Forecast Uncertainty, Disagreement, and Linear Pools of Density Forecasts

Malte Knüppel — Deutsche Bundesbank
Fabian Krüger — Heidelberg University

DEUTSCHE BUNDESBANK EUROSYSTEM
UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Abstract

In many empirical applications, a combined density forecast is constructed using the linear pool which aggregates several individual density forecasts. We analyze the linear pool's implications concerning forecast uncertainty in a mean-variance prediction space framework. Our theoretical results show that, if the variance predictions of the individual density forecasts are unbiased, the well-known 'disagreement' component of the linear pool exacerbates the upward bias of the linear pool's variance prediction. Moreover, we find that disagreement has no predictive content for the true forecast uncertainty under conditions which can be empirically relevant. These findings suggest to remove the disagreement component from the linear pool. The resulting centered linear pool tends to outperform the linear pool in empirical applications based on stochastic volatility models for macroeconomic time series and stock returns.

Introduction

- Density forecasts are an important tool for measuring and communicating forecast uncertainty. They are now widely used in macroeconomics, most popularly in 'fan charts' issued by many central banks
- Following [5] and others, many authors consider forecast combinations via so-called *linear pools* with density

$$f_c = \sum_{i=1}^{n} \omega_i\, f_i, \tag{1}$$

where the $\{f_i\}$ are n forecast densities to be combined (taken from n different models, for example) and the $\{\omega_i\}$ are weights with $\sum_{i=1}^{n} \omega_i = 1$.
- Denote the means and variances of the individual models by $\{m_i\}$ and $\{v_i\}$. Then the variance of the linear pool is

$$V_{LP} = \underbrace{\sum_{i=1}^{n} \omega_i v_i}_{\text{Weighted variances}} + \underbrace{\sum_{i=1}^{n} \omega_i (m_i - m_c)^2}_{\text{Disagreement}}, \tag{2}$$

where $m_c = \sum_{i=1}^{n} \omega_i m_i$ is the mean of the combined forecast.
- Equation (2) is a common but very specific way to quantify the combined forecast uncertainty. We analyze its plausibility from a forecasting perspective, focusing on the role of forecaster disagreement
  - Does Equation (2) lead to unbiased variance forecasts on average?
  - Does Equation (2) track the heteroscedasticity in macroeconomic time series?
- Our analysis builds upon earlier work by [4]. In contrast to the latter paper, we focus on mean and variance forecasts. This restriction seems empirically sensible, and allows us to derive more specific results.

Theoretical Analysis

Model for forecasts and realizations

We view the n mean and variance forecasts as random variables whose joint distribution is given by

$$\begin{bmatrix} M \\ V \\ U \end{bmatrix} \Big| \eta \sim \left(\begin{bmatrix} \mu_M \\ \eta \mu_V \\ 0 \end{bmatrix}, \begin{bmatrix} \eta \Sigma_M & 0 & 0 \\ 0 & \eta^2 \Sigma_V & 0 \\ 0 & 0 & \eta \Sigma_U \end{bmatrix} \right).$$

where the column on the right-hand side denotes the expectations, the matrix is the variance-covariance matrix, and U denotes the residuals of the projection

$$Y = (M - \mu_M)' \gamma + U,$$

where Y is the demeaned predictand. The mean forecasts are collected in $M \in \mathbb{R}^n$, which is an $n-$variate continuous random vector. Since $\mathbb{E}[Y] = 0$, μ_M denotes the biases of the mean forecasts. $V \in \mathbb{R}^n_+$ is an n-variate random vector with positive continuous entries denoting the variance forecasts. The key measure for realized forecast uncertainty is given by squared forecast errors

$$S = (Y - \hat{Y})^2,$$

where $\hat{Y}$ is a generic mean forecast. For a correctly specified forecast model, we have that $\mathbb{E}[V] = \mathbb{E}[S]$. Furthermore, if the process for Y is heteroscedastic, we expect V to be highly correlated with S.

Main assumptions considered

A1 The individual variance forecasts V are unconditionally unbiased, i.e. $\mathbb{E}[V_i] = \mathbb{E}[Y - M_i]^2, i = 1, \ldots, n$.

A2 The mean forecasts are unbiased, i.e. $\mu_M = [0, \ldots, 0]'$.

A3 The combination weights $\omega = [\omega_1, \ldots, \omega_n]'$ are positive.

A4 The combination weights have been chosen such that $\mathbb{E}[S]$ is minimized, subject to the constraint of adding to one (Bates-Granger weights).

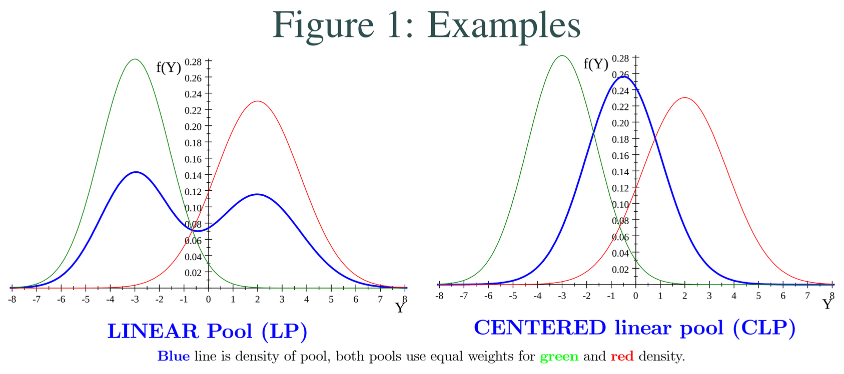
A5 The joint distribution of M and U conditional on η is normal.

A6 It holds that $\Sigma_V = 0$.

Combination methods

We consider the following combination methods:
- The *linear pool (LP)* as introduced in Equation (1).
- The *centered linear pool (CLP)*, obtained by re-centering the n forecast densities at the combined mean m_c, and then applying the LP. The CLP forecast also has mean m_c, but variance $V_{CLP} = \sum_{i=1}^{n} \omega_i v_i$. Hence, the *CLP's variance does not depend on disagreement*. Figure 1 provides an example for both pools.

Figure 1: Examples



Blue line is density of pool, both pools use equal weights for green and red density.

Biases in the variance forecasts of linear pools

Theorem 1. *Under A1 and A3, it holds that*

$$\mathbb{E}[S] = \underbrace{\mathbb{E}[\omega' V] + \mathbb{E}[D]}_{\mathbb{E}[V_{LP}]} - 2\,\mathbb{E}[D] = \underbrace{\mathbb{E}[\omega' V]}_{\mathbb{E}[V_{CLP}]} - \mathbb{E}[D]$$

- Theorem 1 quantifies the relationship between the LP's and CLP's expected squared forecast error, $\mathbb{E}[S]$, and their average variance forecasts $\mathbb{E}[V_{LP}]$ and $\mathbb{E}[V_{CLP}]$.
- The variance forecasts of the LP and the CLP *systematically deviate from* S, with their underconfidence, i.e. their upward biases depending on expected disagreement only.
- The upward bias of the LP is *twice as large* as the upward bias of the CLP.

Theorem 2. *Consider an arbitrary joint distribution of M, V and Y. Under A1 and A3, both the linear pool and the centered linear pool are underconfident, and the centered linear pool is less underconfident than the linear pool.*

- Theorems 1 and 2 have the same qualitative interpretation: If the individual variances are unbiased and the combination weights are positive, then both the LP and the CLP are underconfident. However, Theorem 1 exploits the joint distribution of forecasts and realizations to make a precise quantitative statement on the pools' underconfidence.

Disagreement encompassed by weighted individual variance forecasts

Consider the following linear regressions:

$$S = a_0 + a_1 D + a_2\, \omega' V + \text{error};$$
$$S = b_1 D + b_2\, \omega' V + \text{error}; \tag{3}$$

The regressions are similar to Mincer-Zarnowitz regressions considered in the literature on forecasting financial volatilities. We have the following result.

Theorem 3. *Assume that A2, A4, A5 and A6 hold. Then, $a_1 = b_1 = 0$, i.e. $\omega' V$ encompasses disagreement in the prediction of S.*

Empirical Case Studies

Motivation

- Theorems 1 & 2 state that, if individual variance forecasts are unbiased (or upward biased), LP's variance is too large. Theorem 3 presents conditions under which disagreement is encompassed by weighted variances.
- Theoretical results suggest to remove disagreement from the LP's variance. *We test this strategy empirically*, by comparing the LP to the CLP. Note: Same mean forecast, but different variance specifications.
- In line with the theoretical setup, we use an evaluation tool that focuses on means & variances only: the *Dawid-Sebastiani score*, given by

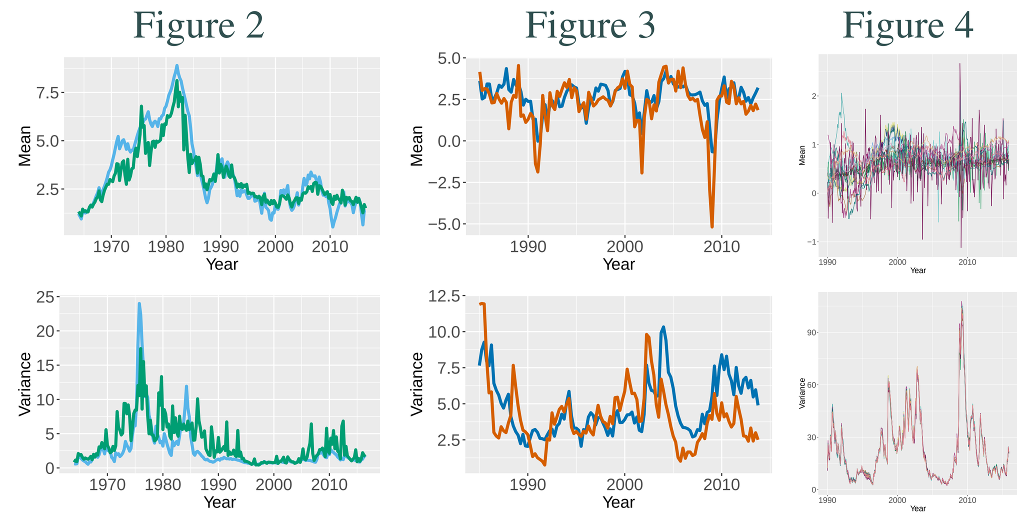
$$DSS(m, v, y) = \log v + \frac{(y - m)^2}{v}, \tag{4}$$

where m and v are mean and variance forecasts, and y is the realization. The lower the score, the better is the forecast.

Data sets

We consider three case studies, with forecasts generated by Bayesian models with stochastic volatility.
- **Case study #1: Inflation.** Forecasts from the univariate unobserved component model with stochastic volatility by [1] and from the bivariate unobserved component model with trends and cycles by [2]. Quarterly US data (1946-2016), forecasts are 1,2,4 and 8 quarters ahead. See Figure 2 for an example.
- **Case study #2: Macro nowcasts.** Following [6], we consider nowcasts of four US macro variables (GDP, INF, UNE, TBI), and employ two nowcasting methods: First, a statistical model using intra-quarterly data. Second, a subjective nowcast from the Survey of Professional Forecasters. See Figure 3 for an example.
- **Case study #3: Excess returns.** We combine 15 Bayesian univariate regression models (with stochastic volatility) for excess returns (c.f. [8] and [7]). US data, one-month ahead forecasts. See Figure 4 for an example.

Figure 2 — Figure 3 — Figure 4



Top row: Time series of forecast means. Bottom row: Time series of forecast variances.
Left column: Inflation, $h = 4$. Middle column: GDP nowcasts. Right column: Excess returns.

Results

- All empirical results consistent with Theorems 1 & 2.
  - Average variance of CLP exceeds mean of squared forecast errors ⇒ average variance of LP exceeds mean of squared forecast errors even more.
  - Size of disagreement determines differences between LP and CLP. Large disagreement observed for nowcasts (up to 64% of LP's variance); almost no disagreement for returns.
- Results on Theorem 3 more mixed: Disagreement encompassed by average variance in six out of nine cases. All three violations occur for nowcast case study.
- Dawid-Sebastiani scores (Table 1): CLP attains smaller (i.e. better) scores than LP in all nine cases. Differences are significant in four cases.

Table 1

	Inflation				Nowcasts				Returns
	$h=1$	$h=2$	$h=4$	$h=8$	GDP	INF	TBI	UNE	
DSS (LP)	1.047	1.308	1.617	2.047	2.129	0.863	-3.878	-3.147	3.776
DSS (CLP)	1.041	1.299	1.604	2.044	2.098	0.825	-4.017	-3.691	3.776
DM test stat	1.44	2.24	2.47	0.48	0.65	1.10	3.97	5.94	0.22

Summary

We analyze whether linear pools of density forecasts imply accurate statements about forecast uncertainty. We find that, in many relevant settings, they do not: The forecast variance in (2) is upward biased, and does not track the heteroscedasticity in macroeconomic time series under conditions which often appear to be empirically relevant. Removing the disagreement component is a simple first step towards improving the variance specification.

References

[1] Joshua C.C. Chan. Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, 176(2):162 – 172, 2013.

[2] Joshua CC Chan, Gary Koop, and Simon M Potter. A bounded model of time variation in trend inflation, nairu and the phillips curve. *Journal of Applied Econometrics*, 3(31):551–565, 2016.

[3] A Philip Dawid and Paola Sebastiani. Coherent dispersion criteria for optimal experimental design. *Annals of Statistics*, pages 65–81, 1999.

[4] Tilmann Gneiting and Roopesh Ranjan. Combining predictive distributions. *Electronic Journal of Statistics*, 7:1747–1782, 2013.

[5] Stephen G Hall and James Mitchell. Combining density forecasts. *International Journal of Forecasting*, 23:1–13, 2007.

[6] Fabian Krüger, Todd E. Clark, and Francesco Ravazzolo. Using entropic tilting to combine bvar forecasts with external nowcasts. *Journal of Business & Economic Statistics*, pages 1–48, forthcoming.

[7] David E Rapach, Jack K Strauss, and Guofu Zhou. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23:821–862, 2010.

[8] Ivo Welch and Amit Goyal. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21:1455–1508, 2008.