Online Appendix for Empirically-transformed Linear Opinion Pools

Anthony Garratt
(University of Warwick)

Timo Henckel
(ANU and CAMA)

Shaun P. Vahey
(University of Warwick and CAMA)

January 6th, 2022

Appendix 1: Histograms of Simulated EtLOP and BLOP CRPS Ratios

In Figures A1.1 and A1.2 we report histograms of the simulated Continuous Rank Probability Score (CRPS) ratios for the Empirically-transformed Linear Opinion Pool (EtLOP, red bars) and Beta Linear Opinion Pool (BLOP, blue bars) inflation forecast densities, one step ahead. The sample-average CRPS scores are relative to the equivalent equal weight benchmark LOP. We report histograms using sample sizes 100, 150, 200, 250 and 500, respectively.

Across all sample sizes, the entire histogram for EtLOP ratio lies to the left of one (the benchmark), with a mean value of approximately 0.87. The BLOP histogram lies to the right of the EtLOP histogram, with no overlap, and an approximate mean value close to one. This suggests a consistent preferred performance for EtLOP over BLOP.

Figure A1.1: Sample Size 100, 150 and 200.

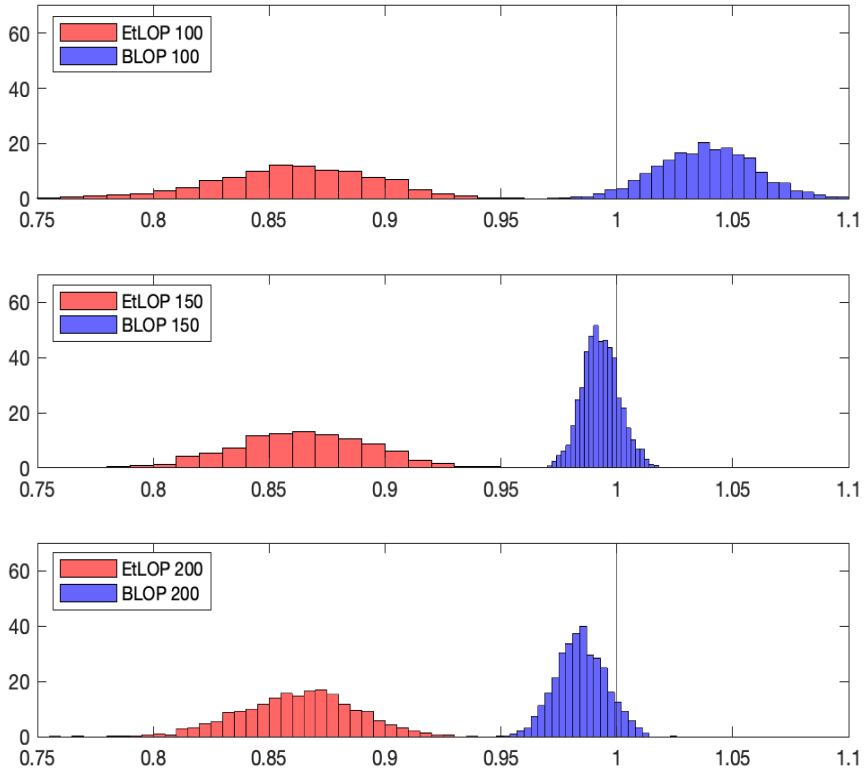
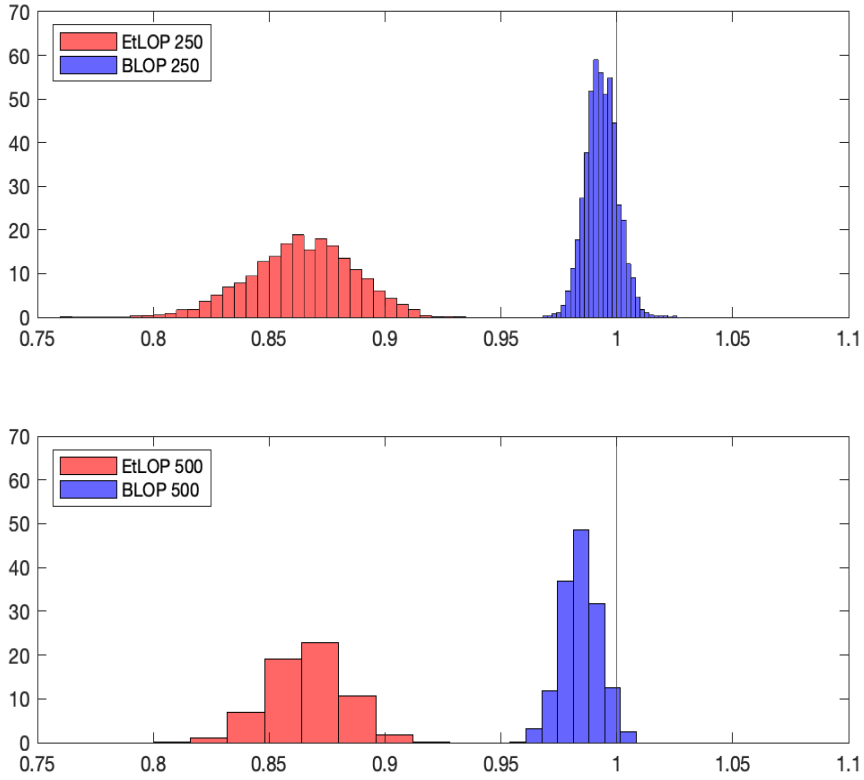


Figure A1.2: Sample Size 250 and 500.



Appendix 2: Application Forecast Evaluation for BLOP and UCSV Models

We contrast the forecast performance of the EtLOP inflation forecast densities with those from the BLOP combination, where we estimate shape parameters at each recursion, and an Unobserved Components Stochastic Volatility (UCSV) model, fitted to the univariate inflation series, as described in Appendix 3.

Tables A2.1 and A2.2 report, for the $h = 1$ and $h = 4$ forecast horizons, the RMSFE and the sample-average CRPS in columns 2 and 3. Columns 4 through 6 give the tail-weighted, right tail-weighted, and left tail-weighted CRPS, respectively. The last column, RS, reports the test statistic of Rossi and Sekhposyan (2019) for uniform PITS. The equal weight benchmark LOP statistics (first row, in italics), are reported as absolute values. Except for the RS statistic in column 7, all other values in the remaining rows are computed as ratios to the benchmark. Ratios less than one indicate an improvement in forecast performance, relative to the benchmark LOP. As a rough guide, we use the Giacomini and White (2006) statistic to test the null that the BLOP, EtLOP and UCSV RMSFE and CRPS (overall average and tail weighted) are equal to that of the benchmark (i.e. the ratios are not statistically different from one).

The RMSFE performance gain at forecast horizon $h = 1$ is largest for the EtLOP at 13%; the corresponding gains for the BLOP and the UCSV model are 9% and 10%, respectively. Turning to the longer forecast horizon, $h = 4$, EtLOP has a 34% gain, versus 26% and 27% for the BLOP and UCSV models, respectively.

The CRPS performance gain is also strongest for the EtLOP, for both forecast horizons $h = 1$ and $h = 4$. Gains for EtLOP are 14% and 33%, compared to gains for the BLOP of 9% and 18% for forecast horizons $h = 1$ and $h = 4$, respectively. The UCSV model is closer to the EtLOP with gains of 10% and 25% for $h = 1$ and $h = 4$, respectively. The tail-weighted CRPS statistics are consistent with the average CRPS results, suggesting the largest gains relative to the equal weight benchmark LOP for the EtLOP, where the gains are larger for the right tail.

Turning to gauging absolute, rather than relative, forecast performance, in terms of the uniformity of the PITS, the Kolmogorov-Smirnov based RS statistics indicate specification failure at the $h = 1$ and $h = 4$ horizons for the equal weight benchmark LOP at the 10% (and 1%) level of significance. In contrast, for $h = 1$, BLOP cannot reject the null of "correct specification" at the 10% level of significance, with an RS test statistic of 0.815. But, for $h = 4$, BLOP rejects the null at the 10% (and 1%) level, with an RS test statistic of 2.022. The $h = 1$ and $h = 4$ EtLOP and UCSV combinations both fail to reject the null of uniform PITS at the 10% level of significance.

The remaining Figures provide additional evidence on the evaluation of the EtLOP (and benchmark) combinations for the forecast horizons $h = 2, 3, 4$. Focusing on horizon $h = 4$, in Figures A2.1 and A2.2, the predictive densities exhibit the same features as the equivalent $h = 1$ densities, namely the EtLOP is less diffuse, generally, with positive and significant skew. Whereas the equal weight benchmark LOP densities are typically more diffuse and symmetric. The figures that plot skew and a p-value testing the null of the zero skew illustrate the difference between the shapes of the EtLOP and benchmark densities. Figures A2.6-A2.7 and A2.11-A2.12 display similar patterns for forecast horizons $h = 2$ and $h = 3$, respectively.

For forecast horizon $h = 4$, Figure A2.3 displays the performance gain of the EtLOP over the benchmark in terms of the recursive RMSFE and the sample-average CRPS throughout the

evaluation. The histograms in Figure A2.4 indicate close to uniform PITs for EtLOP, consistent with the RS tests reported in Table A2.2. We observe similar patterns for $h = 2$ and $h = 3$ in Figures A2.8-A2.9 and A2.13-A2.14, respectively.

Figure A2.5 displays the EtLOP and equal weight benchmark LOP forecast densities for the target observations of 2009:1 through to 2009:4, when inflation was unusually low. As with the $h = 1$ results, the EtLOP densities at $h = 4$ have less probability mass on high inflation realisations and are somewhat less diffuse than their conventional counterparts. Again a similiar pattern is observed for $h = 2$ and $h = 3$ in Figures A2.10 and A2.15, respectively.

Table A2.1: Forecast Performance of Empirically-transformed and Conventional Combinations, $h = 1$

Model	RMSFE	CRPS	TW	RTW	LTW	RS
LOP	*1.061*	*0.589*	*0.131*	*0.183*	*0.179*	2.085†
BLOP	0.914**	0.910	0.924**	0.899**	0.944**	0.815
EtLOP	0.867**	0.860**	0.850**	0.839**	0.877**	0.546
UCSV	0.902**	0.906**	0.909**	0.878**	0.928**	0.717

Notes: Row 1, for the LOP, reports absolute values, whereas the remaining rows report ratios relative to the LOP. Columns 2-6, report RMSFE, the CRPS and the tail-weighted (TW), right-tail weighted (RTW) and left-tail weighted (LTW) CRPS statistics, respectively (see Gneiting and Ranjan (2011)). Ratios less than one indicate an improvement in forecast performance relative to the LOP. The superscript $**$ denotes significantly different from the LOP at the 5% level for RMSFE and CRPS statistics. The last column, denoted RS, reports the test statistic of Rossi and Sekhposyan (2019), where † indicates rejection of the null of good calibration of the densities at the 10% significance level.

Table A2.2: Forecast Performance of Empirically-transformed and Conventional Combinations, $h = 4$

Model	RMSFE	CRPS	TW	RTW	LTW	RS
LOP	*1.333*	*0.743*	*0.163*	*0.259*	*0.203*	3.803†
BLOP	0.740**	0.821**	0.906**	0.710	0.958	2.022†
EtLOP	0.660**	0.670**	0.696**	0.594**	0.751**	1.231
UCSV	0.732**	0.754**	0.776**	0.649**	0.862**	0.886

Notes: See notes to Table A2.1.

Figures for $h = 4$

Figure A2.1: EtLOP and LOP Forecasts, $h = 4$, 1990:1 to 2020:2

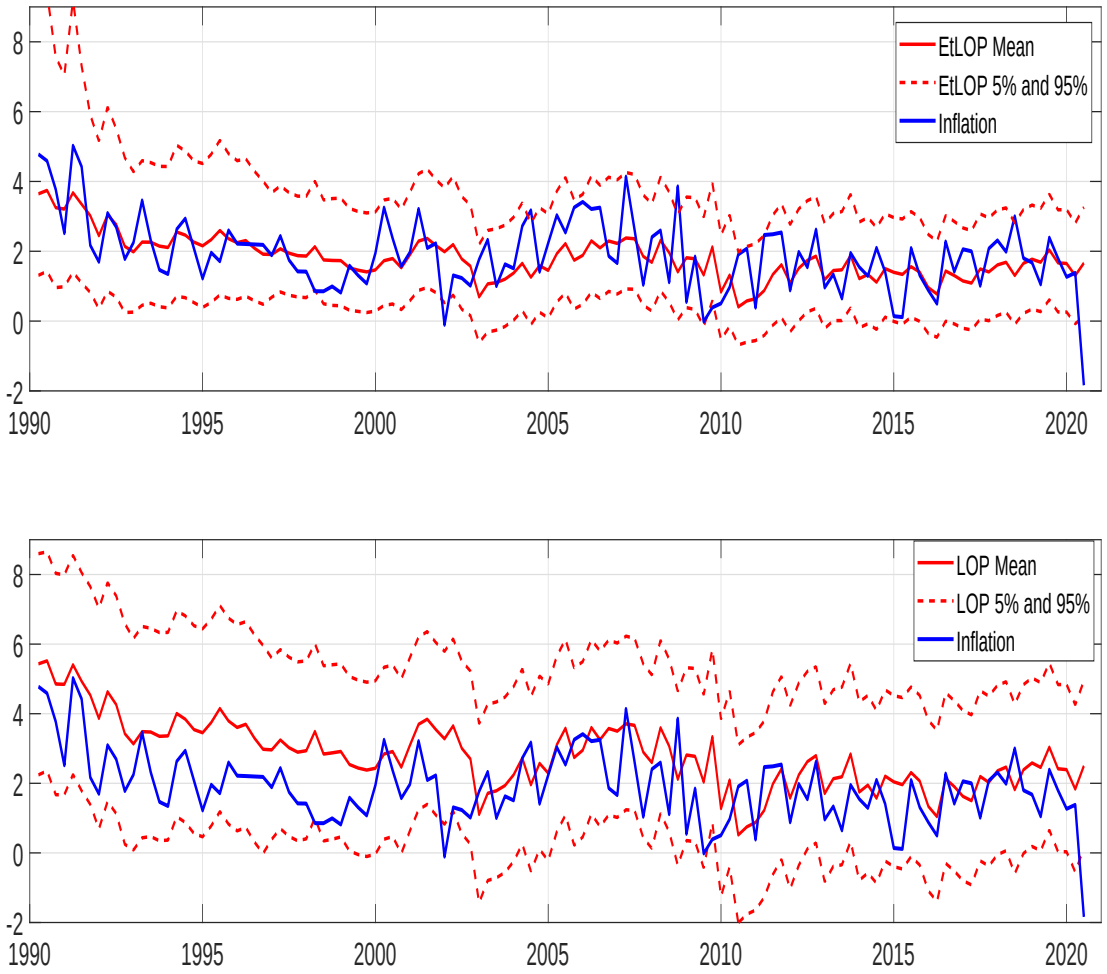
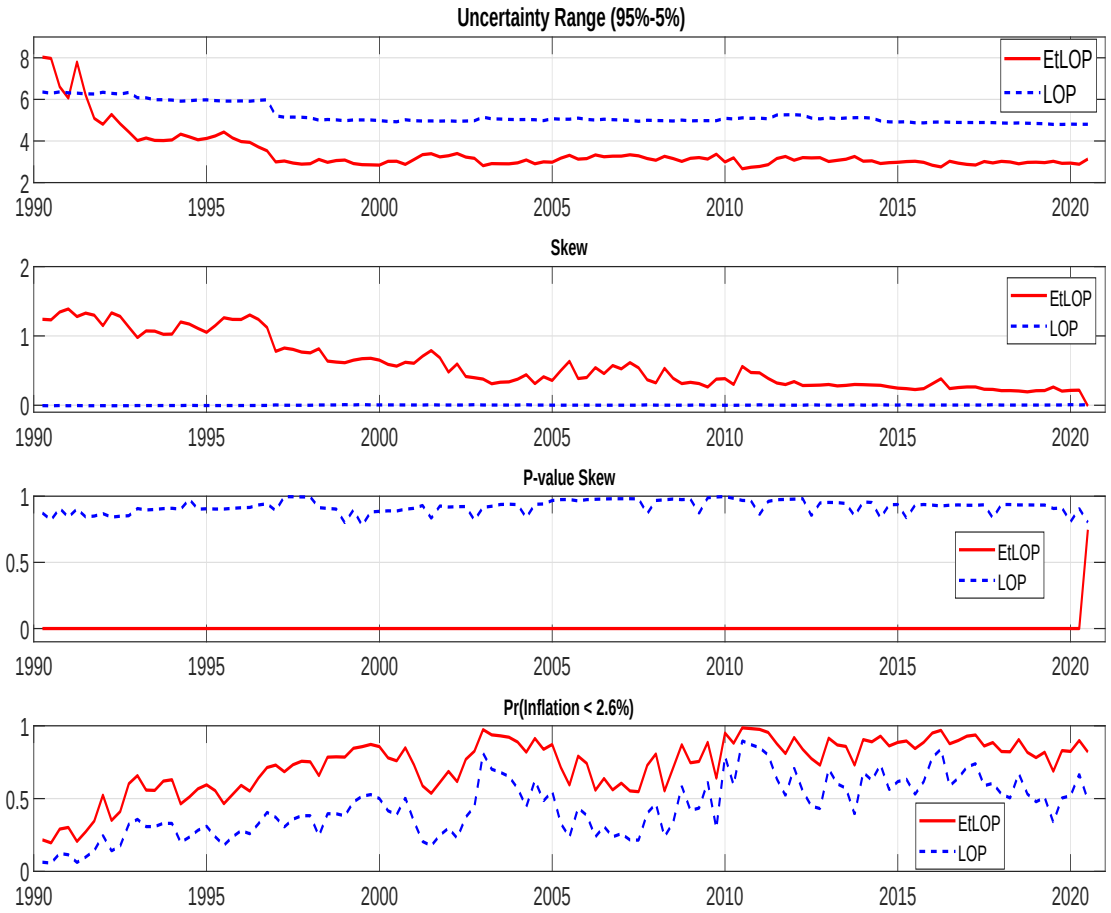


Figure A2.2: Uncertainty Range, Skew and Pr(Inflation$< 2.6\%$), of EtLOP and LOP, $h = 4$, 1990:1 to 2020:2



Notes: Skew is defined as: $s_0 = \sqrt{n(n-1)}s_1/n-2$. Where $n = 5000$, $s_1 = (1/n)\sum_{i-1}^{n}(x_i-\overline{x})^3/\left[\sqrt{(1/n\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^3$ and x_i are the 5000 forecast density iterates at each evaluation period.

Figure A2.3: RMSFE & CRPS, EtLOP & LOP, $h = 4$, 1990:1 to 2020:2

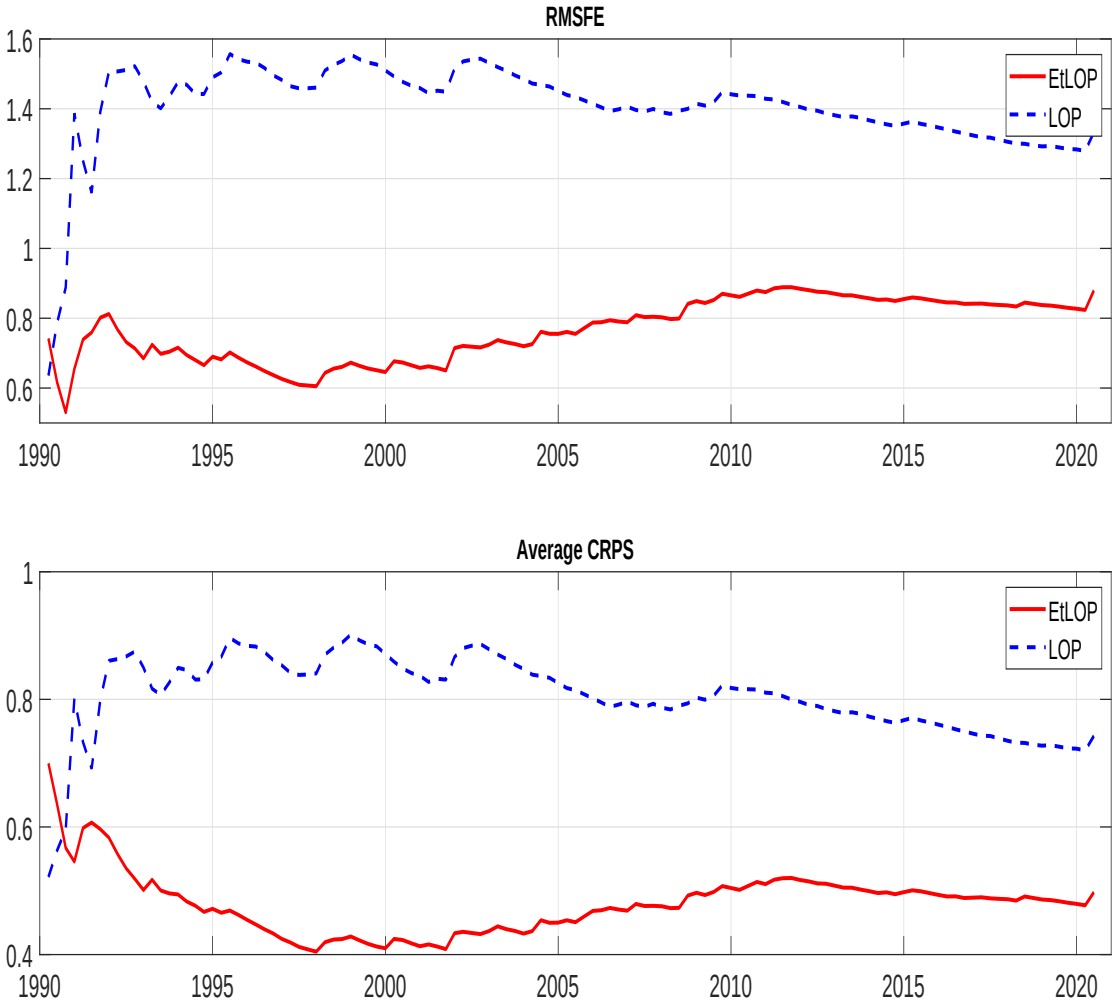
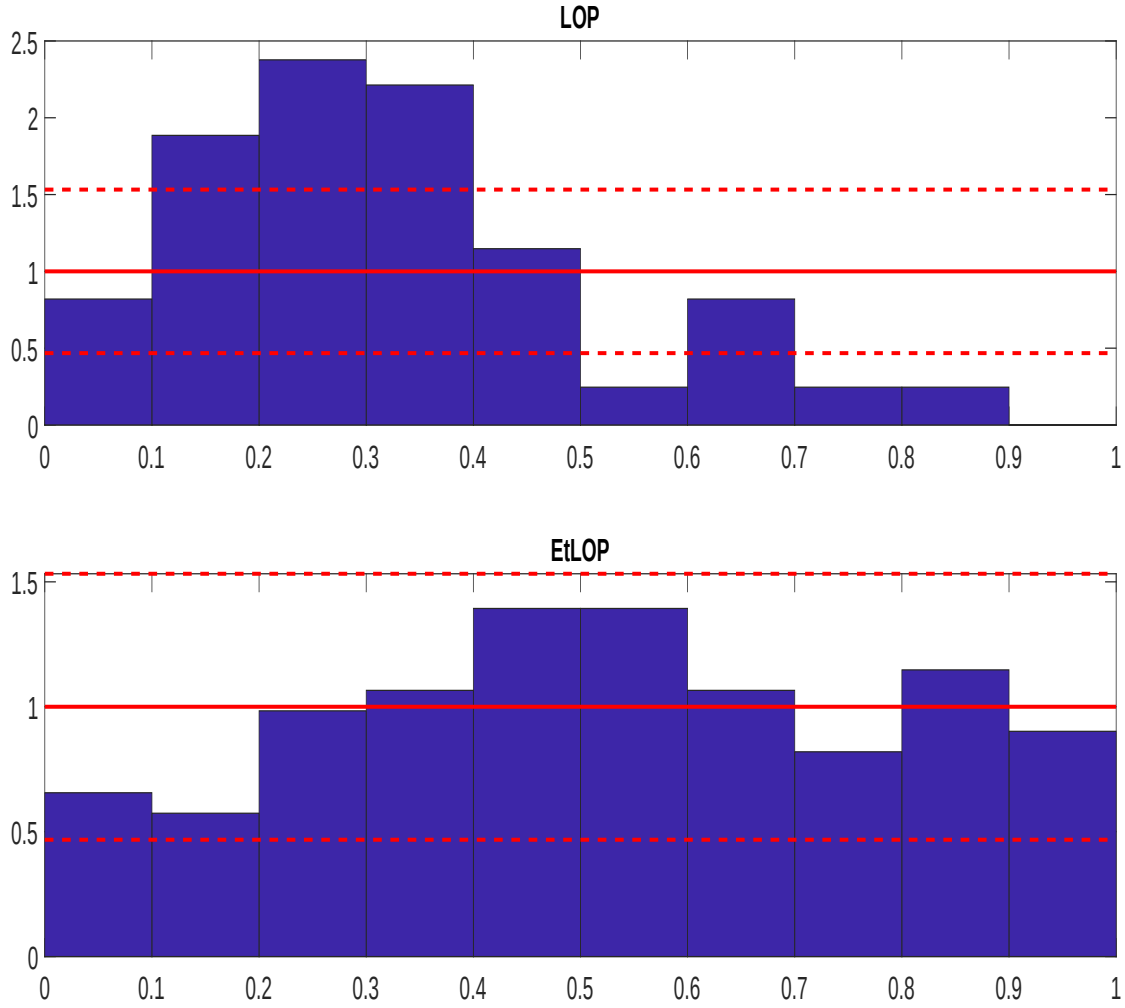
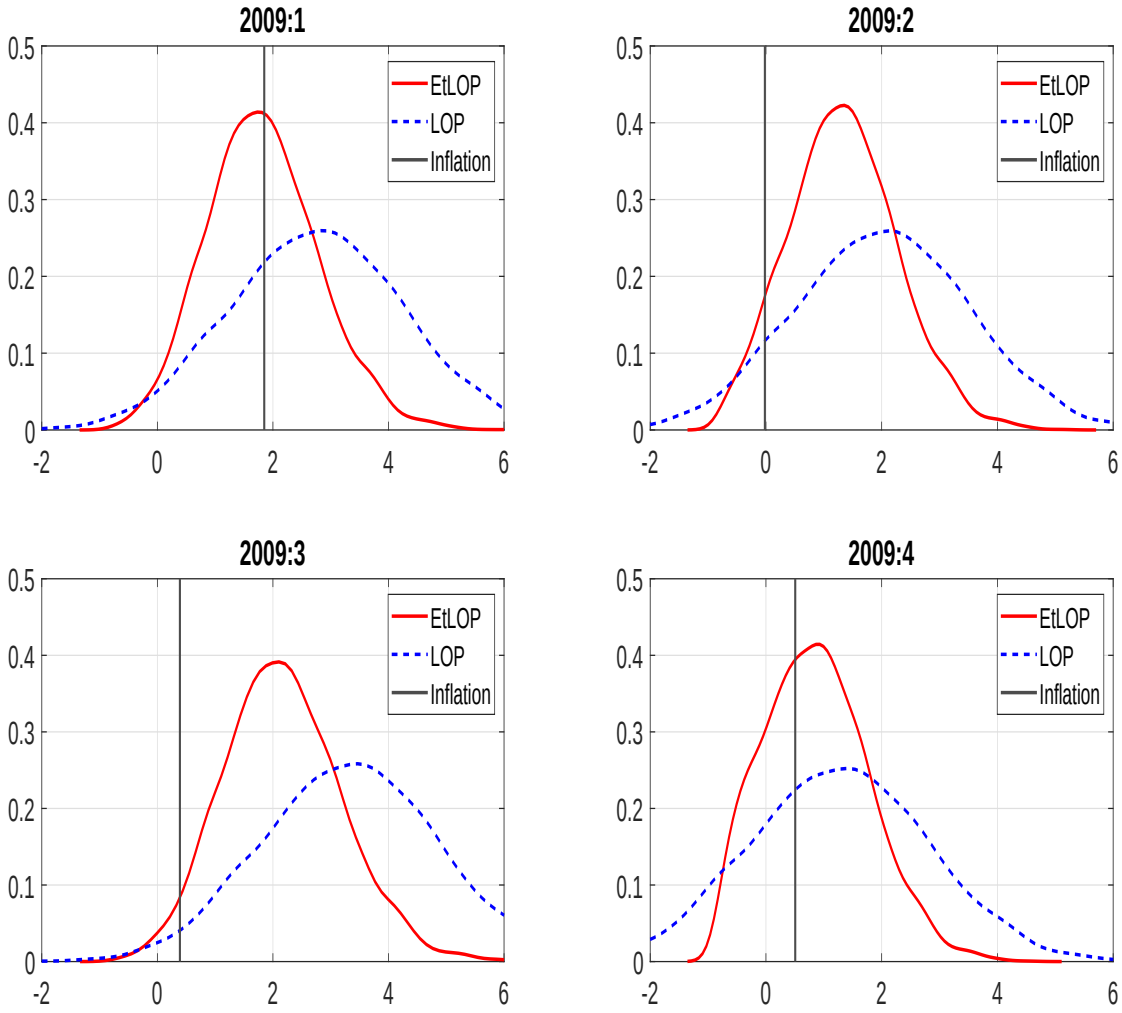


Figure A2.4: PITS, LOP and EtLOP, $h = 4$, 1990:1 to 2020:2



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 A2.5: Forecast Densities for 2009, EtLOP and LOP, $h = 4$



Figures for $h = 2$

Figure A2.6: EtLOP and LOP Forecasts, $h = 2$, 1990:1 to 2020:2

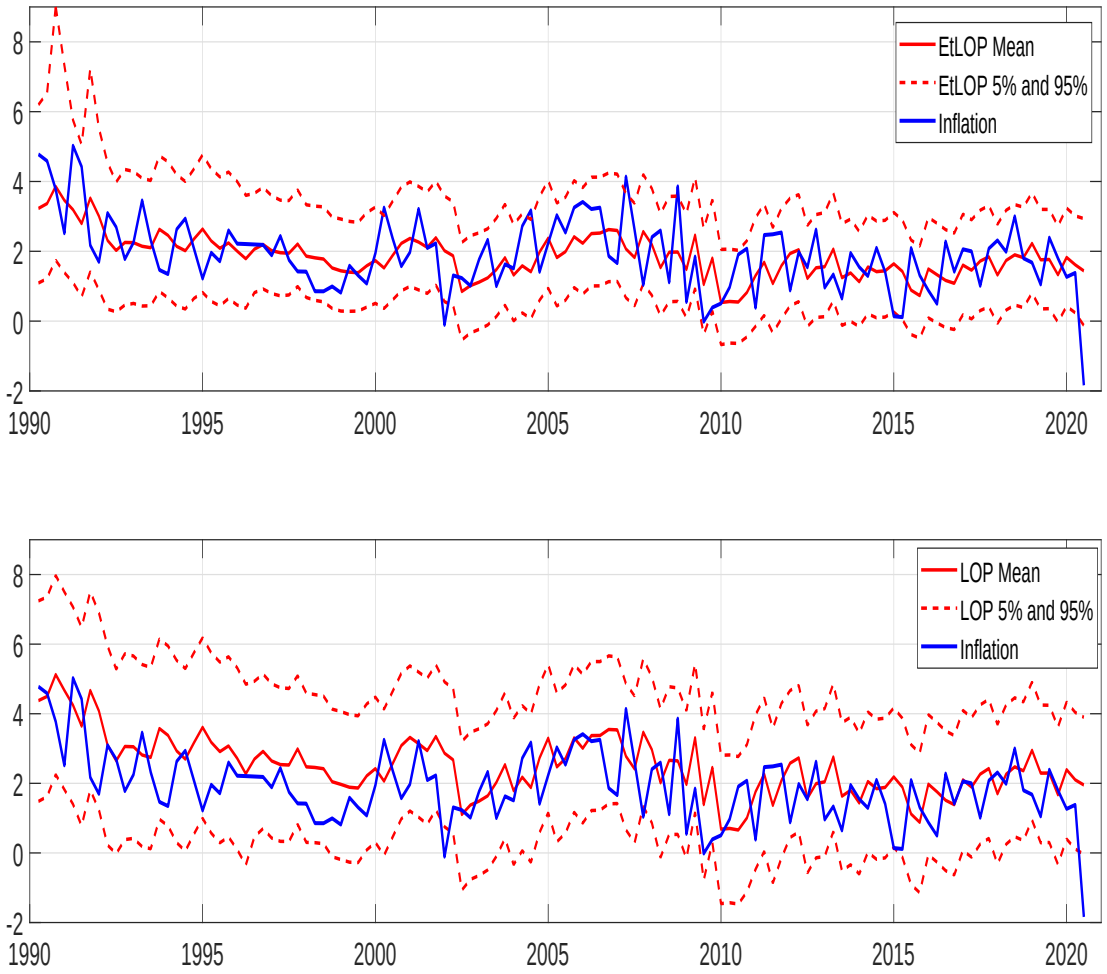
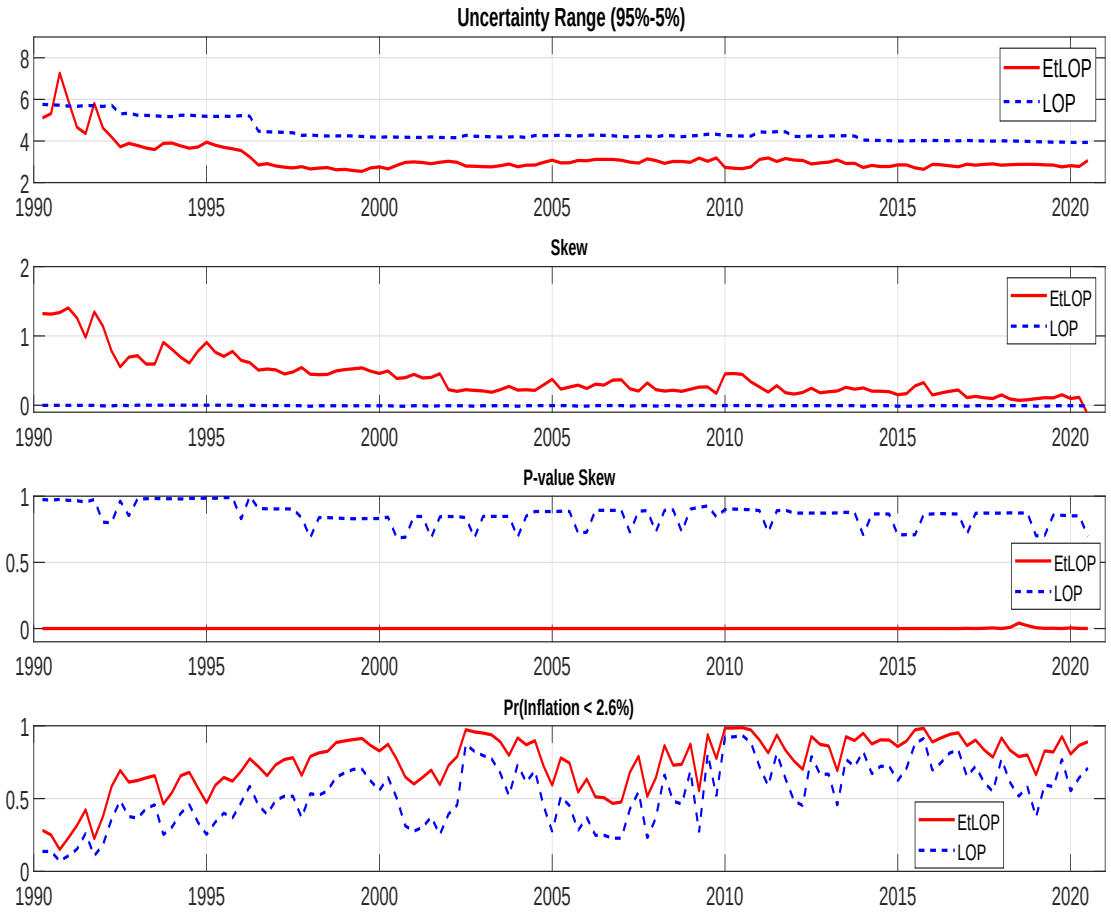


Figure A2.7: Uncertainty Range, Skew and Pr(Inflation$< 2.6\%$), of EtLOP and LOP, $h = 2$, 1990:1 to 2020:2



Notes: Skew is defined as: $s_0 = \sqrt{n(n-1)}s_1/n-2$. Where $n = 5000$, $s_1 = (1/n)\sum_{i-1}^{n}(x_i-\overline{x})^3/\left[\sqrt{(1/n\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^3$ and x_i are the 5000 forecast density iterates at each evaluation period.

Figure A2.8: RMSFE & CRPS, EtLOP & LOP, $h = 2$, 1990:1 to 2020:2

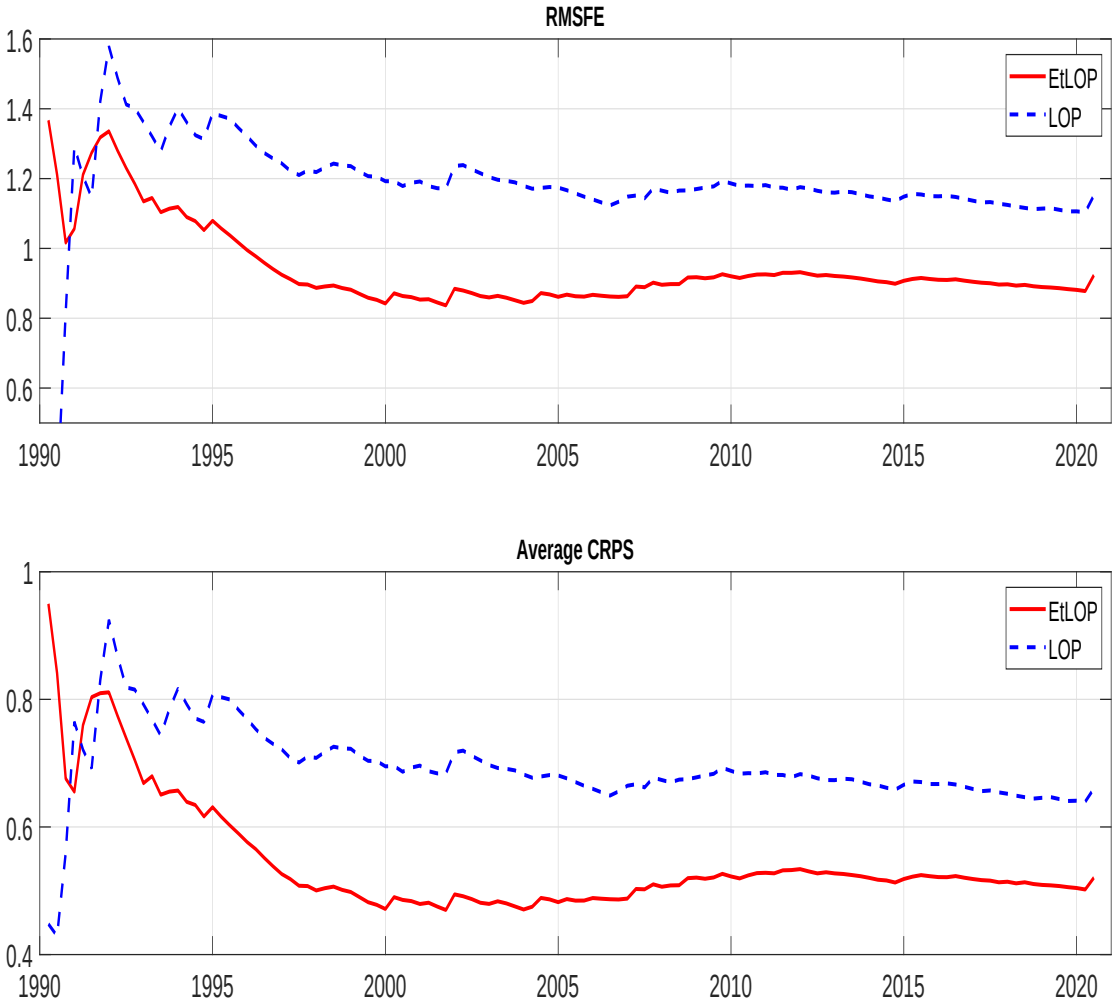
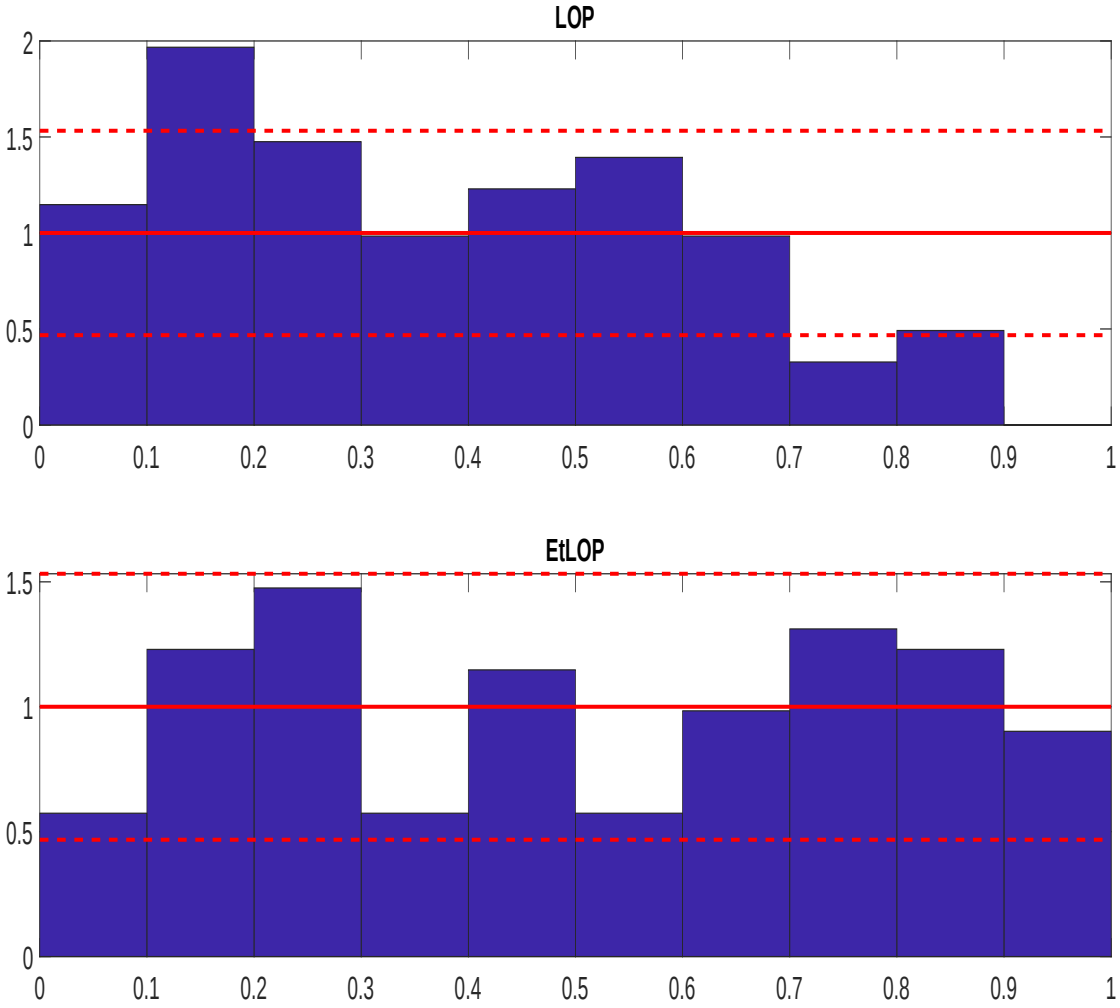
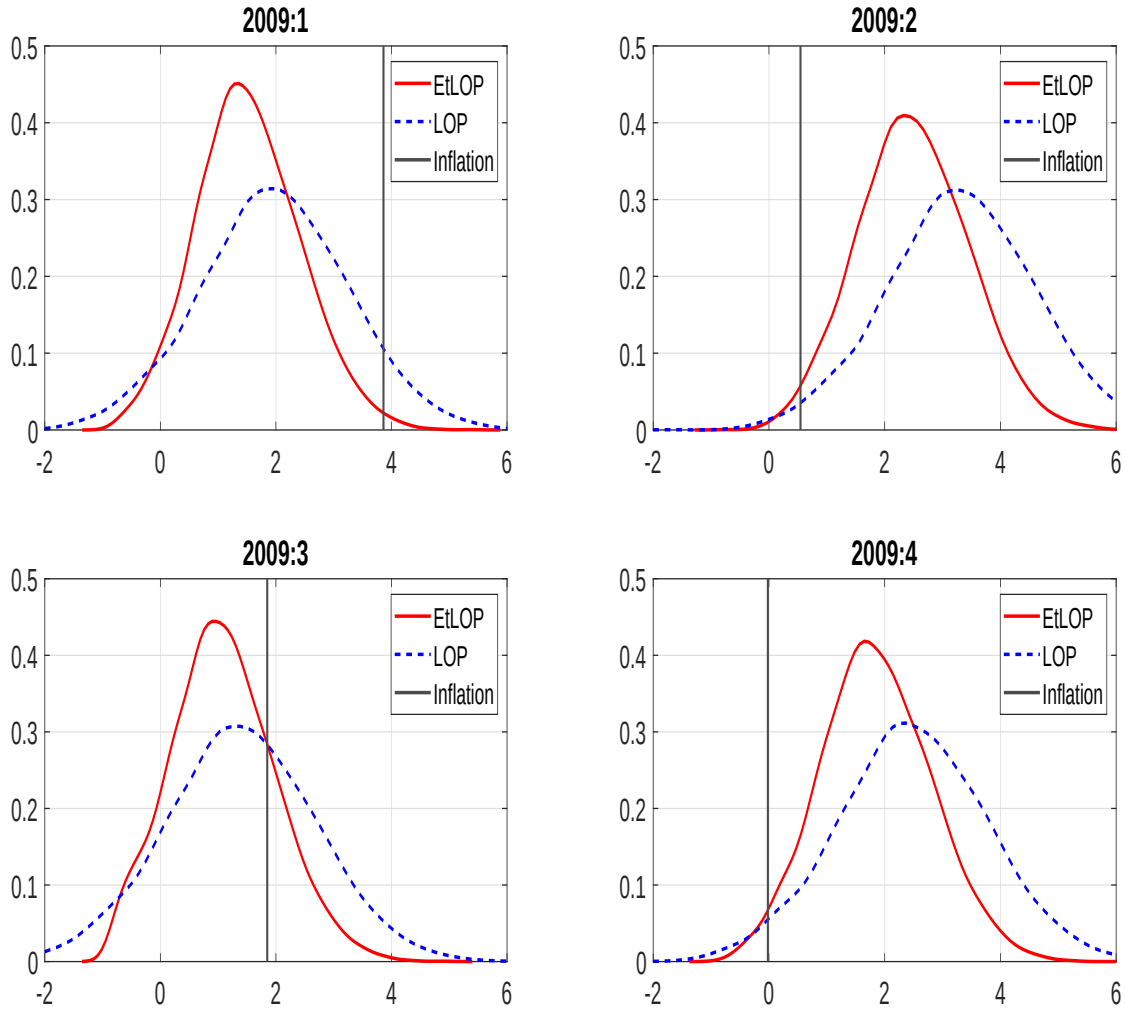


Figure A2.9: PITS, LOP and EtLOP, $h = 2$, 1990:1 to 2020:2



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 A2.10: Forecast Densities for 2009, EtLOP and LOP, $h = 2$



Figures for $h = 3$

Figure A2.11: EtLOP and LOP Forecasts, $h = 3$, 1990:1 to 2020:2

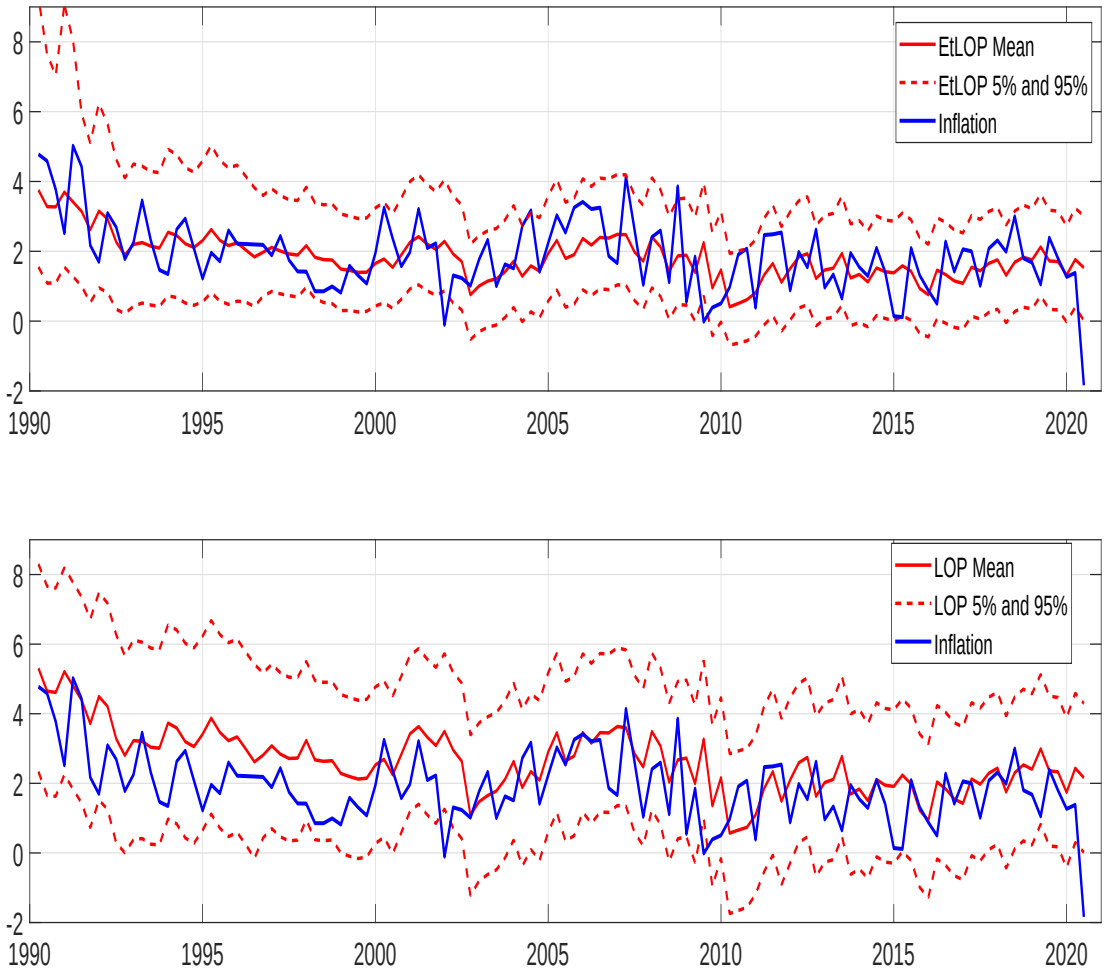
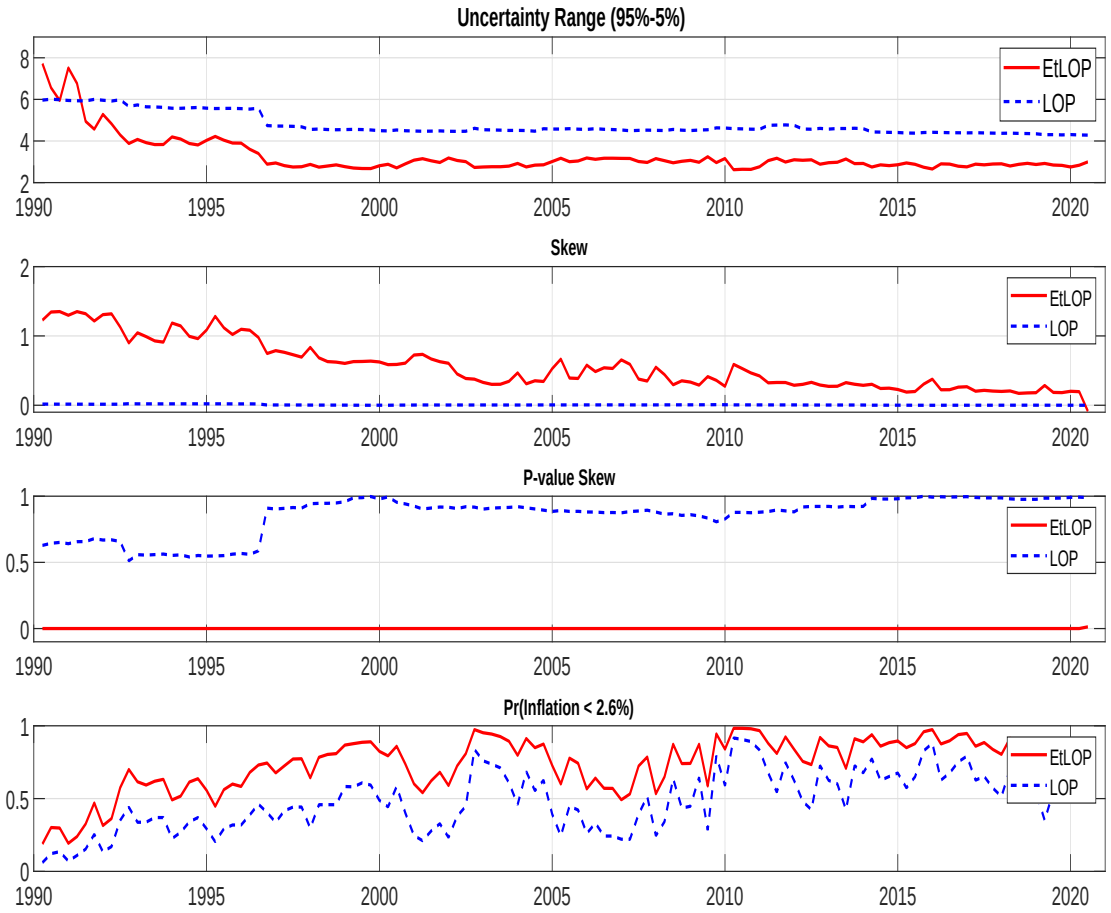


Figure A2.12: Uncertainty Range, Skew and Pr(Inflation$< 2.6\%$), of EtLOP and LOP, $h = 3$, 1990:1 to 2020:2



Notes: Skew is defined as: $s_0 = \sqrt{n(n-1)}s_1/n-2$. Where $n = 5000$, $s_1 = (1/n)\sum_{i-1}^{n}(x_i-\overline{x})^3/\left[\sqrt{(1/n\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^3$ and x_i are the 5000 forecast density iterates at each evaluation period.

Figure A2.13: RMSFE & CRPS, EtLOP & LOP, $h = 3$, 1990:1 to 2020:2

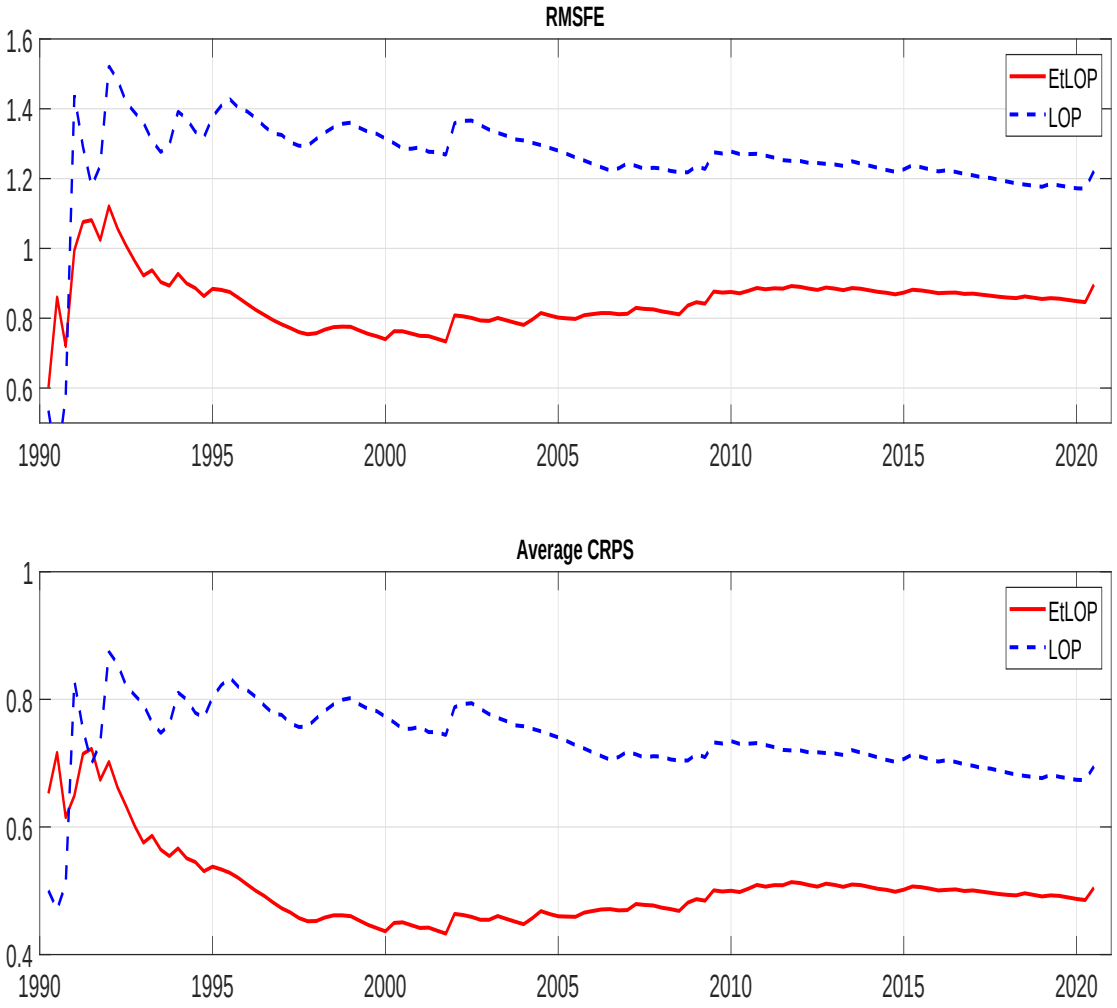
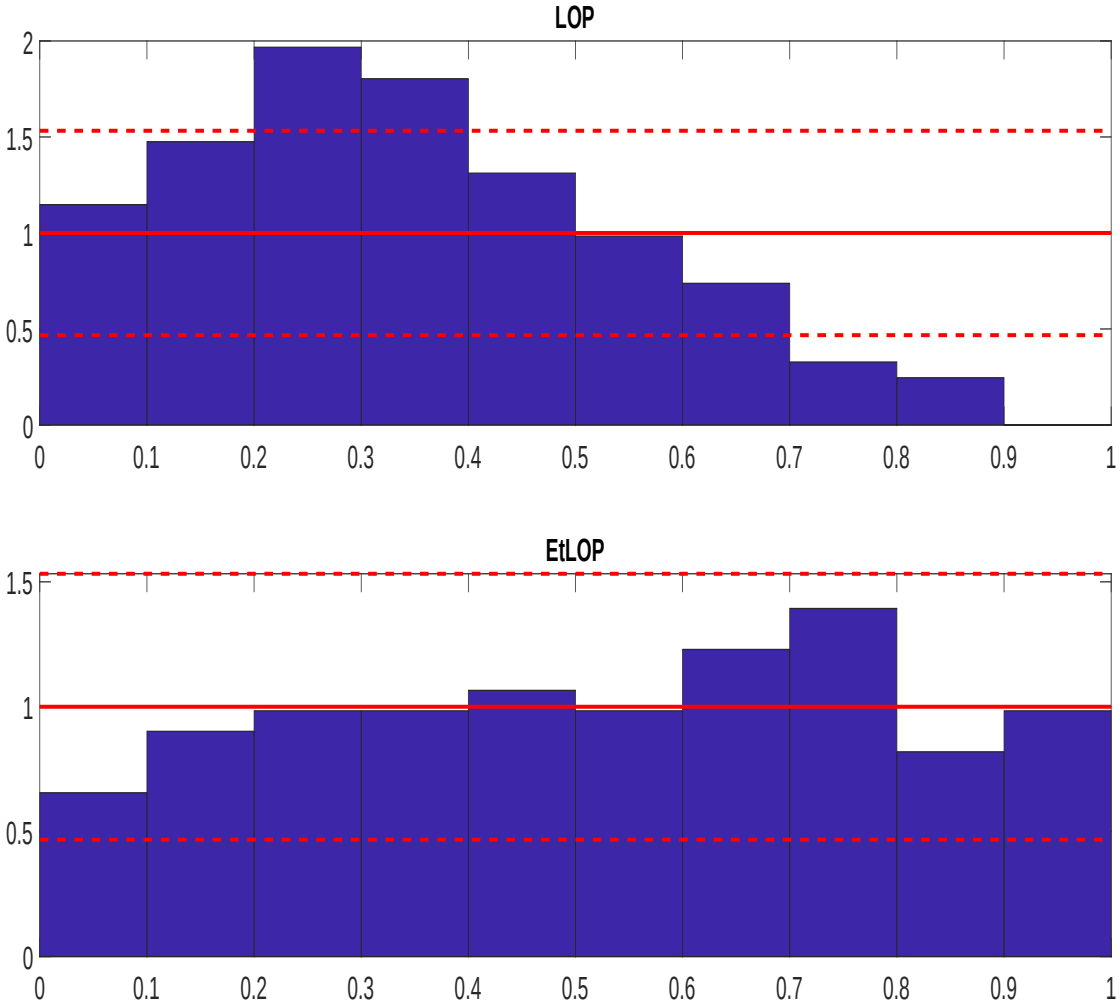
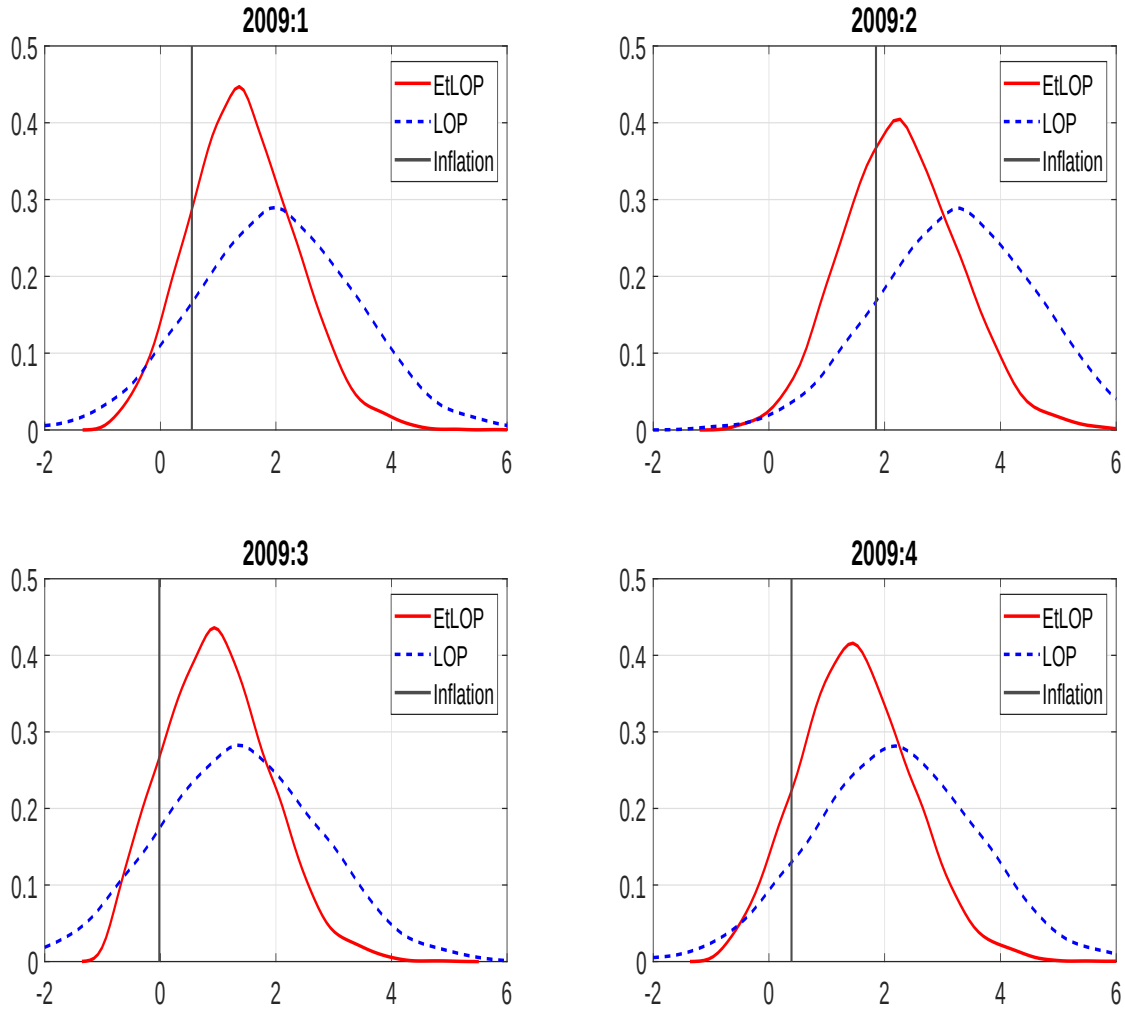


Figure A2.14: PITS, LOP and EtLOP, $h = 3$, 1990:1 to 2020:2



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 A2.15: Forecast Densities for 2009, EtLOP and LOP, $h = 3$



Appendix 3: Unobserved Components with Stochastic Volatility (UCSV)

Gauging the predictive accuracy of candidate methods relative to the UCSV approach has become standard within the inflation forecasting literature. With a focus on density forecasts, it is convenient to adopt the Bayesian approach described in Chan and Song (2018). 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, 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:

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

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. We draw 50,000 iterates, with 5,000 burnin.