Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Empirically-transformed Linear Opinion Pools

CAMA Working Paper 47/2019
July 2019

Anthony Garratt
University of Warwick

Timo Henckel
Centre for Applied Macroeconomic Analysis, ANU

Shaun P. Vahey
University of Warwick
Centre for Applied Macroeconomic Analysis, ANU

Abstract

Many studies have found that combining density forecasts improves predictive accuracy for macroeconomic variables. A prevalent approach known as the Linear Opinion Pool (LOP) combines forecast densities from "experts"; see, among others, Stone (1961), Geweke and Amisano (2011), Kascha and Ravazzolo (2011), Ranjan and Gneiting (2010) and Gneiting and Ranjan (2013). Since the LOP approach averages the experts' probabilistic assessments, the distribution of the combination generally differs from the marginal distributions of the experts. As a result, the LOP combination forecasts sometimes fail to match salient features of the sample data, including asymmetries in risk. In this paper, we propose a computationally convenient transformation for a target macroeconomic variable with an asymmetric marginal distribution. Our methodology involves a Smirnov transform to reshape the LOP combination forecasts using a nonparametric kernel-smoothed empirical cumulative distribution function. We illustrate our methodology with an application examining quarterly real-time forecasts for US inflation based on multiple output gap measures over an evaluation sample from 1990:1 to 2017:2. Our proposed methodology improves combination forecast performance by approximately 10% in terms of both the root mean squared forecast error and the continuous ranked probability score. We find that our methodology delivers a similar performance gain for the Logarithmic Opinion Pool (LogOP), a commonly-used alternative to the LOP.

Keywords

Forecast density combination, Smirnov transform, inflation

JEL Classification

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

Empirically-transformed Linear Opinion Pools*

Anthony Garratt (University of Warwick)
Timo Henckel (ANU and CAMA)
Shaun P. Vahey (University of Warwick and CAMA, ANU)

June 17, 2019

Abstract

Many studies have found that combining density forecasts improves predictive accuracy for macroeconomic variables. A prevalent approach known as the Linear Opinion Pool (LOP) combines forecast densities from "experts"; see, among others, Stone (1961), Geweke and Amisano (2011), Kascha and Ravazzolo (2011), Ranjan and Gneiting (2010) and Gneiting and Ranjan (2013). Since the LOP approach averages the experts' probabilistic assessments, the distribution of the combination generally differs from the marginal distributions of the experts. As a result, the LOP combination forecasts sometimes fail to match salient features of the sample data, including asymmetries in risk. In this paper, we propose a computationally convenient transformation for a target macroeconomic variable with an asymmetric marginal distribution. Our methodology involves a Smirnov transform to reshape the LOP combination forecasts using a nonparametric kernel-smoothed empirical cumulative distribution function. We illustrate our methodology with an application examining quarterly real-time forecasts for US inflation based on multiple output gap measures over an evaluation sample from 1990:1 to 2017:2. Our proposed methodology improves combination forecast performance by approximately 10% in terms of both the root mean squared forecast error and the continuous ranked probability score. We find that our methodology delivers a similar performance gain for the Logarithmic Opinion Pool (LogOP), a commonly-used alternative to the LOP.

Keywords: Forecast density combination; Smirnov transform; inflation.

*We thank Liz Wakerly and Yunyi Zhang for helpful comments and suggestions.

1 Introduction

Monetary policymakers often warn the public about asymmetric economic risks to inflation and other macroeconomic variables. For example, the Federal Open Market Committee, like most central banks, discusses these risks explicitly. Some central banks, including the Bank of England and Norges Bank, publish asymmetric predictive densities for key variables. Among others, Smith and Vahey (2016) and Adrian et al. (2019) examine methods to quantify risks to macroeconomic variables, drawing explicit attention to the asymmetric marginal distributions of some US macroeconomic variables.

A large and growing literature has focused on the forecast performance of density combinations with variants of the Linear Opinion Pool (LOP) for macroeconomic variables. A useful feature of the framework is that the combinations require only density forecasts from the experts (together with the realisations of the target variable)—experts' predictions are treated as opinions to be pooled. The approach facilitates combination of density forecasts from a wide range of sources, including survey respondents and forecasting models. Jore et al. (2010), Geweke and Amisano (2011), Kascha and Ravazzolo (2011), Ranjan and Gneiting (2010), Gneiting and Ranjan (2013), Waggoner and Zha (2012), Billio et al. (2013), Del Negro, Hasegawa and Schorfheide (2016) and Bassetti et al. (2019), among others, provide recent applications and generalisations of the LOP. Hall and Mitchell (2007), Geweke and Amisano (2011), Conflitti, De Mol and Giannone (2015), Kapetanios et al. (2015) and Pettenuzzo and Ravazzolo (2016) discuss various schemes to optimise weights for density combinations.

The LOP approach averages the density forecasts from experts in a manner that preserves expert disagreement about probabilities. Unfortunately, the distribution of the combination density forecast generally differs from the marginal distributions of the experts' forecasts. In several nowcasting systems, such as the System for Averaging Models

(SAM) developed at Norges Bank, the forecasts for inflation (and other variables) being combined are based on linear-Gaussian (or approximately Gaussian) models. In this setting, if the target variable has an asymmetric distribution, the LOP combination forecasts typically fail to match features of the sample data. Arguably, most macroeconomic variables including inflation are distributed asymmetrically.

In this paper, we propose a modification to the standard LOP approach, which overcomes this limitation by explicitly considering combinations where inflation has an asymmetric marginal distribution. Our methodology involves the decision maker applying a Smirnov transform to reshape the LOP combination forecasts. Our proposed Empirically-transformed Linear Opinion Pool (EtLOP) involves fitting a non-parametric Empirical Cumulative Distribution Function (ECDF) to provide the marginal distribution for the target variable. The decision maker uses a (modified) Smirnov transform to reshape the combined density using the inverse of the ECDF. In doing so, the decision maker deploys an algorithm adapted from the pseudo-random number generation methodology sometimes referred to as inverse transform sampling.

An earlier literature on opinion (or predictive) pooling methods in decision making stressed that the LOP approach does not satisfy various axioms usually associated with individual and group rationality of the experts; see, for example, Genest and Zidek (1986). Ranjan and Gneiting (2010) and Gneiting and Ranjan (2013) propose a solution to the tendency of LOP to give over-dispersed aggregate combination forecast densities via a beta transform. Bassetti et al. (2019) extend the approach using Bayesian methods to estimate a mixture of betas. Ganics (2017) proposes re-weighting candidate forecast densities using a measure of calibration.

In contrast, applying our EtLOP approach involves transforming the LOP forecast density using a *non-parametric* marginal distribution fitted to inflation (or another macroeconomic target variable). Recent papers utilising ECDFs to fit margins include, among others, the copula modelling papers by Smith and Vahey (2016), Loaiza-Maya and Smith (2018) and Amengual et al. (2019). In contrast to these earlier studies, however, we do not estimate the dependence parameters. Instead, we propose a method to transform the predictions from an opinion pool to match the marginal distribution of the sample data, without estimating the dependence between experts. Our computationally convenient approach suits either frequentist or Bayesian analysis, where the number of forecasts being combined is large relative to the length of the time series available for the target variable, which often makes the estimation of dependence parameters troublesome in practice.

We illustrate our methodology with an application examining real-time density forecasts for US inflation based on a large number of candidate misspecified models (experts) in the tradition of for example Jore et al. (2010), Garratt et al. (2014) and Rossi and Sekhposyan (2014). To keep things simple but relevant for monetary policymakers, we utilise an expert space similar to Garratt et al. (2011), where each expert uses a bivariate linear-Gaussian time series model for inflation and the output gap to produce "real-time" h-step ahead forecasts for inflation. The decision maker combines the forecasts from the experts using LOP and EtLOP and the approaches are compared using an evaluation sample from 1990:1 to 2017:2. Relative to the more conventional LOP, our approach improves forecasting performance by approximately 10% in terms of both point and density forecasting metrics, namely, the root mean squared forecast error (RMSFE) and the continuous ranked probability score (CRPS).

Since geometric averaging of probabilities via the Logarithmic Opinion Pool (LogOP) is often competitive with the arithmetic averaging approach (LOP) in terms of out-of-sample predictive performance in macroeconomic applications, we replicate our analysis with LogOP and find a similar performance gain from the Empirically-transformed Logarithmic Opinion Pool (EtLogOP).

The remainder of this paper is structured as follows. In Section 2, we set out our methodology for the empirically-transformed opinion pools and contrast with copula modelling approaches. In Section 3, we apply our methodology to forecast densities for US inflation. We present our results in Section 4, and in the final section, we draw some conclusions.

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 the conventional opinion pooling methodology, contrast it with our own approach and then discuss some practical considerations.

2.1 Conventional Opinion Pooling

We begin by describing the LOP approach and contrasting it with LogOP; see Kascha and Ravazzolo (2010) for further discussion of these two types of opinion pool. The decision maker utilises the out-of-sample forecasts for inflation from the many experts (models) in the combined forecast density:

$$p^{LOP}(\pi_\tau) = \sum_{j=1}^{J} w_{j,\tau}\ g(\pi_\tau \mid I_{j,\tau}), \qquad \tau = \underline{\tau}, \dots, \overline{\tau}, \tag{1}$$

where $g(\pi_\tau \mid I_{j,\tau})$ are the one step ahead forecast densities from expert (model) 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 with each recursion in the evaluation period $\tau = \underline{\tau}, \dots, \overline{\tau}$; see the discussion in, for example, Garratt et al. (2014).

In contrast, the decision maker's combined density defined by LogOP is:

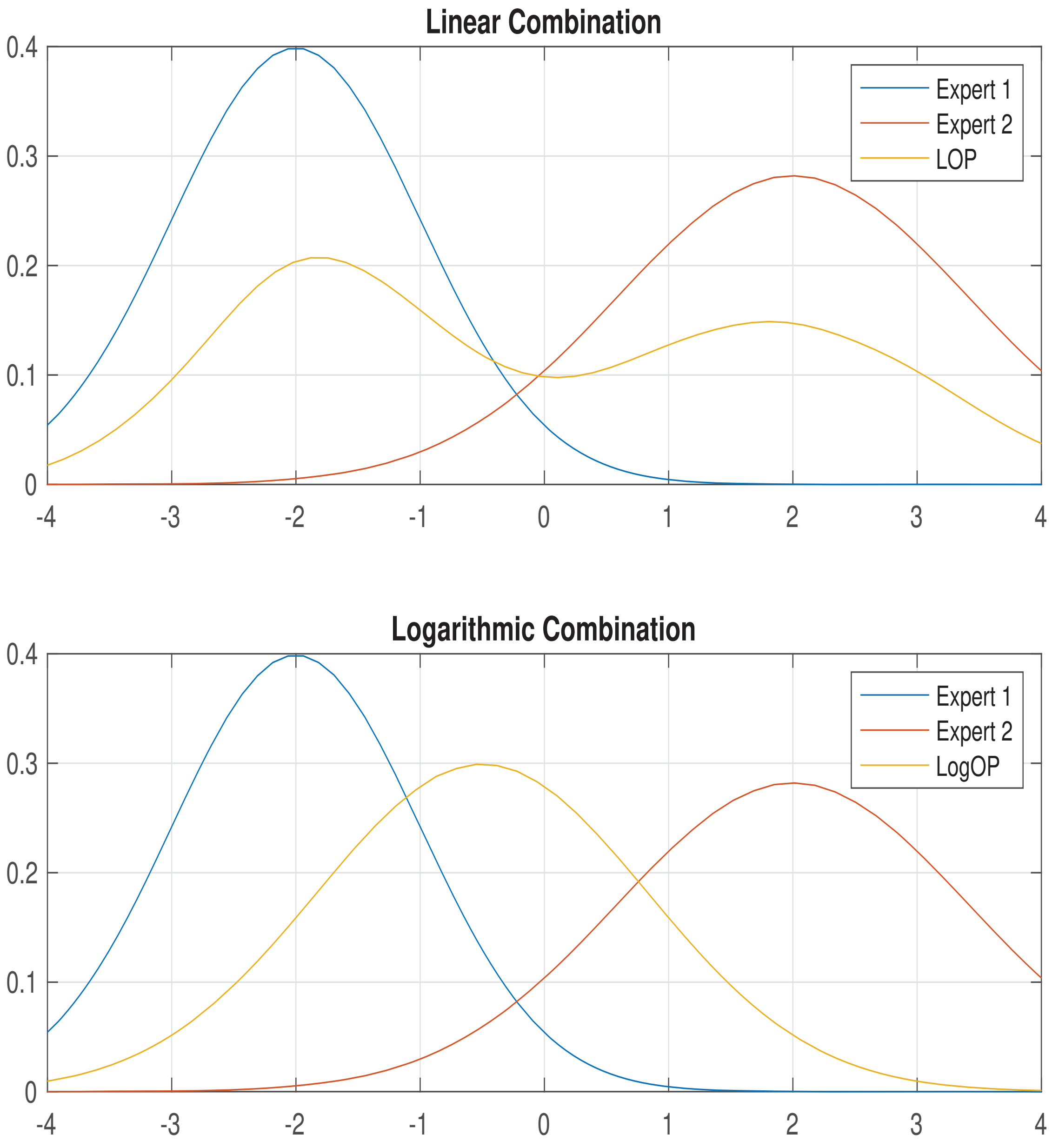
$$p^{LogOP}(\pi_\tau) = \frac{\prod_{j=1}^{J} g(\pi_\tau \mid I_{j,\tau})^{w_{j,\tau}}}{\int \prod_{j=1}^{J} g(\pi_\tau \mid I_{j,\tau})^{w_{j,\tau}} d\pi_\tau}, \quad \tau = \underline{\tau}, \ldots, \overline{\tau}, \tag{2}$$

where the denominator is a constant that ensures that the combination density is proper. The LogOP is linear in its logarithmic form.

Following Kascha and Ravazzolo (2010), to illustrate the idea behind our approach, we consider a simple combination by the decision maker for a single target observation based on the predictive densities supplied by two Gaussian experts. Both panels of Figure 1 plot the experts' forecasts, where the prediction of Expert 1 has mean -2.0 and standard deviation 1 and that of Expert 2 has mean 2.0 and standard deviation 2.0. The top panel also plots the LOP forecast (assuming equal weights) which is bimodal, with a slightly higher peak associated with the forecast mean of Expert 1. The bottom panel plots the LogOP forecast, which is unimodal, but where the central mass sits between the twin peaks of the LOP (shown in the top panel).

This static example illustrates several relevant features of conventional opinion pooling with Gaussian experts. First, in the case of LOP, although the experts' forecasts are individually Gaussian, the combined LOP density does not necessarily match the functional form of the forecasts from the individual experts. Second, the LOP often puts more mass in the tails than LogOP. Third, the LOP tends to preserve disagreement about the central probability mass of the experts, whereas LogOP does not. Regardless of the type of opinion pool, in this example the aggregated forecast does not inherit all the properties of the experts' forecasts being combined.

Figure 1: Conventional Combinations and Experts



2.2 Empirically-transformed Pooling

Our methodology involves non-parametric fitting of the marginal distribution for the target variable. We exploit the separation of the marginal distributions from dependence familiar from copula modelling strategies.

Suppose there is a multivariate time series, $\boldsymbol{Z}$, comprising the target variable, inflation, and the dependent macroeconomic variables.[1] For the K macroeconomic time series variables, Sklar's Theorem (Nelsen, 2006) suggests there exists a copula function C such that the joint distribution function can be written:

$$F(\boldsymbol{z}) = C(\boldsymbol{u}) \tag{3}$$

where $\boldsymbol{z} = (\boldsymbol{z}'_1, \ldots, \boldsymbol{z}'_T)'$, $\boldsymbol{u} = (\boldsymbol{u}'_1, \ldots, \boldsymbol{u}'_T)'$, $\boldsymbol{z}_t = (z'_{1,t}, \ldots, z'_{K,t})'$, $\boldsymbol{u}_t = (u'_{1,t}, \ldots, u'_{K,t})'$, $u_{k,t} = F_{z_k}(z_{k,t})$ with $k = 1, \ldots, K$.

The variables are assumed to be stationary and the margins time invariant. A copula function C is a distribution function on the unit hypercube $[0,1]^K$, where all margins are strictly uniform. Dependence between elements of $\boldsymbol{Z}$ is captured by the copula function.

Differentiate the distribution function to give the density of $\boldsymbol{Z}$:

$$f(\boldsymbol{z}) = c(\boldsymbol{u}) \prod_{t=1}^{T} \prod_{k=1}^{K} f_{z_k}(z_{k,t}) \qquad k = j + 1 \tag{4}$$

where $f_{z_k}(z_{k,t})$ is the marginal density of $z_{k,t}$, $F_{z_k}(z_{k,t})$ is the corresponding distribution function, and $c(\boldsymbol{u}) = \frac{\partial}{\partial \boldsymbol{u}} C(\mathbf{u})$ is the copula density, where $\mathbf{u} = (u_1, \ldots, u_K)'$.

Sklar's Theorem indicates that a copula density capturing the dependence between the K variables exists but it is unknown. Assuming a Gaussian copula, for example, would correspond to linear dependence.[2]

[1] In our subsequent application the experts view the output gaps as misspecified alternatives, each working with a unique bivariate VAR specification to produce forecasts.

[2] Smith and Vahey (2016) discuss the relationship between copulas and Vector Autoregressions.

In our density combination setting for a single target variable, inflation, each expert produces forecasts from a unique model of inflation and the macroeconomic variables, where there are J experts (models) in total. The decision maker aggregates the experts' predictions for inflation but, by assumption, does not estimate the underlying relationship between the K macroeconomic variables. In macroeconomic forecasting applications, the density combination approach has been found to be robust in the presence of model misspecifications.

In our applied work, each expert is defined by a unique forecasting model using a single candidate output gap measure and a particular lag structure. We assume that each expert estimates a linear dependence structure assuming Gaussian margins. Under LOP, the decision maker combines the density forecasts produced by the experts, but does not estimate the dependence structure between the target variable, inflation, and the forecasts from the experts.[3] The resulting LOP combination has an unknown marginal distribution, as illustrated in Figure 1, that need not (and typically will not) match the marginal distribution of the target macroeconomic variable.

We propose transforming the LOP combination forecast using the Empirical Cumulative Distribution Function (ECDF) for inflation. In effect, we use the ECDF to reshape the conventional combination distribution to match the marginal distribution of the sample data. The ECDF is the distribution function for the given sample—a step function that represents the entire history of the outturns.

For expositional ease, in what follows we discuss the approach for a single forecast origin and a single forecast horizon. (Generalisations to multiple horizons and many experts are straightforward.) In our application described in the subsequent section, we deploy the algorithm recursively to mimic real-time forecasting by the decision maker, fitting marginal

[3]The standard LOP approach to weighting forecast densities assumes that experts' information sets are conditionally independent; see, for example, the discussion in DeGroot and Mortera (1991).

distributions on vintages of real-time data to match the approach commonly-utilised by forecasters in practice.

Our algorithm involves the following four steps:

1. Fit an ECDF to the target time series variable, π_t, fitting a non-parametric (time invariant) distribution to the margin, $F(\pi_t)$.

2. Construct a (time-invariant) pseudo marginal distribution for the conventional LOP combination of experts. Fitting an ECDF to the entire history of forecast densities from the experts provides a proxy for the (typically unknown) distribution, $\phi(\pi_t)$.

3. Draw from the LOP combination forecast density and construct the quantile for each draw using $\phi(\cdot)$. In effect, the swarm of forecasts has then been mapped onto the unit interval (the forecast swarm has the scale of a probability).

4. Map the unit interval forecast swarm onto the observable scale using the the fitted marginal distribution, exploiting the inverse ECDF, $F^{-1}(\cdot)$, the quantile function.

Returning to our static example from the previous subsection, we highlight the advantages of empirically-transforming the experts' forecast densities. Figure 2 plots the same individual expert forecast densities as in Figure 1 but instead of showing the linear combination (upper panel) and logarithmic combination (lower panel), these now include the empirically-transformed linear and logarithmic combinations, respectively.

The improvements should be immediately visible. In the case of the linear combination (upper panel), the combined LOP density preserves the uni-modality of the individual (Gaussian) densities, with the peak closer to the forecast mean of Expert 1, with a long right tail. This contrasts with the conventional LOP combination shown in Figure 1 (top panel), which is bimodal. Turning to the empirically-transformed LogOP shown in Figure 2 (lower panel), the combination forecast density is again single peaked. Relative to the EtLOP case (upper panel), the empirically-transformed LogOP, EtLogOP, is more peaked. The EtLogOP modal forecast differs little from EtLOP, with a similar long right tail. Comparing the EtLOP combination forecast to the conventional LogOP displayed Figure 1 (lower panel), the coventional forecast density is more diffuse and symmetric.

2.3 Discussion

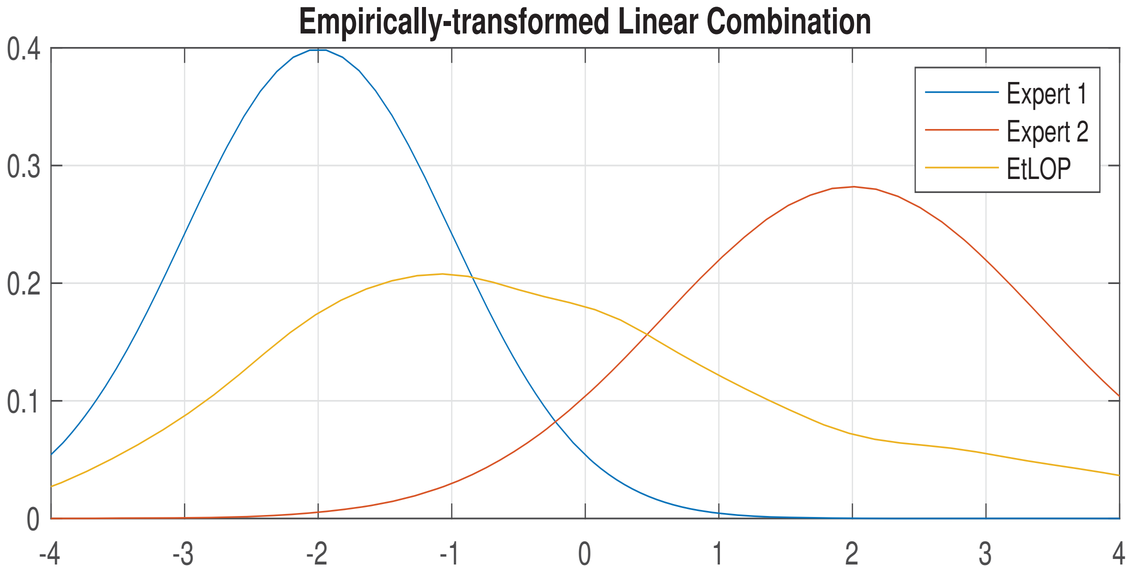
Our algorithm is based on the idea that the (time-invariant) marginal distributions, $F(\cdot)$ and $\phi(\cdot)$, are well calibrated. Using non-parametric ECDFs provides a computationally convenient and pragmatic strategy to fit the relevant margins. Given the well-calibrated marginals, our approach converts the swarm from the conventional forecast combination to the scale of Probability Integral Transforms, PITs, preserving the rank order of the draws within the swarm, then transforms the unit interval swarm to reflect the marginal distribution of the sample data. Among others, Rosenblatt (1952), Diebold, Gunther and Tay (1998), Galbraith and van Norden (2012) and Rossi and Sekhposyan (2019) discuss PITs properties.

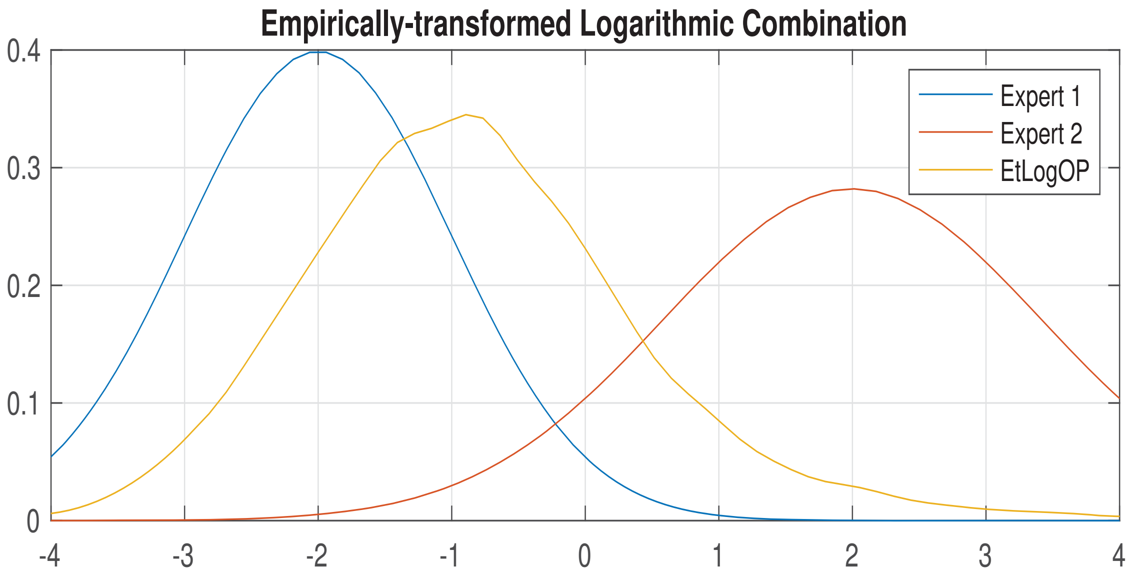
In practice, we use the kernel smoother `ksdensity` from MATLAB to compute the ECDFs, although the selection of the bandwidth is somewhat arbitrary. To deal with this, when fitting $\phi(\cdot)$, we treat the bandwidth as an unknown parameter and in our application below check the robustness over a specified interval. (There exist alternative means of selecting bandwidth, which typically rely on the minimisation of a risk function.)

3 Application: Forecasting US Inflation

To demonstrate the predictive effectiveness of our approach, we apply the algorithm to conventional density forecast combinations for quarterly US inflation using the target dates from 1990:1 to 2017:2 using a modelling space similar to Garratt et al. (2011).

Figure 2: Empirically-transformed Combinations and Experts





3.1 Expert Space, Forecasts and Data Considerations

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, allowing for simultaneity. The j^{th} VAR model takes the form:

$$\begin{bmatrix} \pi_t \\ y_t^j \end{bmatrix} = \begin{bmatrix} a_{\pi\pi}^j & a_{\pi y}^j \\ a_{y\pi}^j & a_{yy}^j \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ y_{t-1}^j \end{bmatrix} + \begin{bmatrix} \epsilon_{\pi t}^j \\ \epsilon_{yt}^j \end{bmatrix}, \qquad t = 1, \ldots, T, \tag{5}$$

where $[\epsilon_{\pi t}^j, \epsilon_{yt}^j]' \sim i.i.d.\ N(\mathbf{0}, \mathbf{\Sigma}^j)$. That is, we consider a baseline VAR specification in which the measure of interest, the output gap, has been varied to give J linear and Gaussian VAR models, indexed $j = 1, \ldots, J$. For expositional ease, we ignore the intercept and restrict the lag order of the J VARs to one.

Following Garratt et al. (2011), our baseline VAR setup uses seven output gap measures derived from the set of univariate off-model filters deployed by Orphanides and van Norden (2002, 2005). Federal Reserve researchers Edge and Rudd (2012) deploy similar univariate detrending methods.

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) actual output in t as q_t, and let μ_t^j be its trend using definition j, where $j = 1, \ldots, J$. 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:

$$q_t = \mu_t^j + y_t^j. \tag{6}$$

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

In our application, we vary a single auxiliary assumption to generate the model space. Specifically, we vary the lag length in the VAR (which for ease of exposition we fixed at one in equation (5)). 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 ensemble system. In our application, we restrict L to a maximum of four and therefore we consider 7×4 models—28 forecasts from the experts to be combined. Therefore, there are 29 margins for the system, 28 experts' forecasts plus the target variable, inflation.

The motivation for deploying these models stems from their common usage by central banks around the world. Nevertheless, as Orphanides and van Norden (2005) emphasise, single VAR models of this type often perform little better than simple univariate autoregressive benchmarks in real-time out-of-sample evaluations of point forecasting accuracy. In contrast, Garratt et al. (2011) note that using an ensemble of VAR models delivers probabilistically well-calibrated forecast densities for inflation in the US and Australia.

3.2 Data

Orphanides and van Norden (2002, 2005) stress that our 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) US dataset has 112 vintages, with the first vintage (dated 1990:1) used for parameter estimation containing time series observations from 1970:1 to 1989:4, and the last vintage used for parameter estimation (dated 2017:2) spanning from 1970:1 to 2017:1. We use the second release of the observed data to evaluate the forecasts. For example, when evaluating the $h = 1$ forecast (nowcast) for 1990:1, we use the 1990:3 vintage observation of 1990:1. US GDP deflator data are released with a one quarter lag.

The raw data for GDP (in practice, Gross National 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.

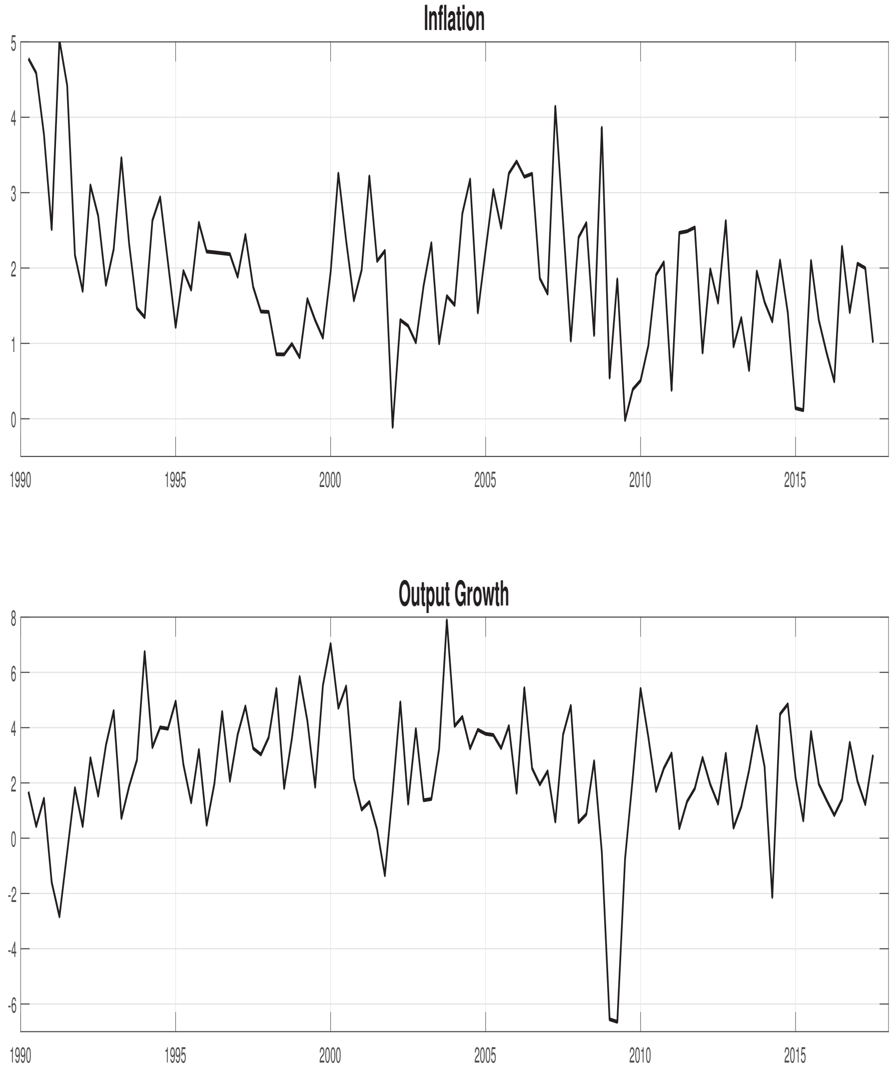
Figure 3 displays inflation (upper panel) and real output growth (lower panel) for our evaluation period, from 1990:1 to 2017:2 based on the final vintage of data. (The empirical analysis which follows uses many vintages.)

For the observations displayed prior to the Great Recession—the Great Moderation—inflation exhibits lower volatility and a higher unconditional mean. A striking feature of the Great Recession and its aftermath is the increased threat of low inflation.

In contrast, during the run up to the Great Recession, between 2003 and 2006, there are several realisations of high inflation. The upward spikes apparent in various inflationary measures for the US during this period are often regarded as (the response to) relative price movements, and, in particular, energy and food costs. See, for example, the discussion in Yellen (2006) and Ravazzolo and Vahey (2014).[4]

[4] Yellen (2006) is downloadable from http://www.frbsf.org/news/speeches/2006/0315.html. Clark and

Figure 3: US Inflation and Real Output Growth



Terry (2010) discuss the absence of pass through from energy prices to other prices during the period.

3.3 Forecast Combination and Empirical Transformation

The decision maker recursively combines the forecasts 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 2017:1 (window size 189 observations).

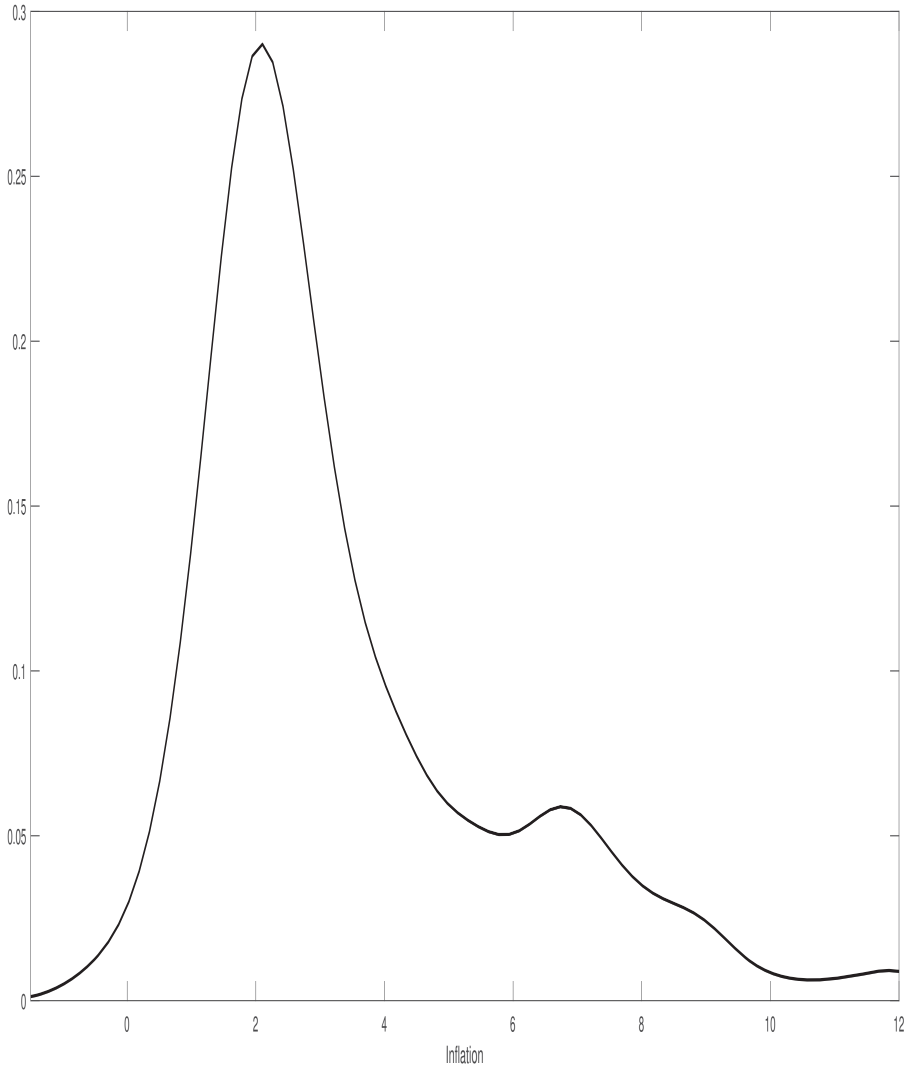
To deploy our algorithm for the EtLOP combination, the decision maker must fit a marginal distribution for inflation. The (smoothed) probability density function corresponding to the fitted ECDF plotted in Figure 4 uses the final vintage of inflation data. The marginal distribution displays some asymmetry, with the right tail extending relatively far from the central mass. Since the decision maker uses recursive fitting based on expanding windows of data, the fitted marginals for inflation vary slightly by forecast origin in practice. The null hypothesis of normality is rejected at the 1% significance level for the full sample of inflation data for all vintages.

It is straightforward to produce forecast densities for both inflation and the (j^{th}) output gap through our evaluation period: $\tau = \underline{\tau}, \ldots, \overline{\tau}$ where $\underline{\tau}$ =1990:1 and $\overline{\tau}$ =2017:2 (110 quarterly observations). Our experts focus on forecasting inflation, given the output gap data as well as lagged inflation data. Recall that we treat the "true" output gap as latent. Following Clark and McCracken (2010) and others, the experts (and the decision maker) use the second estimate as the target "final" data. For consistency, we report results for the same definition of final data for all forecast evaluations.

For each expert, we estimate the parameters using Ordinary Least Squares over an expanding window.

Our decision maker combines the out-of-sample forecasts from the VAR models using the LOP and EtLOP. In the following section we compare the forecasting performances of the various combinations and show the superiority of the empirically transformed opinion pools. The performance metrics gauge point forecast accuracy, using the Root Mean

Figure 4: ECDF Inflation



Squared Forecast Error (RMSFE) and density forecast accuracy, using the average Continuous Ranked Probability Score (CRPS) proposed by Gneiting and Raftery (2007).

4 Results

In this section, we report results for the one-step ahead forecasts from the forecast origins. Appendix 2 provides results for the four-step ahead case which display quantitively similar forecast performance. All results reported refer to equal weighted LOP (and EtLOP) combinations. (Recursively weighted combinations based on the time-averaged logarithmic score or the time-averaged CRPS weights give similar results.)

4.1 Bandwidth Selection

Table 1a presents out-of-sample forecast metrics for EtLOP for a variety of values of the bandwidth for $F(\cdot)$, denoted bw. Table 1b provides similar information for EtLogOP. The first column reports how the Brier score (Brier, 1950) varies with bw, for the event defined as inflation below its unconditional mean:

$$BS = (1/M)\sum_{\tau=\underline{\tau}}^{\overline{\tau}}(o_\tau - f_\tau)^2, \tag{7}$$

where M is the number of periods in the out-of-sample evaluation, $\overline{\tau}$ minus $\underline{\tau}$, o_τ is the outcome (1 if inflation below mean, 0 otherwise), and f_τ is the real-time probability, generated by EtLOP (EtLogOP), that inflation is below its mean. Low Brier scores are preferred. The second column reports the Root Mean Squared Forecast Error (RMSFE) as a ratio to the benchmark LOP (LogOP). The third column displays the (time-averaged) logarithmic score and the fourth column the (time-averaged) CRPS. The fifth column reports p-values for the Knüppel (2015) test of calibration, based on the PITs, that allows for autocorrelation and is flexible in the choice of raw moments. We employ the first two raw moments for the test statistic.

The first four columns of Table 1a suggest a relatively flat response over bw for the Brier score, RMSFE, the logarithmic score and the CRPS for the EtLOP. The fifth column indicates that the forecast densities are well calibrated typically, but here there is some sensitivity to the bw value.

Turning to Table 1b, the EtLogOP generally shares the characteristics exhibited by the EtLOP (described in Table 1a). Given the fairly robust performance across bandwidths, we present results below for a representative case, where $bw = 0.975$, with the corresponding rows in Table 1 in bold for emphasis.

4.2 Comparing Density Combinations

The first four rows of Table 2 display the main results of the paper, with the principle concern being the contrast between EtLOP and the benchmark LOP, with EtLogOP and LogOP provided for context. The EtLOP (EtLogOP) results refer to the $bw = 0.975$ case. We also provide results in the fifth row for an univariate unobserved components stochastic volatility (UCSV) model for inflation in the last row of Table 2, $\Delta\pi_t = \alpha + \varepsilon_t + \theta\varepsilon_{t-1}$. This model has been found by Chan and Song (2018) and others to provide competitive forecasting performance for US inflation. See Appendix 3 for details.

The first column of Table 2 reports the RMSFE and the second column reports the time-averaged CRPS over the evaluation sample. Columns 3 to 5 give the tail-weighted, right tail-weighted, and left tail-weighted CRPS, respectively. Except for the RMSFE and CRPS statistics for the LOP model (first row, in italics), which are reported as absolute values, all other values in columns 1 to 5 are computed as ratios to the LOP benchmark. Ratios less than one, for both RMSFE and CRPS, indicate an improvement of forecast performance, relative to the LOP benchmark model. The sixth column reports the p-values of the Knüppel test of density calibration.

The RMSFEs reported in the first column indicate a gain for both the EtLOP (and the EtLogOP) of just under 10% over the LOP benchmark. The CRPS values reported in column 2 likewise indicate a performance gain for EtLOP (and EtLogOP) of around 9% over the LOP benchmark. This performance gain is slightly greater when considering the tail-weighted CRPS, either using both tails or just the right tail (columns 3 and 4), whereas the left tail-weighted CRPS has a slightly smaller performance gain (column 5).

The forecast densities for the EtLOP (and EtLOgOP) are well calibrated, with p-values for the Knüppel test exceeding 5%. In contrast, the benchmark LOP (and LogOP) are poorly calibrated at the 1% significance level by the same test.

To summarise the results so far, empirically transformed opinion pools outperform their more conventional counterparts in terms of point and density forecasting performance. Turning to the UCSV model (relative to the benchmark), we see a broadly competitive performance across metrics to the EtLOP, with a very modest performance improvement generally, except for the calibration test where UCSV performs particularly well.

The upper and lower panels of Figure 5 show the 5% and 95% uncertainty bands of the forecast densities, as well as the mean forecasts, for inflation for the LOP and the EtLOP, respectively. The mean forecasts of the EtLOP model are quite close to those of the LOP.

There are differences in the shape of the forecast densities. For example, in 2010 the 5% threshold of the forecast density band in the LOP model extends below -1%, whereas in the EtLOP model it is around -0.5%. A similar pattern can be seen elsewhere, for example in 2002 and in the second half of 2015. This difference reflects the EtLOP's ability to match the skewness of the sample inflation data. Furthermore, a careful comparison of the two panels (e.g. along the vertical lines) reveals that the EtLOP's 90% band is narrower than the LOP's 90% band.

The upper and lower panels of Figure 6 display similar characteristics when comparing the LogOP and EtLogOP combinations, with the empirically-transformed combination

Figure 5: LOP and EtLOP Forecasts

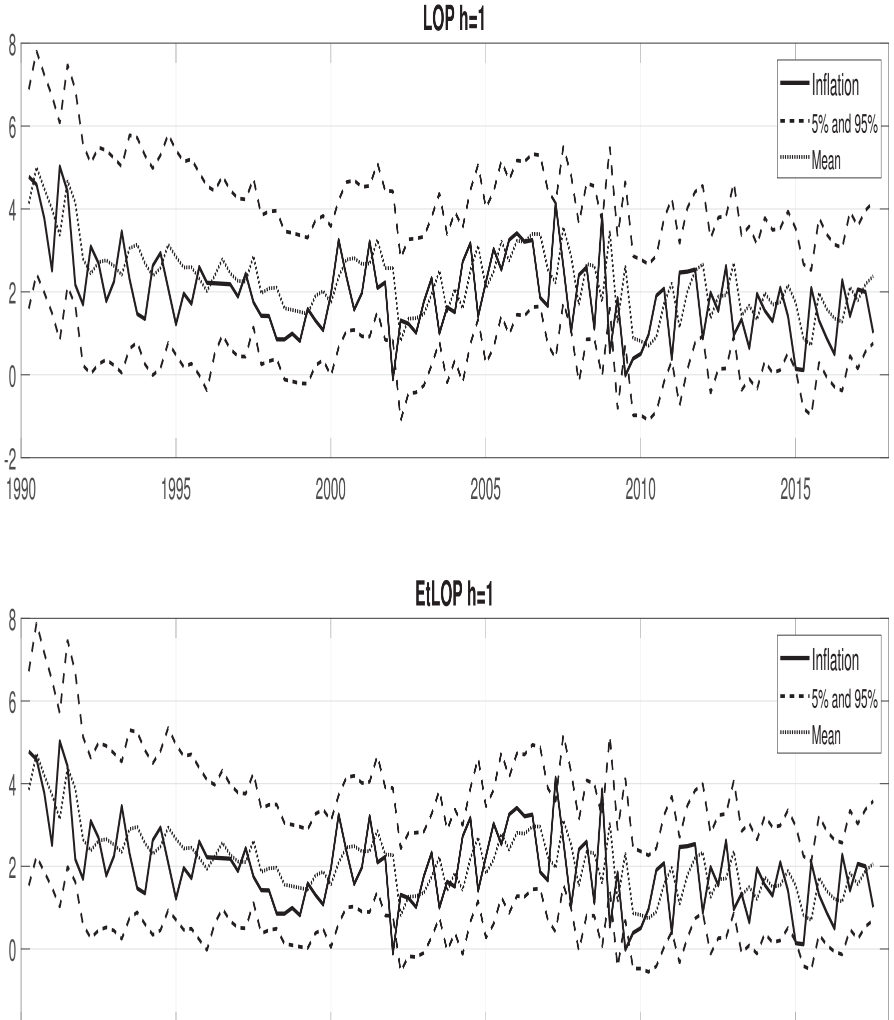
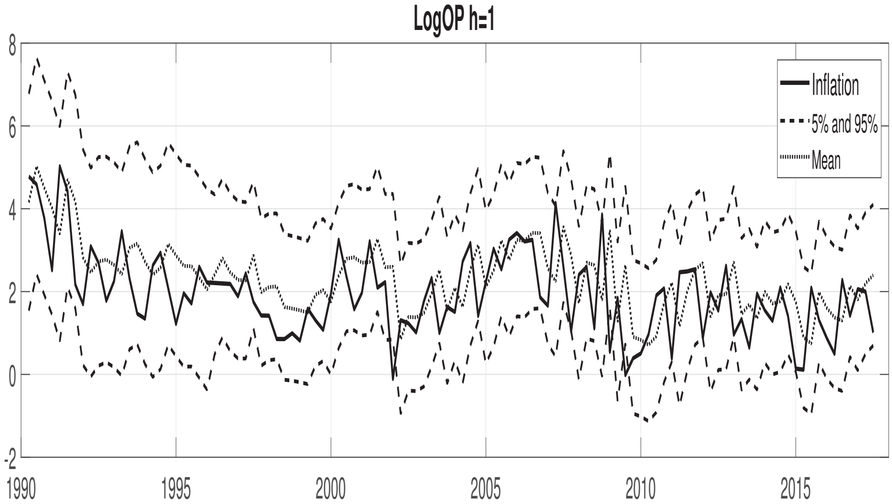


Figure 6: LogOP and EtLogOP Forecasts



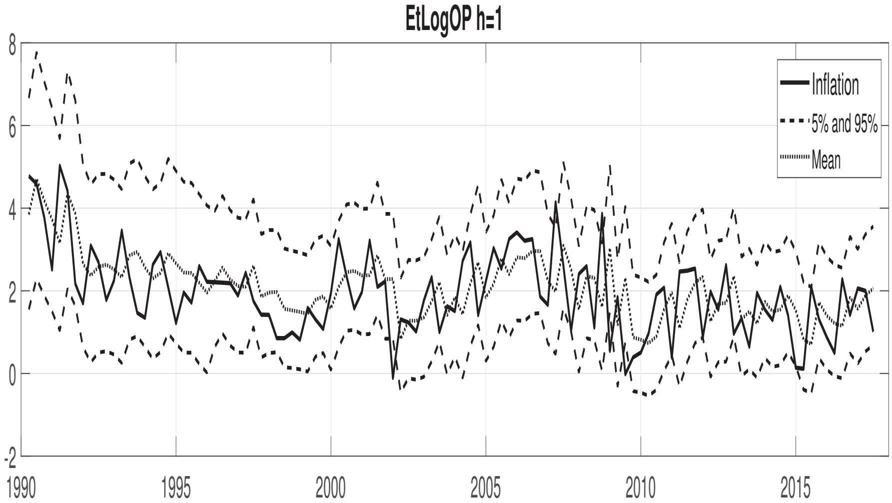


Figure 7: Forecast Densities for 2009:02

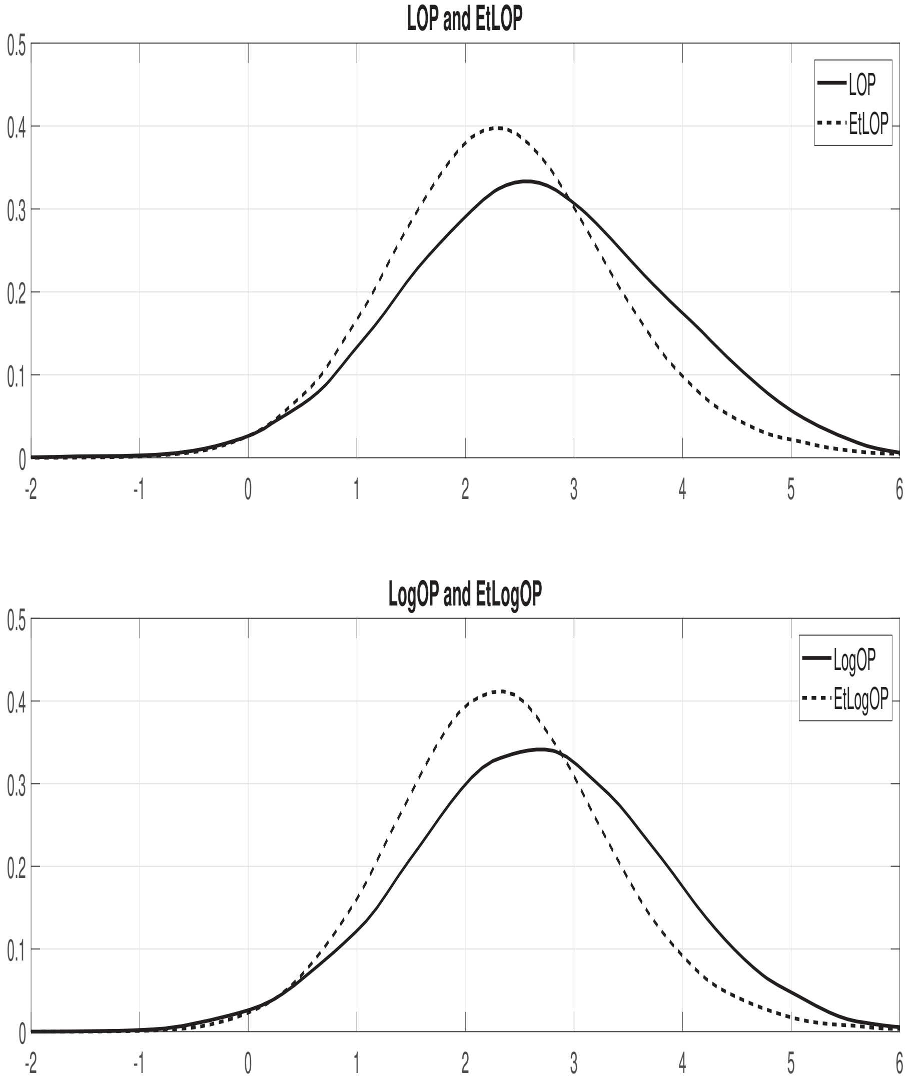
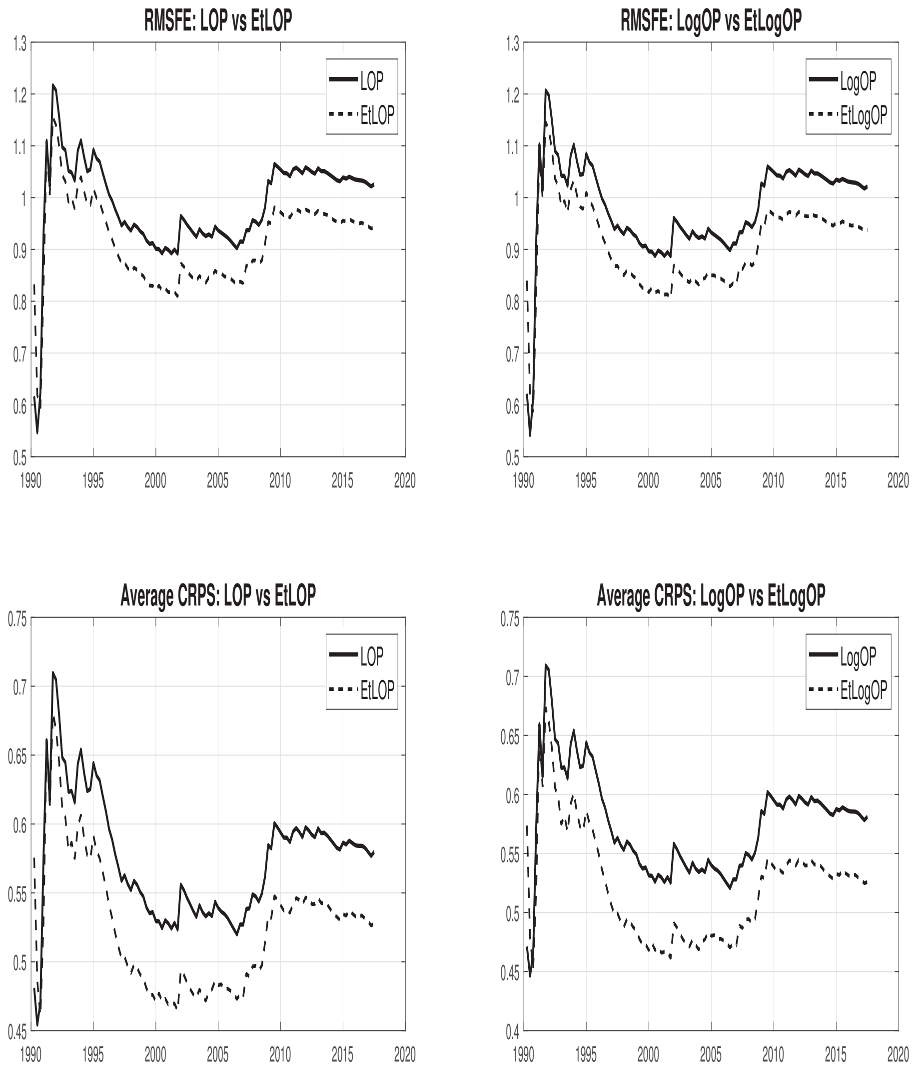


Figure 8: Recursive RMSFE and CRPS



displaying more skew and being somewhat sharper than the conventional counterpart.

Figure 7 displays the forecast densities resulting from the four specifications for the target observation 2009:2, when inflation was unusually low. The densities of both empirically-transformed opinion pools have less probability mass on high inflation outturns and are somewhat less diffuse than their conventional counterparts.

Figure 8 provides a visual confirmation of the performance differentials in terms of the levels of RMSFE and CRPS computed recursively through the evaluation sample. The EtLOP (EtLogOP) dominates the LOP (LogOP) throughout, apart from during the first year of the evaluation sample.

5 Conclusions

In this paper, we have proposed a methodology to improve the accuracy of combination inflation forecasts based on the Linear Opinion Pool. Our approach involves transforming the conventional combination forecast densities using an ECDF to match the marginal distribution of the sample data. In our application, we have analysed US inflation, using an evaluation sample considering the Great Recession, combining forecast densities produced from a system of 28 VAR models of inflation and output gaps. We have demonstrated that the Empirically-transformed Linear Opinion Pool results in considerably improved forecast performance relative to the more conventional opinion pool and that our enhanced approach is broadly competitive with a UCSV model. Further work in this area should focus on applying the methodology to survey forecasts and investigating the scope for improving the accuracy of joint forecast distributions.

References

[1] Adrian, T., N. Boyarchenko and D. Giannone (2019), "Vulnerable Growth ", *Federal Reserve Bank of New York Staff Reports, no. 794*, forthcoming *American Economic Review*.

[2] Amengual, D., E. Sentana and Z. Tian (2019), "Gaussian Rank Correlation and Regression ", unpublished manuscript, CEMFI, Spain.

[3] Bassetti, F., R. Casarin and F. Ravazzolo (2019), "Bayesian Nonparametric Calibration and Combination of Predictive Distributions", *Journal of the American Statistical Association*, forthcoming.

[4] 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.

[5] Berkowitz, J. (2001), "Testing Density Forecasts, with Applications to Risk Management", *Journal of Business and Economic Statistics*, 19, 465-474.

[6] 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.

[7] Billio, M., R. Casarin, F. Ravazzolo and H. van Dijk (2013), "Time-varying Combinations of Predictive Densities using Nonlinear Filtering", *Journal of Econometrics*, 177, 2, 213-232.

[8] Brier, G.W. (1950), "Verification of Forecasts Expressed in Terms of Probability", *Monthly Weather Review*, 78, 1-3.

[9] 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.

[10] Chan, J.C.C., G. Koop, and S.M. Potter (2013), "A New Model of Trend Inflation", *Journal of Business and Economic Statistics*, 31, 94-106.

[11] Christiano, L. and T.J. Fitzgerald (2003), "The Band Pass Filter", *International Economic Review*, 44, 2, 435-465.

[12] Clark, T.E. and T. Doh (2014), "Evaluating Alternative Models of Trend Inflation", *International Journal of Forecasting*, 30, 426-448.

[13] Clark, T.E. and S.J. Terry (2010), "Time Variation in the Inflation Passthrough of Energy Prices", *Journal of Money Credit and Banking*, 42, 7, 1419-1433.

[14] Conflitti, C., C. De Mol and D. Giannone (2015), "Optimal Combination of Survey Forecasts", *International Journal of Forecasting*, 31, 1096-1103.

[15] Croushore, D. and T. Stark (2001), "A Real-time Data Set for Macroeconomists", *Journal of Econometrics*, 105, 111-130.

[16] DeGroot M.H. and J. Mortera (1991), "Optimal Linear Opinion Pools", *Management Science*, 37, 5, 546-558.

[17] Del Negro, M., R.B. Hasegawa and F. Schorfheide (2016), "Dynamic Prediction Pools: An investigation of financial frictions and forecasting performance", *Journal of Econometrics*, 192, 2, 391-405.

[18] Diebold, F.X., T.A. Gunther and A.S. Tay (1998), "Evaluating Density Forecasts; with Applications to Financial Risk Management", *International Economic Review,* 39, 863-83.

[19] Diebold, F.X. and R.S. Mariano (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, 253-263.

[20] Diks, C., V. Panchenko, V. and D. van Dijk (2011), "Likelihood-Based Scoring Rules for Comparing Density Forecast in Tails", *Journal of Econometrics*, 163(2), 215-230.

[21] Edge, R.M and J.B. Rudd (2012), "Real-Time Properties of the Federal Reserve's Output Gap", Finance and Economics Discussion Series 2012-86. Board of Governors of the Federal Reserve System (U.S.).

[22] 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.

[23] Ganics, G. (2017), "Optimal Density Forecast Combinations", Bank of Spain Working Paper No. 1751.

[24] Garratt, A., Lee, K., Mise, E. and K. Shields (2008), "Real-Time Representations of the Output Gap", *Review of Economics and Statistics*, 90, 4, 792-804.

[25] 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.

[26] Garratt, A., J. Mitchell, and S.P. Vahey (2014), "Measuring Output Gap Nowcast Uncertainty", *International Journal of Forecasting*, 30, 2, 268-279.

[27] Geweke, J. and G. Amisano (2011), "Optimal Prediction Pools", *Journal of Econometrics*, 164, 130-141.

[28] Genest, C. and J.V. Zidek (1986), "Combining Probability Distributions: A critique and an annotated bibliography", *Statistical Science*, 1, 1, 114-135.

[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 Ranjan, R. (2011), "Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules", *Journal of Business and Economic Statistics*, Volume 29, Issue 3.

[31] Gneiting, T. and R. Ranjan (2013), "Combining Predictive Distributions", *Electronic Journal of Statistics*, 7, 7471782.

[32] Hall, S.G. and J. Mitchell (2007), "Density Forecast Combination", *International Journal of Forecasting*, 23, 1-13.

[33] Harvey, A.C. (2006), "Forecasting with Unobserved Components Time Series Models". In G. Elliot, C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting,* North-Holland, 327-412.

[34] Harvey, D., S. Leybourne and P. Newbold (1997), "Testing the Equality of Prediction Mean Squared Errors", *International Journal of Forecasting*, 13, 281-291.

[35] Hodrick, R. and E. Prescott (1997), "Post-War U.S. Business Cycles: An Empirical Investigation", *Journal of Money, Banking and Credit*, 29, 1-16.

[36] 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.

[37] Kascha, C. and F. Ravazzolo (2010), "Combining Inflation Density Forecasts", *Journal of Forecasting*, 29, 231-250.

[38] Kapetanios, G., J. Mitchell, S. Price and N. Fawcett (2015), "Generalised Density Forecast Combinations", *Journal of Econometrics*, 188, 1, 150-165.

[39] Knüppel, M. (2015), "Evaluating the Calibration of Multi-Step Ahead Density Forecasts Using Raw Moments", *Journal of Business Economics and Statistics*, 33, 270-281.

[40] Loaiza-Maya, R. and M.S. Smith (2018), "Real-Time Macroeconomic Forecasting with a Heteroskedastic Inversion Copula", *Journal of Business and Economic Statistics*, forthcoming.

[41] Marcellino, M., J. Stock and M.W. Watson (2006), "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-steps Ahead", *Journal of Econometrics*, 135, 499-526.

[42] 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.

[43] Mitchell, J. and S.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, 995-1033.

[44] Nelsen, R (2006), *An Introduction to Copulas*, 2nd ed., New York:NY: Springer.

[45] 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.

[46] 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.

[47] Pettenuzzo, D. and F. Ravazzolo (2016), "Optimal Portfolio Choice Under Decision-Based Model Combinations", *Journal of Applied Econometrics*, 31, 1312-1332.

[48] Ranjan, R. and T. Gneiting (2010), “Combining Probability Forecasts”, *Journal of the Royal Statistical Society Series B*, 72, 71-91.

[49] Ravazzolo, F. and S.P. Vahey (2014), “Forecast Densities for Economic Aggregates From Disaggregates Ensembles”, *Studies in Nonlinear Dynamics and Econometrics*.

[50] Rosenblatt, M. (1952), “Remarks on a Multivariate Transformation”, *The Annals of Mathematical Statistics*, 23, 470-472.

[51] Rossi, B. and T. Sekhposyan (2014), “Evaluating Predictive Densities of US Output Growth and Inflation in a Large Macroeconomic Data Set”, *International Journal of Forecasting*, 30(3), 662682.

[52] Rossi, B. and T. Sekhposyan (2019), “Alternative Tests for Correct Specification of Conditional Predictive Densities”, *Journal of Econometrics*, forthcoming.

[53] 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.

[54] Stock, J.H. and M.W. Watson (2007), “Has Inflation Become Harder to Forecast?”, *Journal of Money, Credit, and Banking*, 39, 3-34.

[55] Waggoner, D.F. and T. Zha (2012) “Confronting Model Misspecification in Macroeconomics”, *Journal of Econometrics*, 171, 2, 167-184.

[56] Yellen, J. (2006), “The US Economy in 2006”, Speech before an Australian Business Economists luncheon Sydney, Australia, March 15, 2006.

[57] Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley and Sons.

Table 1: Performance of Empirically-transformed Combinations, by Bandwidth, *bw*

(a) EtLOP

bw	BS	RMSFE Ratio	LS	CRPS	Knup
0.800	0.927	0.834	-3.757	0.530	0.014
0.825	0.923	0.836	-3.746	0.529	0.022
0.850	0.917	0.832	-3.737	0.529	0.016
0.875	0.919	0.831	-3.729	0.528	0.024
0.900	0.916	0.830	-3.722	0.528	0.034
0.925	0.908	0.830	-3.715	0.528	0.045
0.950	0.912	0.829	-3.709	0.528	0.057
0.975	**0.906**	**0.829**	**-3.704**	**0.528**	**0.068**
1.000	0.907	0.829	-3.699	0.528	0.079
1.025	0.908	0.829	-3.695	0.529	0.089
1.050	0.905	0.829	-3.691	0.529	0.096
1.075	0.898	0.829	-3.688	0.530	0.100
1.100	0.899	0.829	-3.684	0.530	0.103
1.125	0.912	0.830	-3.681	0.531	0.103
1.150	0.894	0.830	-3.678	0.532	0.101
1.750	0.903	0.831	-3.675	0.533	0.097
1.200	0.899	0.831	-3.673	0.531	0.091

(b) EtLogOP

bw	BS	RMSFE Ratio	LS	CRPS	Knup
0.800	0.926	0.830	-3.717	0.528	0.013
0.825	0.921	0.829	-3.705	0.528	0.022
0.850	0.917	0.828	-3.696	0.527	0.017
0.875	0.925	0.827	-3.687	0.526	0.027
0.900	0.917	0.827	-3.679	0.526	0.039
0.925	0.913	0.826	-3.672	0.526	0.053
0.950	0.911	0.826	-3.666	0.526	0.070
0.975	**0.910**	**0.826**	**-3.661**	**0.526**	**0.087**
1.000	0.909	0.825	-3.656	0.526	0.104
1.025	0.910	0.825	-3.651	0.527	0.120
1.050	0.903	0.826	-3.647	0.527	0.134
1.075	0.908	0.826	-3.644	0.528	0.144
1.100	0.901	0.826	-3.641	0.528	0.152
1.125	0.907	0.826	-3.638	0.529	0.156
1.150	0.898	0.827	-3.635	0.530	0.157
1.175	0.897	0.827	-3.633	0.531	0.155
1.200	0.904	0.828	-3.630	0.531	0.150

Notes: The first column reports the scaled Brier Score (BS), for the inflation event of less than the sample mean, calculated by dividing the Brier Score by the measure of data uncertainty, defined as: $\overline{o}(1-\overline{o})$, where $\overline{o}$ is the observed frequency of the event in question over the evaluation. The second column, RMSFE, reports the ratio of the root mean squared forecast error relative to the benchmark. Column three reports the logarithmic score (LS), column four the CRPS and column five the p-values of the Knüppel (2015) test for density calibration.

Table 2: Forecast Performance of Empirically-transformed and Conventional Combinations, $bw = 0.975$

Model	**RMSFE**	**CRPS**	**CRPS-TW**	**CRPS-RT**	**CRPS-LT**	**Knup**
LOP	*1.026*	*0.580*	*0.129*	*0.184*	*0.171*	0.000
EtLOP	0.918**	0.910**	0.899**	0.891**	0.924**	0.070
LogOP	0.996**	1.002**	0.992**	0.989**	1.000	0.000
EtLogOP	0.914**	0.907**	0.891**	0.886**	0.924**	0.089
UCSV	0.904**	0.903**	0.888**	0.867**	0.935**	0.994

Notes: The first row in italics reports absolute values of the RMSFE and CRPS statistics for the LOP, excluding the final column which reports the p-values of the Knüppel calibration test. The remaining rows report ratios, relative to the LOP, for RMSFE and CRPS. Ratios less than one, for both RMSFE and CRPS, indicate an improvement in forecast performance relative to the LOP. As a rough guide, using a Harvey et al. (1997) small-sample adjustment of the Diebold and Mariano (1995) test, the superscript $**$ denotes significantly different from the LOP at the 5% level for RMSFE. The corresponding statistic for CRPS is denoted similarly. See notes to Table 1 for columns labelled CRPS and Knup. Columns three (CRPS-TW), four (CRPS-RT) and five (CRPS_LT) report tail-weighted, right tail weighted and left tail weighted CRPS; see Gneiting and Ranjan (2011) and Diks et al. (2011).

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 US data.[5]

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 seek to 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 US 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 modeling of the data. Their approach implicitly assumes that the series is captured reasonably well by a random walk model and

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

that, if there is drift present, this can be proxied by the average growth rate over the sample.

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. Watson (2007) reviews band-pass filtering methods.

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

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

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:

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

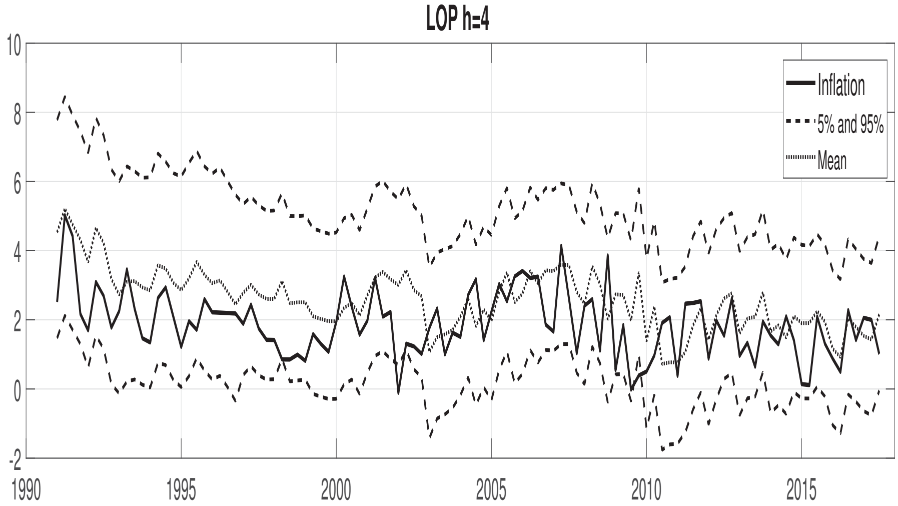
where λ is the frequency in radians, ρ is a damping factor and κ_t and κ_t^* are two independent white noise Gaussian disturbances with common variance σ_κ^2. We estimate this model by maximum likelihood, exploiting the Kalman filter, and estimates of the trend and cyclical components are obtained using the Kalman smoother. For a detailed description of the unobserved components approach see Harvey's (2006) literature review and the references therein.

Appendix 2: Forecast Performance of Empirically-transformed and Conventional Combinations, $h = 4$

Model	RMSFE	CRPS	CRPS-TW	CRPS-RT	CRPS-LT	Knup
LOP	*1.199*	*0.682*	*0.148*	*0.232*	*0.184*	*0.000*
EtLOP	0.894**	0.881**	0.889**	0.871**	0.895**	0.000
LogOP	0.980**	0.996**	0.986**	0.986**	0.996**	0.000
EtLogOP	0.873**	0.864**	0.867**	0.842**	0.887**	0.000
UCSV	0.794**	0.818**	0.834**	0.734**	0.929**	0.075

Notes: The first row in italics reports absolute values of the RMSFE and CRPS statistics for the LOP, excluding the final column which reports the p-values of the Knüppel calibration test. The remaining rows report ratios, relative to the LOP, for RMSFE and CRPS. Ratios less than one, for both RMSFE and CRPS, indicate an improvement in forecast performance relative to the LOP. As a rough guide, using a Harvey et al. (1997) small-sample adjustment of the Diebold and Mariano (1995) test, the superscript $**$ denotes significantly different from the LOP at the 5% level for RMSFE. The corresponding statistic for CRPS is denoted similarly. See notes to Table 1 for columns labelled CRPS and Knup. Columns three (CRPS-TW), four (CRPS-RT) and five (CRPS_LT) report tail-weighted, right tail weighted and left tail weighted CRPS; see Gneiting and Ranjan (2011) and Diks et al. (2011).

Figure 9: LOP and EtLOP Forecasts



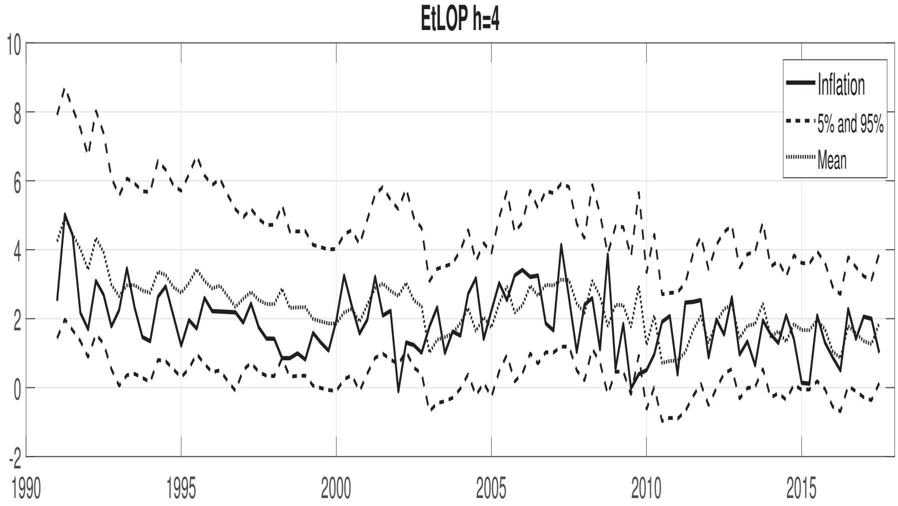
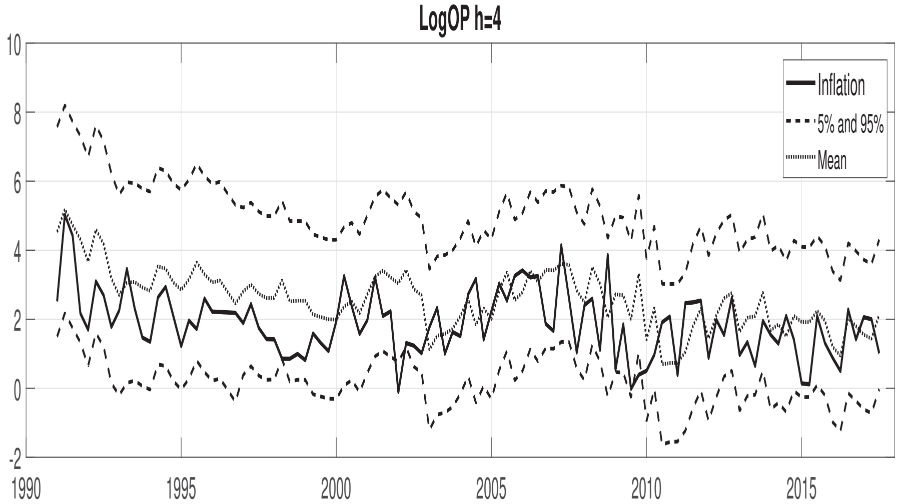


Figure 10: LogOP and EtLogOP Forecasts



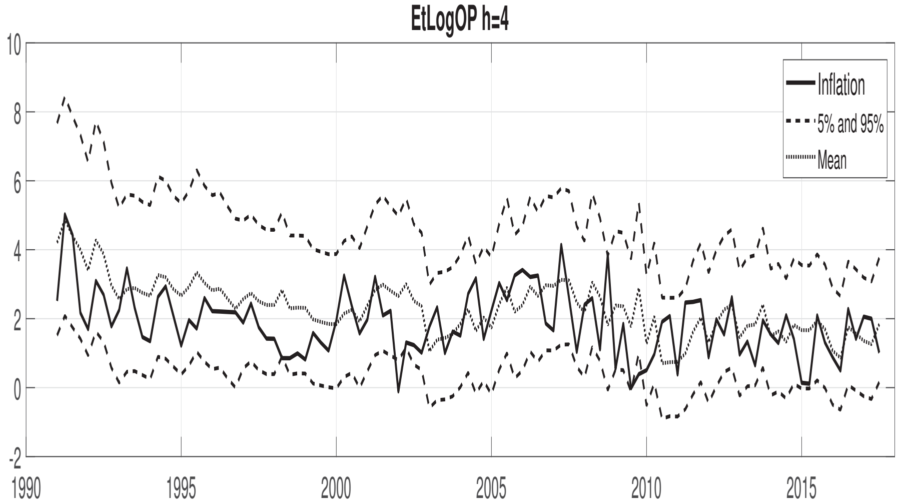


Figure 11: Forecast Densities for 2009:02

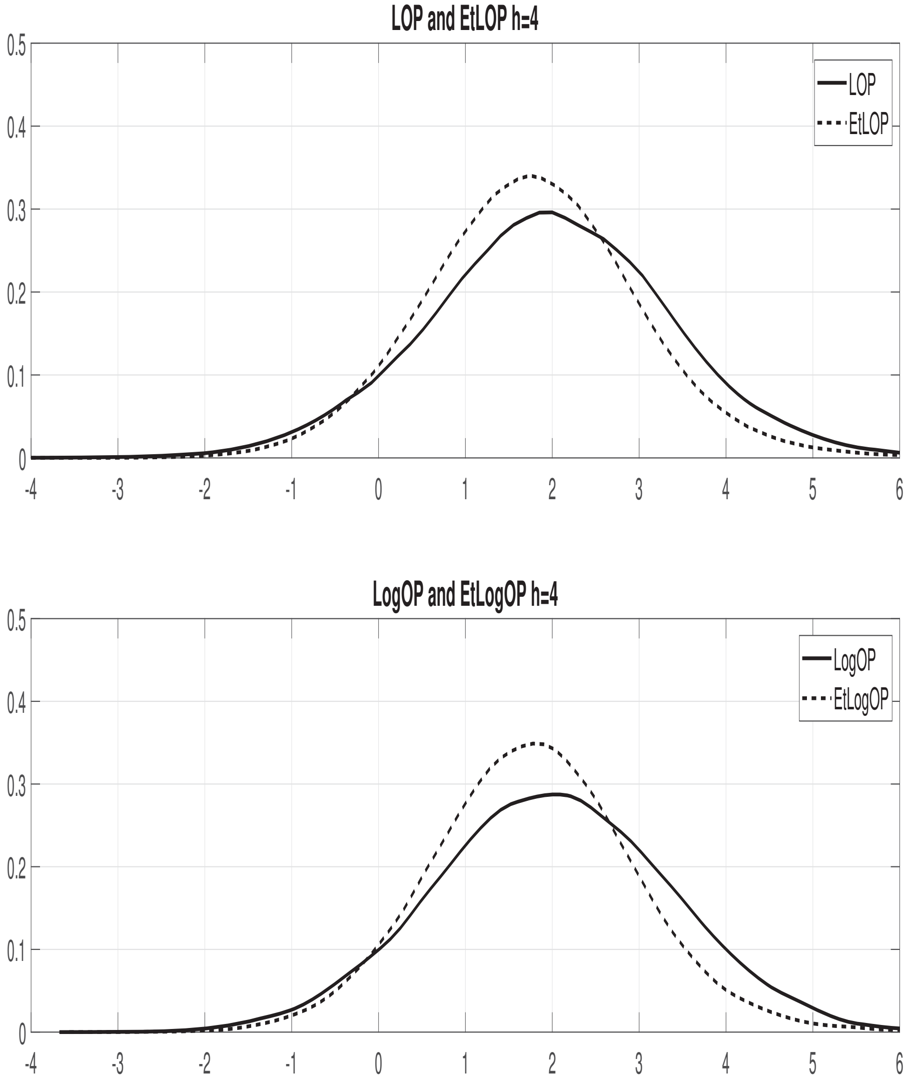
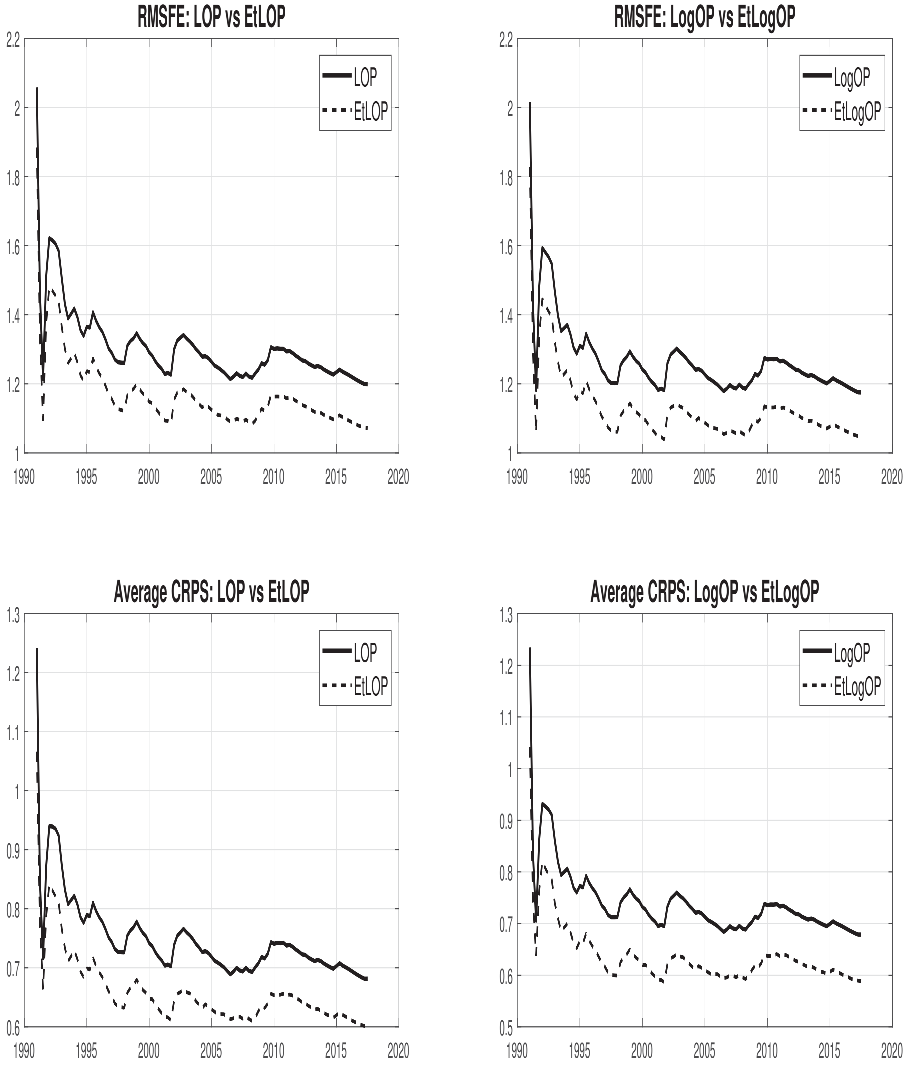


Figure 12: Recursive RMSFE and CRPS, $h = 4$



Appendix 3: Unobserved Components with Stochastic Volatility (UCSV)

Since Stock and Watson (2007) the predictive accuracy of candidate methods relative to the UCSV approach has become standard within the inflation forecasting literature. With a focus on forecast densities, it is convenient to adopt the Bayesian approach described in Chan and Song (2018), Chan, Koop, and Potter (2013) and Clark and Doh (2014). The model specifies the following trend-cycle decomposition for inflation, π_t:

$$\pi_t = \pi_t^* + u_t^\pi, \qquad u_t^\pi \sim N(0, e^{h_t})$$

where π_t^* represents trend inflation and u_t^π is a transitory deviation from the trend, often referred to as the inflation gap. To define the UCSV model of Stock and Watson (2007), we augment the trend-cycle decomposition with equations specifying AR(1) processes for the inflation trend, π_t^*, and the log volatilities of the transitory and trend components, h_t and g_t respectively:

$$\begin{aligned}
\pi_t^* &= \pi_{t-1}^* + u_t^{\pi^*}, \qquad u_t^{\pi^*} \sim N(0, e^{g_t}) \\
h_t &= h_{t-1} + u_t^h, \qquad u_t^h \sim N(0, \sigma_h^2) \\
g_t &= g_{t-1} + u_t^g, \qquad u_t^g \sim N(0, \sigma_g^2).
\end{aligned}$$

The model is estimated using Markov Chain Monte Carlo (MCMC) methods, implemented using the Gibbs sampler that sequentially draws from the full conditional distributions of the parameters and the latent states. The parameters here are σ_h^2 and σ_g^2 where the latent states are g, h and π^*. The priors are non-informative with estimated smoothing parameters σ_h^2 and σ_g^2; see Chan and Song (2018) Appendix A. In contrast, Stock and Watson (2007) set these parameters to 0.2 for US inflation. We draw 50,000 iterates, with 5,000 burnin.