2005), “The Reliability of Inflation Forecasts Based on Output-Gap Estimates in Real Time”, *Journal of Money Credit and Banking*, 37, 3, 583-601.
- [43] Ranjan, R. and T. Gneiting (2010), “Combining Probability Forecasts”, *Journal of the Royal Statistical Society Series B*, 72, 71-91.
- [44] Rosenblatt, M. (1952), “Remarks on a Multivariate Transformation”, *The Annals of Mathematical Statistics*, 23, 470-472.
- [45] Rossi, B. (2019), “Forecasting in the Presence of Instabilities: How Do We Know Whether Models Predict Well and How to Improve Them”, Barcelona GSE Working Paper Series Working Paper 1161, November.
- [46] Rossi, B. and T. Sekhposyan (2014), “Evaluating Predictive Densities of U.S. Output Growth and Inflation in a Large Macroeconomic Data Set”, *International Journal of Forecasting*, 30, 3, 662-682.
- [47] Rossi, B. and T. Sekhposyan (2019), “Alternative Tests for Correct Specification of Conditional Predictive Densities”, *Journal of Econometrics* 208, 2, 638-657.

- [48] Smith, M.S. and S.P. Vahey (2016), “Asymmetric Forecast Densities for U.S. Macroeconomic Variables from a Gaussian Copula Model of Cross-Sectional and Serial Dependence”, *Journal of Business and Economic Statistics*, 34, 416-434.
- [49] Shimazaki, H. and S. Shinomoto (2010), “Kernel Bandwidth Optimization in Spike Rate Estimation”, *Journal of Computational Neuroscience*, 29, 171-182.
- [50] Timmermann, A. (2006), “Forecast Combinations”, *Handbook of Economic Forecasting*, 135-196.
- [51] Velásquez-Giraldo, M., Canavire-Bacarreza, G., Huynh, K. and D. T. Jacho-Chavez (2018), “Flexible Estimation of Demand Systems: A Copula Approach”, *Journal of Applied Econometrics*, 33, 1109-1116.

Appendix 1: Output trend definitions

We summarise the seven univariate detrending specifications below.

1. For the quadratic trend based measure of the output gap we use the residuals from a regression (estimated recursively) of output on a constant and a squared time trend.
2. Following Hodrick and Prescott (1997, HP), we set the smoothing parameter to 1600 for our quarterly U.S. data.¹⁹
3. Since the HP filter is a two-sided filter it relates the time- t value of the trend to future and past observations. Moving towards the end of a finite sample of data, the HP filter becomes progressively one-sided and its properties deteriorate with the Mean Squared Error (MSE) of the unobserved components increasing and the estimates ceasing to be optimal in a MSE sense. We therefore follow Mise et al. (2005) and mitigate this loss near and at the end of the sample by extending the series with forecasts. At each recursion the HP filter is applied to a forecast-augmented output series (again with smoothing parameter 1600), with forecasts generated from an univariate AR(8) model in output growth (again estimated recursively using the appropriate vintage of data). The implementation of forecast augmentation when constructing real-time output gap measures for the U.S. is discussed at length in Garratt et al. (2008). We deliberately select a high lag order to ensure no important lags are omitted—favouring unbiasedness over efficiency.
4. Christiano and Fitzgerald (2003) propose an optimal finite-sample approximation to the band-pass filter, without explicit modelling of the data. Their approach implicitly assumes that the series is captured reasonably well by a random walk model and that, if there is drift present, this can be proxied by the average growth rate over the sample.

¹⁹We could, of course, allow for uncertainty in the smoothing parameter. We reduce the computational burden in this application by fixing this parameter at 1600.

5. We also consider the band-pass filter suggested by Baxter and King (1999). We define the cyclical component to be fluctuations lasting no fewer than six, and no more than thirty-two quarters—the business cycle frequencies indicated by Baxter and King (1999)—and set the truncation parameter (the maximum lag length) at three years. As with the HP filter we augment our sample with AR(8) forecasts to fill in the ‘lost’ output gap observations at the end of the sample due to truncation.
6. The Beveridge and Nelson (1981) decomposition relies on a priori assumptions about the correlation between permanent and transitory innovations. The approach imposes the restriction that shocks to the transitory component and shocks to the stochastic permanent component have a unit correlation. We assume the ARIMA process for output growth is an AR(8), the same as that used in our forecast augmentation.
7. Finally, our Unobserved Components model assumes q_t is decomposed into trend, cyclical and irregular components

$$q_t = \mu_t^7 + y_t^7 + \xi_t, \quad \xi_t \sim i.i.d. N(0, \sigma_\xi^2), \quad t = 1, \dots, T \quad (\text{A1.1})$$

where the stochastic trend is specified as

$$\begin{aligned} \mu_t^7 &= \mu_{t-1}^7 + \beta_{t-1} + \eta_t, \quad \eta_t \sim i.i.d. N(0, \sigma_\eta^2) \\ \beta_t &= \beta_{t-1} + \zeta_t, \quad \zeta_t \sim i.i.d. N(0, \sigma_\zeta^2). \end{aligned}$$

Letting $\sigma_\zeta^2 > 0$ but setting $\sigma_\eta^2 = 0$, gives an integrated random walk. The cyclical component

is assumed to follow a stochastic trigonometric process:

$$\begin{bmatrix} y_t^7 \\ y_t^{7*} \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} y_{t-1}^7 \\ y_{t-1}^{7*} \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \quad (\text{A1.2})$$

where λ is the frequency in radians, ρ is a damping factor and κ_t and κ_t^* are two independent white noise Gaussian disturbances with common variance σ_κ^2 . We estimate this model by maximum likelihood, exploiting the Kalman filter, and estimates of the trend and cyclical components are obtained using the Kalman smoother.