Forecast Uncertainty, Disagreement, and the Linear Pool*

Malte Knüppel   Fabian Krüger
Deutsche Bundesbank   Karlsruhe Institute of Technology

April 8, 2021

Abstract

The linear pool is the most popular method for combining density forecasts. We analyze its implications concerning forecast uncertainty, using a new framework that focuses on the means and variances of the individual and combined forecasts. Our results show that, if the variance predictions of the individual forecasts are unbiased, the well-known 'disagreement' component of the linear pool exacerbates the upward bias of its variance prediction. This finding suggests the removal of the disagreement component from the linear pool. The resulting centered linear pool outperforms the linear pool in simulations and an empirical application to inflation.

Keywords: Density forecasting; Forecast evaluation; Inflation forecasting

JEL classification: C32, C53, E37.

1 Introduction

There is a growing recognition that measuring forecast uncertainty matters for economic policy. For example, many central banks have followed the Bank of England's lead in publishing probabilistic forecasts of inflation and related variables; see Franta et al. (2014, Table 1). Similarly, Manski (2015) calls for systematic measurement and communication of uncertainty in official statistics. In statistical terms, confronting uncertainty suggests to issue density forecasts, rather than traditional point forecasts. An immediate question is how to make 'good' density forecasts. In light of many available forecasting methods and data sources, it is often a combination of several individual forecasts, rather than a single forecast, that is considered for this purpose.

While various combination methods have been proposed, the comprehensive survey by Aastveit et al. (2019, p. 20) argues that "[..] most applications still focus on the linear opinion pool [..]". Given a set of n individual density forecasts $f_1, \ldots, f_n$, the linear opinion pool, or simply linear pool (LP), is computed as $f_{lp} = \sum_{i=1}^{n} \omega_i f_i$, where $\{\omega_i\}_{i=1}^{n}$ are the combination weights (Stone, 1961). The concept of the LP is, for instance, employed to produce aggregate probability distributions in the Survey of Professional Forecasters (SPF) conducted by the Federal Reserve Bank of Philadelphia and, in similar form, by the European Central Bank.

In the present paper, we analyze the LP's implications concerning forecast uncertainty. For this purpose, we consider the joint behavior of mean forecasts, variance forecasts, and the target variable in terms of their first two moments. We derive several new results, focusing on the LP's 'disagreement' component which quantifies differences between the mean forecasts of the individual densities. While disagreement has received considerable attention as a potential proxy for economic uncertainty (e.g. Dovern et al., 2012), its role turns out to be problematic in the context of the LP.

First, we show that if the individual density forecasts are variance-unbiased (as defined in Assumption 1 below), the LP's variance is upward biased by twice the expected disagreement. This result sharpens and quantifies a related qualitative finding by Gneiting and Ranjan (2013) concerning the LP's overdispersion. Second, under a set of conditions including joint normality, we show that, within the LP, disagreement has no predictive content for squared forecast errors, thereby violating a desideratum of good uncertainty forecasts. Third, we argue that choosing combination weights for the LP entails a trade-off since the weights

affect both the mean forecast and the variance forecast of the LP. Weights that are optimal for the mean are typically not optimal for the corresponding variance.

The first two results indicate that disagreement harms the LP's variance forecasts, which suggests that a variance specification without disagreement should be considered. We therefore propose the centered linear pool (CLP), a simple modification of the LP which achieves this goal and alleviates the trade-off between mean-optimal and variance-optimal weights. We illustrate our results and investigate the performance of the CLP in simulations and an empirical application to inflation forecasts.

The remainder of this paper is structured as follows: Section 2 derives properties of an optimal variance forecast. These properties form a benchmark for evaluating any variance forecast, including that of the LP. Section 3 presents a baseline example which motivates our analysis of the LP and previews our main results. Section 4 presents more general results on bias in the LP's variance forecast, and on the comovement between disagreement and squared forecast errors. Sections 5 and 6 contain simulation and empirical results, and Section 7 concludes.

2 Properties of an optimal variance forecast

We first derive simple yet crucial properties of an optimal variance forecast. As a measure of forecast accuracy, we consider the Dawid and Sebastiani (1999) scoring rule which depends only on the mean and variance of a forecast distribution, in line with the focus of our analysis. The Dawid-Sebastiani score (DSS) equals the negative logarithmic score (log score) of a Gaussian forecast density $f_{\mathcal{N}}$ with mean m and variance v, i.e.

$$
\begin{aligned}
DSS(y; m, v) &= -\log(f_{\mathcal{N}}(y; m, v)) \\
&= 0.5\log(2\pi) + 0.5\log(v) + \frac{(y-m)^2}{2\,v},
\end{aligned}
\tag{1}
$$

where y denotes the realization of the target variable.[1] A smaller score corresponds to a better forecast. Consider forecasting the parameters m and v of a random variable Y, conditional

---

[1]In the literature, the equivalent score $\log(v) + \frac{(y-m)^2}{v}$ is more common. However, we stick to the variant in equation (1) for better comparability with the log score.

on some information set $\mathcal{I}$. Then the expectation $\mathbb{E}\left[DSS(Y;m,v)|\mathcal{I}\right]$ is minimized by setting

$$m = \mathbb{E}\left[Y|\mathcal{I}\right], \tag{2}$$

$$v = \mathbb{E}\left[S|\mathcal{I}\right], \tag{3}$$

where $S = (Y - m)^2$ denotes the squared forecast error. Hence the DSS rewards forecast densities f (that need not be Gaussian) with a correctly specified conditional mean forecast m and conditional variance forecast v, where the latter depends on the former.[2]

In the terminology of Gneiting and Raftery (2007), the DSS is a proper but not a strictly proper scoring rule. This is because the DSS focuses on the first two moments of f only. A density forecast with misspecified higher order moments may hence perform equally well as the ideal density forecast. For example, suppose that the ideal forecast density is skewed with zero mean, unit variance and nonzero median. Then stating a standard normal forecast density yields the same expected DSS as the ideal density. Hence, under the DSS, forecasters have no incentive to correctly model the third or higher moments of the predictand Y. The situation is different for other scoring rules like the log score, the continuous ranked probability score (Matheson and Winkler, 1976), or certain weighted scoring rules (e.g. Lerch et al., 2017), which are strictly proper and thus incentivize forecasters to model the entire forecast density correctly. That said, the optimality conditions at (2) and (3) are not specific to the DSS, but are shared by all strictly proper scoring rules.[3]

In a multi-observation setup, we treat the mean and variance forecasts as random variables M and V. Variation in M and V may be informative (resulting from variation in the conditioning information set $\mathcal{I}$) or not. The optimality condition in equation (3) has two main implications in this context: First, V and S should be equal on average, i.e.

$$\mathbb{E}[S] = \mathbb{E}[V]. \tag{4}$$

Clements (2014) uses this equality in an empirical analysis of subjective probability distributions. We refer to the unconditional expectation $\mathbb{E}[S]$ as the mean squared forecast error

---

[2]The studies listed in Krüger et al. (2020, Table 1 of the Online Supplement) use the log score to evaluate a Gaussian approximation to their forecast densities. This approach is equivalent to applying the DSS to the forecast densities.

[3]To see this point, note that each strictly proper scoring rule is optimized by stating the true forecast density $f(Y|\mathcal{I})$. The forecast mean and variance implied by this density are given by (2) and (3).

(MSFE). Second, from the law of total covariance, the requirement that $V = \mathbb{E}[S|\mathcal{I}]$ implies that

$$\mathbb{C}ov[V, S] = \mathbb{V}\left[\mathbb{E}[S|\mathcal{I}]\right] \geq 0, \tag{5}$$

where $\mathbb{C}ov[\bullet]$ and $\mathbb{V}[\bullet]$ denote covariance and variance, respectively. The inequality is strict only in the presence of predictable heteroskedasticity, because in this case $\mathbb{E}[S|\mathcal{I}]$ varies with $\mathcal{I}$.

From an empirical perspective, our setup seems appropriate at least for macroeconomic time series which motivate the present paper. In particular, our focus on the *conditional* forecast mean M and variance V (given an information set $\mathcal{I}$) allows for considerable flexibility through the joint distribution of M, V and the outcome Y. Empirical evidence by Carriero et al. (2020) indicates that this flexibility is sufficient for capturing asymmetries in the unconditional distribution of Y that have been emphasized by Adrian et al. (2019) and others. As a robustness check, we further present results on the log score in addition to the DSS in our Monte Carlo simulations and empirical analysis.

3   The linear pool's variance forecast: Baseline example

We next provide a simple example in which the LP violates one or both of the implications of an optimal variance forecast stated in equations (4) and (5). Consider a variable Y given by $Y = X_1 + X_2 + U$, where X_1, X_2 and U are distributed as

$$\begin{bmatrix} X_1 \\ X_2 \\ U \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho\sigma_X^2 & 0 \\ \rho\sigma_X^2 & \sigma_X^2 & 0 \\ 0 & 0 & \sigma_U^2 \end{bmatrix} \right).$$

Forecaster 1 only observes X_1, and forecaster 2 only observes X_2. Both forecasters aim to predict the distribution of Y and state the correct forecast distribution given their information sets. Each forecaster $i \in \{1, 2\}$ thus issues a Gaussian forecast density with mean M_i and variance V_i. Table 1 lists the formulas for M_i and V_i, as well as all other relevant formulas for this example. The LP of the two forecasts is given by $f_{lp} = \omega_1 f_1 + (1 - \omega_1)f_2$, where f_{lp} is the density of the combined forecast, f_1 and f_2 are the individual densities, and

$0 \leq \omega_1 \leq 1$ is the weight on the first forecast. Here and throughout the paper, we take the combination weights to be fixed, non-stochastic quantities. We denote the mean and variance of this combined density by M_{lp} and V_{lp}, respectively.

As shown in Table 1, both forecasters fulfill the requirements mentioned in Section 2. First, their variance forecasts V_i and squared forecast errors S_i are equal in expectation. Second, the covariance between each variance forecast and the corresponding squared error $\mathbb{C}ov[V_i, S_i]$ is equal to the variance of the expected (conditional) squared forecast error $\mathbb{V}[\mathbb{E}[S_i]] = \mathbb{V}[\mathbb{E}[S_i|\mathcal{I}]]$. Due to the homoskedasticity in this example, both terms are equal to zero.

The LP's variance forecast is of the form $V_{lp} = a + D$, where a is a constant and D is the well-known measure of disagreement between the two point forecasts (see e.g. Wallis 2005). Strikingly, the LP's variance V_{lp} fails the requirement (4). The LP's expected variance, $\mathbb{E}[V_{lp}]$, exceeds its MSFE, $\mathbb{E}[S]$, by $2\mathbb{E}[D]$. The LP can therefore be labeled underconfident, and the disagreement term $D \geq 0$ contributes to the LP's underconfidence. In addition, if $\omega_1 = 0.5$, the LP's variance V_{lp} has no predictive content for its squared forecast error, such that requirement (5) is not fulfilled. It can be shown that with $\omega_1 = 0.5$, D is not only uncorrelated with S, but also independent of S (see Online Appendix, Section I). Note that equal weighting, here corresponding to $\omega_1 = 0.5$, is a popular choice in practice, and minimizes the MSFE of the combined mean forecast in the present example. For other choices of ω_1, the relation between D and S depends on ρ, σ_X^2 and σ_U^2, but often implies weak correlation between D and S. In the case $\omega_1 = 0.5$, disagreement can be regarded as a noise term which deteriorates the LP's variance forecast.

The example illustrates that the LP's variance exceeds the average variance of the individual forecasts. Of course, this is also problematic if the individual variances are too large instead of unbiased. By contrast, the increase in variance due to linear pooling can be desirable if the individual variances are too small. Given the multitude of empirical settings in which the LP is used (featuring different scientific disciplines, data sources, and individual models), each of these situations can be empirically relevant. The aim of the present paper is to analyze the LP in the situation where the individual variances are unbiased. This situation seems especially relevant if each individual forecast is based on a statistical model, such that variance unbiasedness is implicitly pursued in the model fitting process.

Object	Formula		
Individual forecasters			
Mean forecasts	M_i	$=$	$(1+\rho)X_i$
Variance forecasts	V_i	$=$	$V = (1-\rho^2)\sigma_X^2 + \sigma_U^2$
Squared error of M_i	S_i	$=$	$(-\rho X_i + X_j + U)^2,\ i \neq j$
MSFE of M_i	$\mathbb{E}[S_i]$	$=$	$(1-\rho^2)\sigma_X^2 + \sigma_U^2$
Covariance of V_i and S_i	$\mathbb{C}ov[V_i, S_i]$	$=$	0
Linear pool			
Mean forecast	M_{lp}	$=$	$\omega_1 M_1 + (1-\omega_1)M_2$
Disagreement	D	$=$	$\omega_1(M_{lp}-M_1)^2 + (1-\omega_1)(M_{lp}-M_2)^2$
		$=$	$\omega_1(1-\omega_1)(M_1-M_2)^2$
Variance forecast	V_{lp}	$=$	$V + D$
Squared error of M_{lp}	S	$=$	$\left((1-\omega_1(1+\rho))X_1 + (\omega_1(1+\rho)-\rho)X_2 + U \right)^2$
Expected disagreement	$\mathbb{E}[D]$	$=$	$2\omega_1(1-\omega_1)(1-\rho^2)(1+\rho)\sigma_X^2$
Expected variance forecast	$\mathbb{E}[V_{lp}]$	$=$	$(1-\rho^2)\sigma_X^2 + \sigma_U^2 + 2\omega_1(1-\omega_1)(1-\rho^2)(1+\rho)\sigma_X^2$
		$=$	$V + \mathbb{E}[D]$
MSFE of M_{lp}	$\mathbb{E}[S]$	$=$	$(1-\rho^2)\sigma_X^2 + \sigma_U^2 - 2\omega_1(1-\omega_1)(1-\rho^2)(1+\rho)\sigma_X^2$
		$=$	$V - \mathbb{E}[D]$
Covariance of D and S	$\mathbb{C}ov[D, S]$	$=$	$2\omega_1(1-\omega_1)(2\omega_1-1)^2(1-\rho)^2(1+\rho)^4\sigma_X^4$
Covariance of V_{lp} and S	$\mathbb{C}ov[V_{lp}, S]$	$=$	$\mathbb{C}ov[D, S]$

Table 1: Formulas for the baseline example. Moments of linear pool follow from aggregation of individual forecast densities according to $f_{lp} = \omega_1 f_1 + (1-\omega_1)f_2$. MSFE denotes the mean squared forecast error.

4 The linear pool's variance forecast: General framework

We now consider the more general situation where n density forecasts f_i with corresponding mean and variance forecasts $\{m_i, v_i\}_{i=1}^n$ are available, where the index i denotes an individual forecast. The LP determines the combined density as $f_{lp} = \sum_{i=1}^n \omega_i f_i$, implying the mean and the variance forecast

$$m_{lp} = \sum_{i=1}^n \omega_i m_i \tag{6}$$

$$v_{lp} = \sum_{i=1}^n \omega_i v_i + d \tag{7}$$

with $d = \sum_{i=1}^n \omega_i (m_i - m_{lp})^2$, such that $\sum_{i=1}^n \omega_i v_i$ is the weighted average variance component and d is the disagreement component of the LP's variance forecast. For the case of equal weights, this formula appeared, for instance, in Lahiri et al. (1988) and Wallis (2005), and its general form is well-known in the context of moments of mixture distributions (see Frühwirth-Schnatter, 2006, chap. 1.2.4). We generally restrict attention to weights $\{\omega_i\}_{i=1}^n$ that are nonnegative and sum to one. This ensures that the density f_{lp} is well-defined and, in particular, that the LP's variance v_{lp} is nonnegative. Furthermore, we mostly take the mean specification in (6) as given, and investigate the properties of the variance specification in (7) conditional on (6).

We will also consider a simple modification of the LP, the centered linear pool (CLP), with $m_{clp} = m_{lp}$ and

$$v_{clp} = \sum_{i=1}^n \omega_i v_i = v_{lp} - d. \tag{8}$$

Hence the CLP has the same mean forecast as the LP, but its variance forecast does not contain the disagreement term. Clements (2018) suggests using the CLP variance in the context of combining survey forecasts, on the grounds that it equals the expected variance of a randomly selected individual forecaster. Denoting a density f_i with mean m_i and variance v_i by $f_i(m_i, v_i)$, the CLP is constructed as $f_{clp} = \sum_{i=1}^n \omega_i f_i(m_{clp}, v_i)$. Thus, each individual density is relocated such that its mean equals $m_{clp} = m_{lp}$ instead of m_i before

being combined. If the individual densities are symmetric, then the CLP density is symmetric as well. This behavior is distinct from the LP density which may be asymmetric even if the individual densities are symmetric. We elaborate on this aspect in Section II of the Online Appendix.

Equations (6) to (8) are formulated in terms of given mean and variance forecasts $\{m_i, v_i\}_{i=1}^n$. From an ex-ante perspective, these objects are random variables which we denote by $\{M_i, V_i\}_{i=1}^n$. We proceed using the ex-ante perspective.

We next relate the squared error of the individual forecast i, $S_i = (Y - M_i)^2$, to the squared error of the combined mean forecast M_{lp}, $S = (Y - M_{lp})^2$. By definition, the weighted sum of the individual squared forecast errors equals

$$
\begin{aligned}
\sum_{i=1}^n \omega_i S_i &= \sum_{i=1}^n \omega_i (Y - M_i)^2 \\
&= (Y - M_{lp})^2 + \sum_{i=1}^n \omega_i (M_i - M_{lp})^2 \\
&= S + D
\end{aligned}
\tag{9}
$$

where $D = \sum_{i=1}^n \omega_i (M_i - M_{lp})^2$ denotes disagreement among the point forecasts. The second equality follows from subtracting and adding M_{lp}, and from noting that $\sum_{i=1}^n \omega_i (M_i - M_{lp})$ equals zero. Engle (1983, eq. 11) and Lahiri et al. (2015, eq. 3) state the identity at (9) for the case of equal weights; Page (2007, 2018) refers to it as the 'prediction diversity theorem' and discusses its broader implications. Equation (9) refines the inequality $S \leq \sum_{i=1}^n \omega_i S_i$ that has been emphasized by McNees (1992), Manski (2010), Lichtendahl Jr et al. (2013) and others.

The following assumption relates V_i to its ex-post counterpart S_i.

Assumption 1. *The individual variance forecasts $\{V_i\}_{i=1}^n$ are unconditionally unbiased, i.e. they satisfy $\mathbb{E}[V_i] = \mathbb{E}[S_i], i = 1, \ldots, n$, where $S_i = (Y - M_i)^2$ is the squared error of forecast i.*

Note that Assumption 1 imposes the unconditional notion of variance unbiasedness formulated in (4), which is weaker than the DSS optimality condition in (3).

Using Assumption 1, denoting the vector of weights $(\omega_1, \omega_2, \ldots, \omega_n)'$ by ω, and the corresponding vector of the individual variance forecasts $(V_1, V_2, \ldots, V_n)'$ by V, taking expec-

tations of (9) gives

$$\mathbb{E}[S] = \mathbb{E}[\omega'V] - \mathbb{E}[D]. \tag{10}$$

Taking expectations of the ex-ante versions of the variance forecasts (7) and (8) implied by the LP and the CLP yields

$$\mathbb{E}[V_{lp}] = \mathbb{E}[\omega'V] + \mathbb{E}[D] \tag{11}$$

$$\mathbb{E}[V_{clp}] = \mathbb{E}[\omega'V], \tag{12}$$

and equations (10), (11), and (12) directly lead to

Proposition 1. *Under Assumption 1, the LP's and CLP's variance forecasts systematically deviate from* $\mathbb{E}[S]$ *according to*

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

$$= \underbrace{\mathbb{E}[\omega'V]}_{\mathbb{E}[V_{clp}]} - \mathbb{E}[D].$$

This result shows that the variance forecasts of the LP and the CLP are upward biased, even though the individual variance forecasts are unbiased. The LP's bias is twice as large as the CLP's bias, which is equal to the expected disagreement.

Despite the popularity of the linear pool, to the best of our knowledge, this important result has not been documented in the literature yet. We are aware of two related findings, though. First, Lahiri and Sheng (2010) derive a similar result in the context of a factor model of forecaster disagreement.[4] By contrast, we work under a single generic assumption (Assumption 1) and consider implications for combining density forecasts. Second, Gneiting and Ranjan (2013, Theorem 3.1(c)) show that an LP of 'neutrally dispersed' forecast densities is underconfident. In contrast to our focus on the mean and variance, Gneiting and Ranjan (2013) define 'neutral dispersion' and underconfidence in terms of the Probability Integral Transform (PIT) which depends on the entire forecast density. While considering the entire

---

[4]Their equation (8) yields the same statement as our Proposition 1 if their expression $\sigma^2_{\lambda|th}$ is equal to our $\mathbb{E}[S]$. This condition is satisfied, for example, if the number of forecasters (N in their notation) diverges to infinity and equal combination weights are used.

density is appealing in principle, it can be hard to identify which particular features of the density prohibit calibration. Our approach of defining calibration in terms of forecast variance allows us to sharpen the result of Gneiting and Ranjan (2013) by identifying disagreement as a variance-bias augmenting term and by precisely quantifying the bias. Obviously, if the individual variance forecasts are too large, i.e. if they satisfy $\mathbb{E}[V_i] > \mathbb{E}\left[(Y - M_i)^2\right]$, the upward biases from Proposition 1 are exacerbated. By contrast, the biases are reduced and could even become negative if the individual variance forecasts are too small.

Proposition 1 states that disagreement causes an upward bias in the LP's variance, and as such is not desirable. This poses an intriguing contrast to the fact that disagreement improves the LP's mean forecast according to (9), and as such is desirable. Interestingly, assuming identical MSFEs across all mean forecasts, (9) implies that the weights ω that maximize expected disagreement $\mathbb{E}[D]$ actually minimize the combined MSFE, $\mathbb{E}[S]$.

While the LP produces biased variance forecasts, disagreement might be positively correlated with S. Therefore, removing disagreement could be detrimental to density forecast accuracy. The correlation of D and S is difficult to quantify in general, but a relevant result emerges under

Assumption 2. *(i) The vector $(M', Y)'$ with $M = (M_1, M_2, \ldots, M_n)'$ follows a multivariate normal distribution. (ii) For all $i = 1, \ldots, n$, it holds that $\mathbb{E}[M_i] = \mathbb{E}[Y]$. (iii) The combination weights ω^* are chosen to minimize the MSFE, subject to the constraint that they sum to one. (iv) The covariance matrix of $(M', Y)'$ is such that the MSFE-optimal weights ω^* are all nonnegative.*

Assumption 2 (i) is a common but strict requirement; in practice, it seems far more restrictive than the assumption of conditional normality discussed at the end of Section 2. Assumption 2 (ii) requires the mean forecasts to be unconditionally unbiased, analogous to Assumption 1 for variances. MSFE-optimal weights ω^*, as described in Assumption 2 (iii) and first considered in Bates and Granger (1969), are also optimal in terms of the DSS. Note that ω^* can have negative elements, which is problematic in the context of density combination (see Section 2). We therefore rule out negative elements in ω^* by invoking Assumption 2 (iv). However, the statement of the proposition remains valid without Assumption 2 (iv).

Proposition 2. *Under Assumption 2, $\mathbb{C}ov(D, S) = 0$ holds.*

Proof. See Appendix A. □

Proposition 2 shows that disagreement D and the squared error of the combined forecast S can be uncorrelated under a set of strict (albeit empirically potentially relevant) conditions. We next provide an intuition on why D and S are correlated under non-optimal weights: Consider two mean forecasts, where forecast A is very accurate, while forecast B is very imprecise, such that the MSFE-optimal weight on forecast A is close to one in the combined forecast. If equal weights are used for the combined forecast, and if A and B are very different, i.e. if there is strong disagreement, $\mathbb{E}[S|\mathcal{I}]$ is large, because the combined forecast then differs strongly from the accurate forecast A. If A and B are similar, i.e. if there is little disagreement, the combined forecast is close to the accurate forecast A, leading to a relatively low value of $\mathbb{E}[S|\mathcal{I}]$. This illustrates why D and S may well be correlated under non-optimal weights. The correlation of D and S under non-normality but optimal weights, however, is hard to grasp intuitively.[5] In the Monte Carlo simulations in Section 5, we therefore illustrate the impact of t-distributed forecasts on this correlation.[6]

Propositions 1 and 2 have important implications for the choice of combination weights ω. The MSFE-optimal weights ω^* are not optimal for the LP's variance, because disagreement becomes a bias-augmenting noise term, and is hence undesirable. Since disagreement vanishes if one forecast receives a weight of one, the LP faces a trade-off between accurate mean forecasts achieved by using ω^* and accurate variance forecasts achieved by using a weight vector $\iota_{[i]}$ that places a weight of one on a single forecast i, and a weight of zero on all other forecasts. The DSS-optimal weights for the LP will differ from ω^* if the gain in variance forecast accuracy obtained by moving from ω^* towards some $\iota_{[i]}$ exceeds the corresponding loss in mean forecast accuracy. The variance forecast will become more accurate in this case because its bias is reduced and its disagreement component becomes correlated with the squared forecast error. The CLP faces a similar trade-off. However, moving from ω^* towards $\iota_{[i]}$ will yield smaller gains in variance forecast accuracy for the CLP, because the initial bias is smaller and, hence, the bias reduction will be smaller. Moreover, there is no disagreement

---

[5]As detailed in the proof, the correlation between D and S depends on the correlation between two quadratic forms of the vector $Z = (M', Y)'$.

[6]If the variance of the target variable itself changes over time, and the variances of the mean forecasts change accordingly, disagreement is correlated with the squared error of the combined forecast S even if Assumption 2 holds. However, also in this case, disagreement might not convey any information about S beyond those contained in $\omega'V$. For details, see the prediction space approach in Section 5 of Knüppel and Krüger (2019).

component which becomes correlated with the squared forecast error. Therefore, the DSS-optimal weights of the LP can be expected to differ more strongly from the MSFE-optimal weights ω^* than the DSS-optimal weights of the CLP.

By removing the disagreement component, the CLP reduces the variance of the LP in a simple and transparent way. It thus offers a possible alternative to several existing density combination methods that aim to improve upon the LP. We next sketch some of these methods for the simple case that n Gaussian densities are combined; see Gneiting and Ranjan (2013) and Aastveit et al. (2019) for more general treatments.

The logarithmic pool (LogP; see Wallis, 2011, eq. 9) is identical to the CLP if all Gaussian forecast densities have identical variances v_i. If the variances are not all identical, the variance of the LogP is smaller than the CLP's variance. Similar to the CLP, the LogP's variance does not depend on the individual means m_i (see Wallis, 2011, Section III). Studies such as Lichtendahl Jr et al. (2013) and Busetti (2017) consider averaging the quantiles implied by n forecast densities (QP). In the Gaussian case, the QP density is again Gaussian with mean equal to the LP (and CLP), and variance given by $v_{qp} = (\sum_{i=1}^{n} \omega_i \sqrt{v_i})^2 \leq v_{clp} \leq v_{lp}$ (see Busetti, 2017, Section II). The QP's tendency to generate a combined density with lower variance than the LP holds beyond the Gaussian case (see Lichtendahl Jr et al., 2013, Proposition 8).

The spread-adjusted linear pool (sLP; see Gneiting and Ranjan, 2013, Section 3.2) produces the combined density $f_{slp}(y) = \sum_{i=1}^{n} \omega_i \, f_{\mathcal{N}}(y; m_i, \kappa^2 v_i)$, where $f_{\mathcal{N}}(y; m, v)$ is the density of an $\mathcal{N}(m, v)$ variable at y, and $\kappa > 0$ is a parameter to be estimated. The sLP's variance prediction is then given by $v_{slp} = d + \kappa^2 \, v_{clp}$. Hence if $\kappa < 1$, the sLP reduces the LP's variance by down-scaling its average variance component while retaining disagreement. The beta-transformed linear pool (bLP; see Gneiting and Ranjan, 2013, Section 3.3) uses the density $f_{blp}(y) = f_{lp}(y) \, \mathrm{b}_{\alpha,\beta}(F_{lp}(y))$, where $\mathrm{b}_{\alpha,\beta}$ is the density of the beta distribution with parameters $\alpha > 0, \beta > 0$ to be estimated, and F_{lp} is the CDF of the linear pool. The bLP's forecast density is particularly flexible, in that its shape can differ from the LP in various ways. However, it implicitly contains D, which can be a noise term.

Note that, in contrast to the other combination methods mentioned, implementing the sLP and bLP requires the choice of additional parameters. Moreover, the sLP and bLP approaches can also easily be applied to the CLP instead of the LP. While our analytical

and simulation results focus on the LP and CLP, we return to the sLP and bLP in the empirical analysis of Section 6.

5   Monte Carlo simulations

Here we illustrate our results by simulating variants of the baseline example in Section 3. The target variable is given by $Y = X_1 + X_2 + U$, with

$$
\begin{bmatrix} X_1 \\ X_2 \\ U \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma_{X_2}^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right).
$$

In all simulations, U is normally distributed. The mean forecasts are given by $M_i = X_i$ for $i = 1, 2$. The corresponding variance forecasts equal $V_1 = \sigma_{X_2}^2 + 1$ and $V_2 = 2$, hence satisfying Assumption 1. The mean forecast of each combination scheme we consider equals $M_c = \omega' M = \omega_1 M_1 + (1 - \omega_1) M_2$, where ω_1 is the weight for the first forecast.

We employ three combination schemes, the linear pool (LP), the centered linear pool (CLP) and, additionally, a variance-unbiased linear pool (VULP). The latter is difficult to apply in practice, but it is useful to illustrate some of the theoretical results. We denote the ith density forecast by $f_i(M_i, V_i)$, where M_i denotes the mean and V_i the variance of f_i. The density f_i need not be normally distributed, but our notation suppresses the potential dependence on additional parameters for simplicity. The density of the LP is given by $f_{lp} = \omega_1 f_1(M_1, V_1) + (1 - \omega_1) f_2(M_2, V_2)$ whereas the density of the CLP equals $f_{clp} = \omega_1 f_1(M_c, V_1) + (1 - \omega_1) f_2(M_c, V_2)$. Thus, the CLP relocates both individual density forecasts at M_c before combining them. Finally, the density of the VULP is $f_{vulp} = \omega_1 f_1(M_c, V_1 - \mathbb{E}[D]) + (1 - \omega_1) f_2(M_c, V_2 - \mathbb{E}[D])$. Hence, in addition to relocating, the VULP rescales both densities such that the individual variance forecasts are reduced by $\mathbb{E}[D]$. While all three combined densities have the same mean forecast (i.e.

$M_{lp} = M_{clp} = M_{vulp} = M_c$), the variance forecasts are as follows:

$$
\begin{aligned}
V_{lp} &= \omega'V + D = \omega_1 \left(\sigma_{X_2}^2 + 1\right) + (1 - \omega_1)\, 2 + D, \\
D &= \omega_1 \left(M_1 - M_c\right)^2 + (1 - \omega_1) \left(M_2 - M_c\right)^2 \\
&= \omega_1(1 - \omega_1)(M_1 - M_2)^2\,; \\
V_{clp} &= \omega'V = \omega_1 \left(\sigma_{X_2}^2 + 1\right) + (1 - \omega_1)\, 2\,; \\
V_{vulp} &= \omega'V - \mathbb{E}\left[D\right], \\
\mathbb{E}\left[D\right] &= \omega_1 \left(1 - \omega_1\right) \left(\sigma_{X_2}^2 + 1\right).
\end{aligned}
$$

Note that V_{clp} and V_{vulp} are constant, whereas V_{lp} contains a stochastic component. The individual and combined densities we consider do not involve parameter estimation. This 'population' perspective allows for clear comparisons to our theoretical results. All reported results are averages over 1,000 Monte Carlo iterations. Each iteration, in turn, comprises 10,000 observations of X_1, X_2, U and Y, which we use to compute the individual and combined densities and to assess their forecast performance.

In the first case we consider, all random variables and forecast densities are normal, and $\sigma_{X_2}^2$ equals 1.5. Thus, f_2 has a lower variance than f_1, and M_2 produces a lower MSFE than M_1. We consider a range of weights $\omega_1 \in [0, 1]$ placed on the first forecast. The top left panel of Figure 1 displays the variance of the three pools. In line with Proposition 1, the CLP's constant variance, V_{clp}, lies halfway between the constant optimal variance V_{vulp} and the LP's expected variance $\mathbb{E}[V_{lp}]$. Furthermore, the variance of the VULP is minimal at $\omega_1^* = 0.4$ which is the MSFE-optimal weight. The bias term $\mathbb{E}\left[D\right]$ is maximal at $\omega_1 = 0.5$. The correlation coefficients displayed in the bottom left panel of Figure 1 illustrate the result of Proposition 2, in that D and S are uncorrelated at ω_1^*.

The top left panel of Figure 2 refers to the DSS. As shown there, the optimal combination weights differ for each pool. For the VULP, the minimal score is attained at $\omega_1^* = 0.4$. For the other pools, however, it is optimal to reduce the bias of their variance forecasts at the cost of lower accuracy of their mean forecasts. For the CLP, the smallest DSS is reached at $\omega_1 = 0.37$. For the LP, which has a larger variance bias than the CLP, a considerably smaller weight of $\omega_1 = 0.24$ turns out to be optimal. Similar observations apply to the log score (bottom left panel of Figure 2). Here the optimal weights for the VULP and the CLP

are virtually the same as for the DSS, although their forecast densities are non-normal. For the LP, the optimal weight equals $\omega_1 = 0.3$. Since the LP's scores are flatter due to its higher variance bias, these simulation results indicate that finding optimal combination weights for the LP is likely to be more difficult than for the CLP in practice. For both scores, the LP performs worst for a wide range of weights around ω_1^*, and the VULP performs best.

In the second case we consider, X_1 and X_2 both follow t-distributions with 5 degrees of freedom, rescaled such that $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1$. Each individual forecast continues to be ideal – conditional on the respective information set – in terms of its mean and variance prediction. However, the forecast densities f_i are chosen to be rescaled t-distributions with 5 degrees of freedom, while the correct density would be given by the density of $M_j + U$, i.e. the sum of a (rescaled) t-distributed and a normal random variable. The results in terms of combined variance (top right panel of Figure 1) correspond to Proposition 1. The bottom right panel of Figure 1 shows the correlation between D and S. While the correlation still reaches its minimum at ω_1^*, this minimum differs from zero because the assumption of joint normality of the vector $(M_1, M_2, Y)'$ required by Proposition 2 is violated.

For each pool, the DSS (top right panel of Figure 2) and the log score (bottom right panel) differ because all combined densities are non-normal. All pools attain their lowest values at ω_1^*, and the LP performs worse than the CLP and the VULP at ω_1^* and for a certain range of weights around ω_1^*. This range covers more than the central 50% of all weights considered. The VULP outperforms the CLP with respect to the DSS. For the log score, the VULP and CLP attain similar values, with the CLP performing marginally better.[7]

6   Pooled density forecasts for inflation

We next investigate the properties of the LP and other combination methods for density forecasts of US inflation. We consider two distinct individual forecasting approaches: First, the model by Clark et al. (2020, henceforth CMM) which constructs a forecast distribution based on point forecasts from the Survey of Professional Forecasters (SPF). Briefly, CMM fit a Bayesian stochastic volatility model for the conditional distribution of the SPF point

---

[7]The latter result is basically due to the fact that a combination of misspecifications (CLP: wrong variance and wrong kurtosis) can yield a density that is closer (in terms of the Kullback-Leibler divergence) to the true density than a density with a single misspecification (VULP: same wrong kurtosis as the CLP, but correct variance).
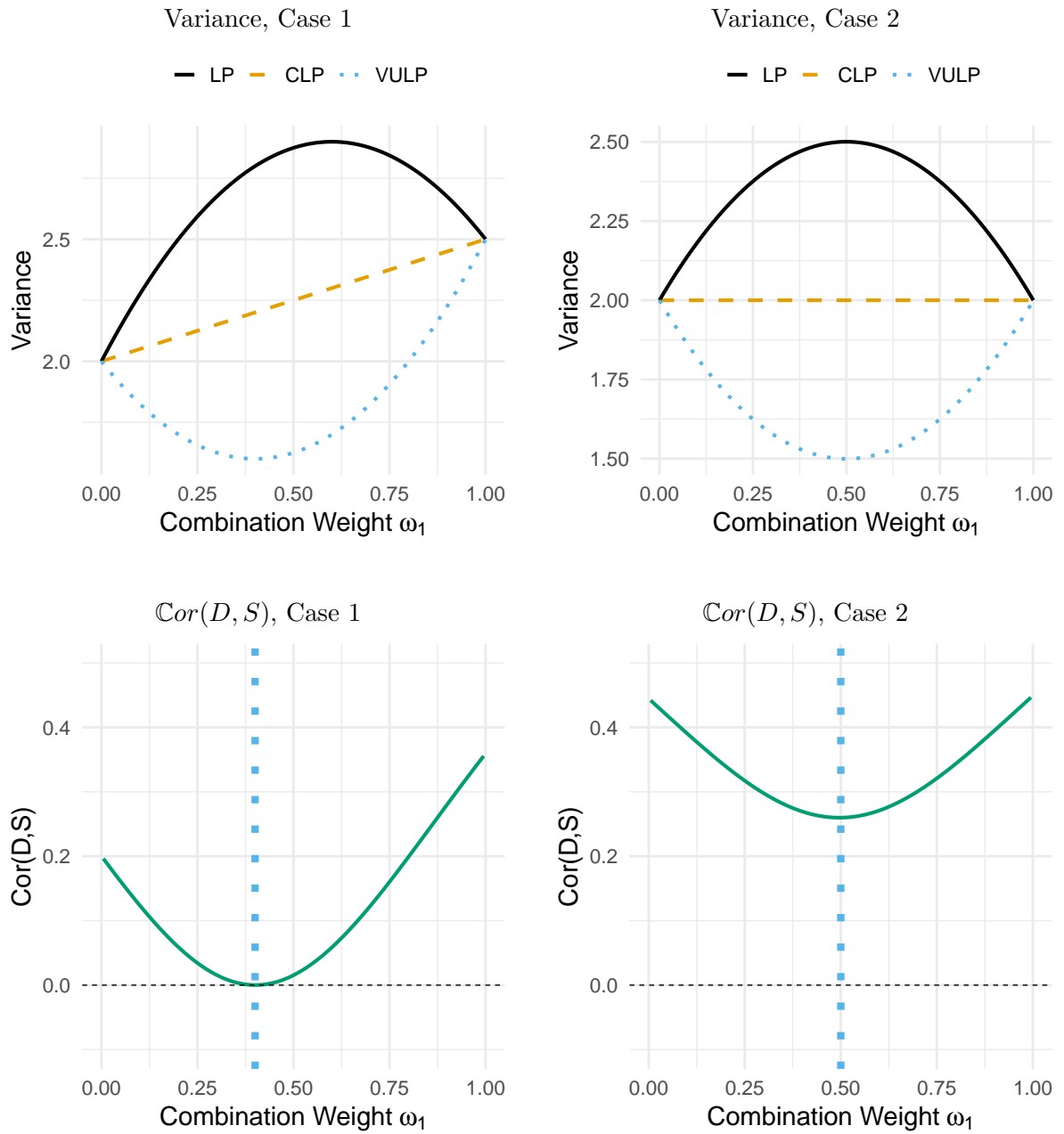
Figure 1: Left column: Simulation results for case 1 (Gaussian forecasts, $\sigma^2_{X_2} = 1.5$). Right column: Simulation results for case 2 (rescaled t-distributed density forecasts, $\sigma^2_{X_2} = 1$). The first row shows the forecast variance of the linear pool (LP), the centered linear pool (CLP) and the variance-unbiased linear pool (VULP), plotted against ω_1. The second row shows the correlation between disagreement D and the squared forecast error S of the combined mean forecast, again plotted against ω_1. The vertical blue line indicates the MSFE-optimal weight ω_1^*.
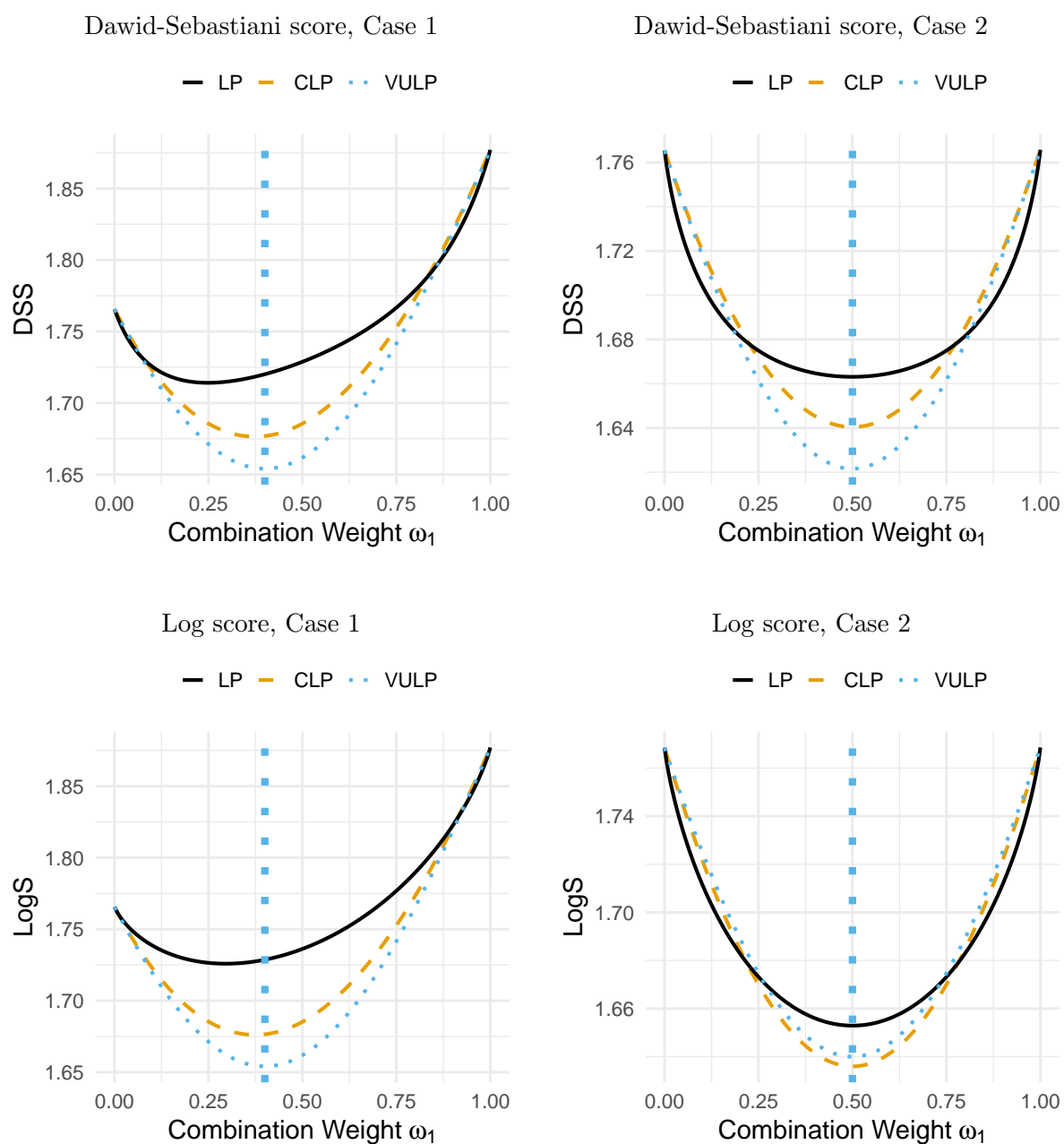
Figure 2: Left column: Simulation results for case 1 (Gaussian forecasts, $\sigma^2_{X_2} = 1.5$). Right column: Simulation results for case 2 (rescaled t-distributed density forecasts, $\sigma^2_{X_2} = 1$). The first row shows the Dawid-Sebastiani score for the linear pool (LP), the centered linear pool (CLP) and the variance-unbiased linear pool (VULP), plotted against the combination weight ω_1. A lower score indicates a more accurate forecast. The vertical blue line indicates the MSFE-optimal weight ω_1^*. The second row shows analogous plots for the log score.

forecast errors. This distribution, together with the point forecasts themselves, implies a forecast distribution for inflation. We use the 'baseline' variant that is described in detail in CMM's Sections III.A, III.B and III.D. Second, we consider the unobserved component model with stochastic volatility (UCSV) by Stock and Watson (2007). Following Chan (2013), we estimate two variance terms appearing in the model via Bayesian methods, rather than setting them to a fixed value. We provide details and prior parameters for both models in Appendix B.

The CMM model harnesses survey point forecasts which are likely to contain judgmental elements, and are often found to perform well compared to purely statistical approaches (e.g. Faust and Wright, 2013). The UCSV model, by contrast, is a prominent example of a purely statistical approach. It can be viewed as a flexible filtering technique that accommodates smooth time variation in the level and volatility of inflation. The use of stochastic volatility in both the CMM and the UCSV model reflects possible conditional heteroscedasticity in macroeconomic time series (e.g. Clark and Ravazzolo, 2015).

We estimate both models recursively using the annualized quarterly growth rate of the GDP deflator, based on first-release data available in the Federal Reserve Bank of Philadelphia's Real-Time Database. Using real-time vintages for model estimation is one of the approaches advocated by Clements and Galvao (2020) to account for the effects of data uncertainty when making probabilistic forecasts. We investigate forecasts at horizons $h = 1, 2, \ldots, 5$,[8] and our evaluation sample ranges from 1976:Q2 to 2018:Q3 (170 observations).[9] We denote the combination weight for the CMM model by ω_1.

Figure 3 presents the forecasts for the mean and the variance of inflation at horizons $h = 1$ and $h = 5$; the corresponding plots for the other (intermediate) horizons are displayed in Section III of the Online Appendix. The mean forecasts of both models are more strongly correlated for $h = 1$ than for $h = 5$, and the UCSV model tends to forecast larger variances especially around 1980. Figures 4 and 5 display the results of the LP and the CLP for all positive weights. In contrast to the Monte Carlo simulations, the variance-unbiased linear pool cannot be used because expected disagreement is unknown. The top row of Figure

---

[8]For the CMM model, forecasts at $h = 1$ utilize SPF 'nowcasts' that are released around the middle of the quarter to be predicted (e.g., the middle of February 2020 for 2020:Q1). This timing conforms roughly to the first release of the previous quarter's inflation rate, justifying the notion of a one-step ahead forecast.

[9]For forecasts at horizon $h > 1$, the first $(h - 1)$ observations cannot be used for evaluation since a corresponding forecast is not available.

4 summarizes the pools' variance forecasts and their MSFEs. The results for the weights $\omega_1 = 0$ and $\omega_1 = 1$ reveal that the variance forecasts of both models exceed their respective MSFEs; this upward bias is more pronounced for the UCSV model. Average disagreement (vertical distance between black and orange lines) depends on the combination weight ω_1. Furthermore, the MSFE of the pools' mean forecast is minimized at $\omega_1^* \approx 0.95$ for $h = 1$ and at $\omega_1^* \approx 0.6$ for $h = 5$, reflecting the better point forecast performance of the CMM model which, in turn, reflects the predictive content of the SPF.[10] The good point forecasting performance of the SPF, especially at short horizons, is well known in the literature (see e.g. Krüger et al., 2017).

The second row of Figure 4 shows the correlation between V_{lp} and S, and between V_{clp} and S. The results for $\omega_1 = 0$ and $\omega_1 = 1$ indicate that the variance forecasts are somewhat more strongly correlated with the squared mean forecast errors in the case of the CMM model. The differences between the correlations of the CLP and LP are caused by disagreement. For $h = 1$, disagreement appears to be helpful for predicting S especially if the combination weight differs sufficiently from ω_1^*, as discussed in the context of Proposition 2. For $h = 5$, the two pools' correlation with S is similar (albeit slightly larger for the LP) across all weights $\omega \in [0, 1]$.

Figure 5 display the DSS and log scores of both pools. The LP tends to attain slightly better scores than the CLP at $h = 1$, while the CLP often performs moderately better at $h = 5$. At $h = 1$, the best scores are obtained simply by using the CMM model. At $h = 5$, the CLP with a weight of roughly 0.75 would yield the best scores. The LP would prefer a weight closer to one.

The results of Diebold and Mariano (1995) tests for equally weighted forecasts (i.e. $\omega_1 = 0.5$), reported in Table 2, indicate that at $h = 1$, the differences between the scores of the LP and the CLP are insignificant. At $h = 5$, the CLP attains better scores than the LP, with the difference being significant at the 5% level for both the DSS and the log score.[11] Similarly, the first row of Table 2 shows that the CLP performs significantly better than the LP at conventional levels for both scores at horizons $h = 2, 3, 4$ with one exception.

In addition to the LP and CLP, we also investigate the spread-adjusted linear pool (sLP) and the beta-transformed linear pool (bLP) discussed in Section 4. We estimate the

---

[10]The MSFE-optimal choice of ω_1 equals $0.73, 0.59$ and 0.57 for $h = 2, 3$ and 4, respectively.

[11]Here and throughout, our statements on significance refer to two-sided tests.

required tuning parameters (κ for the sLP, α and β for the bLP) based on an expanding window of data, using a minimum of ten observations. Implementation details are described in Appendix B.3.

As indicated by Table 2 (second and third row), the sLP attains significantly better DSS scores than the CLP at horizons 3 and 5, using a 5% significance level. The performance of sLP and CLP is statistically indistinguishable at conventional levels in all other comparisons. Similarly, the score differences between the bLP and CLP are insignificant in all cases. These results indicate that more sophisticated combination methods are of limited help for addressing the LP's calibration problems, possibly due to limited data available for estimating the combination tuning parameters.

We also apply the spread adjustment and the beta transformation to the CLP's density f_{clp} instead of the LP's density f_{lp}, denoting these combination methods by sCLP (spread-adjusted centered linear pool) and bCLP (beta-transformed centered linear pool), respectively. The motivation of these variants is to first remove the potentially noisy disagreement component via centering, and then use other methods in order to handle (any remaining) miscalibration. As shown by the fourth and fifth row of Table 2, the performance of these methods (as compared to sLP and bLP) is modestly encouraging: While applying the recalibration methods to f_{clp} instead of f_{lp} leads to significant improvements at the 5% level in some instances, it never causes significant deteriorations.

Finally, Table 3 shows the average scores of all combination methods considered. While the LP attains the best scores for $h = 1$, the sCLP method mostly performs best for $h \in \{2, 3, 4, 5\}$.

	Dawid-Sebastiani score					Log score				
h	1	2	3	4	5	1	2	3	4	5
LP-CLP	-0.21	2.55	3.84	4.30	3.23	-0.81	1.10	3.30	2.62	2.30
sLP-CLP	0.23	-0.44	-2.02	-1.49	-2.32	0.43	-0.21	-1.28	-0.72	-0.69
bLP-CLP	1.71	0.27	-0.95	-0.21	-0.59	1.53	0.46	-0.63	0.20	0.16
sLP-sCLP	-0.42	0.51	1.34	2.35	1.76	-1.00	-0.31	1.43	1.38	1.21
bLP-bCLP	-0.57	0.22	2.01	2.55	1.27	-1.05	-0.25	1.58	1.37	0.87

Table 2: t-statistics for comparisons between combination methods, using the Andrews (1991) variance estimator as implemented in the R package 'sandwich' (Zeileis, 2004). A negative statistic indicates that the first method outperforms the second method and vice versa. Test statistic is standard normally distributed under the null of equal performance. All combinations are based on equal weighting ($\omega_1 = 0.5$). Evaluation sample ranges from 1976:Q2 to 2018:Q3.

	Dawid-Sebastiani score					Log score				
h	1	2	3	4	5	1	2	3	4	5
LP	**1.400**	1.519	1.554	1.610	1.680	**1.391**	1.502	1.534	1.589	1.669
CLP	1.402	1.512	1.543	1.597	1.669	1.396	1.498	1.521	1.576	1.656
sLP	1.404	1.507	1.508	1.577	1.644	1.399	**1.496**	1.504	1.569	1.650
bLP	1.422	1.517	1.521	1.592	1.654	1.411	1.505	1.509	1.581	1.660
sCLP	1.409	**1.505**	**1.501**	**1.566**	**1.636**	1.406	1.498	**1.495**	**1.560**	**1.641**
bCLP	1.428	1.516	1.514	1.583	1.648	1.418	1.506	1.502	1.574	1.655

Table 3: Average scores of combination methods, based on equal weighting ($\omega_1 = 0.5$). Score of best performing method printed in bold. Evaluation sample ranges from 1976:Q2 to 2018:Q3.
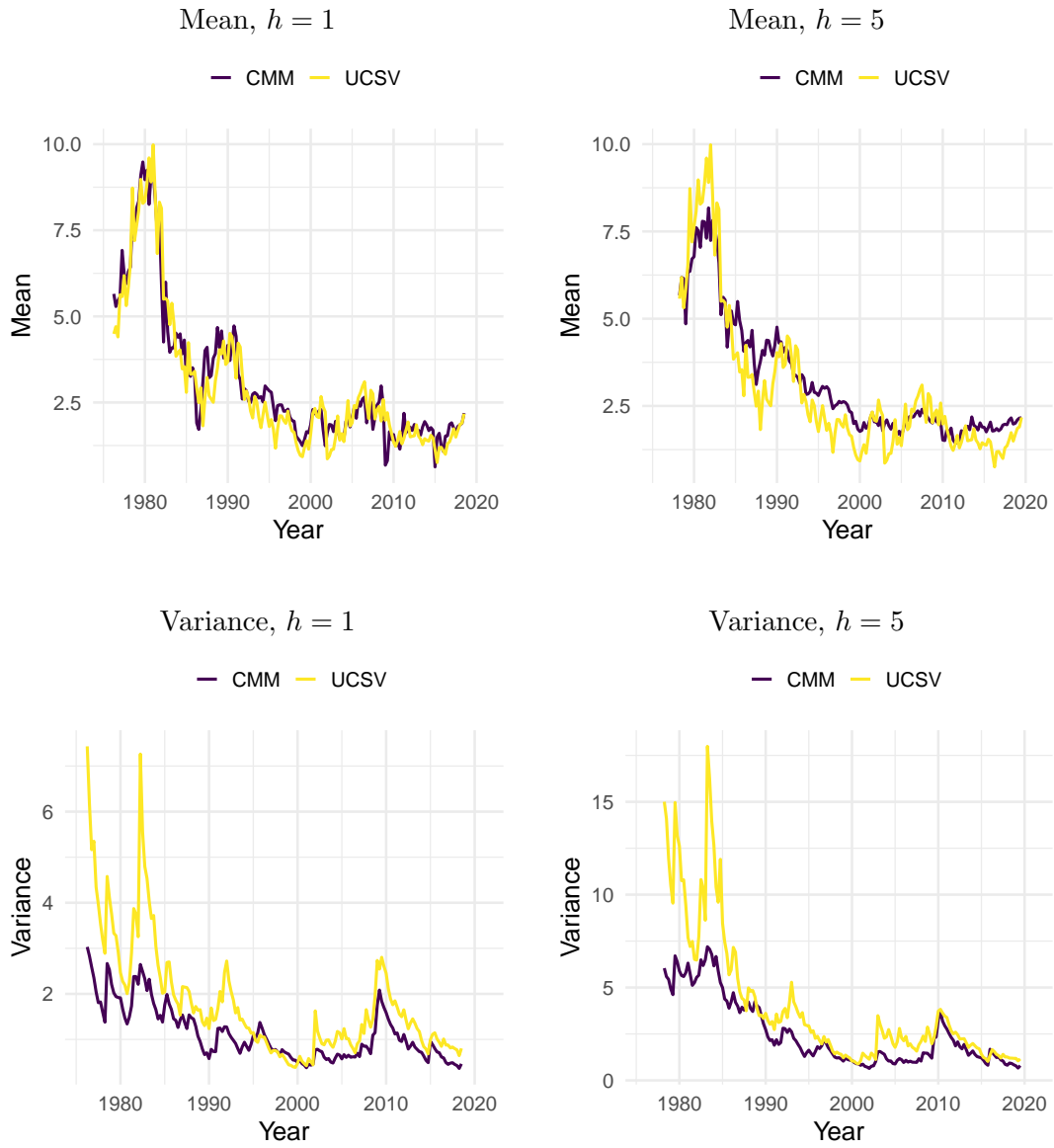
Figure 3: Mean (top row) and variance (bottom row) of the forecast distributions for the CMM and UCSV model. Left and right column correspond to shortest and longest forecast horizon ($h = 1$ and $h = 5$). Evaluation sample ranges from 1976:Q2 to 2018:Q3.
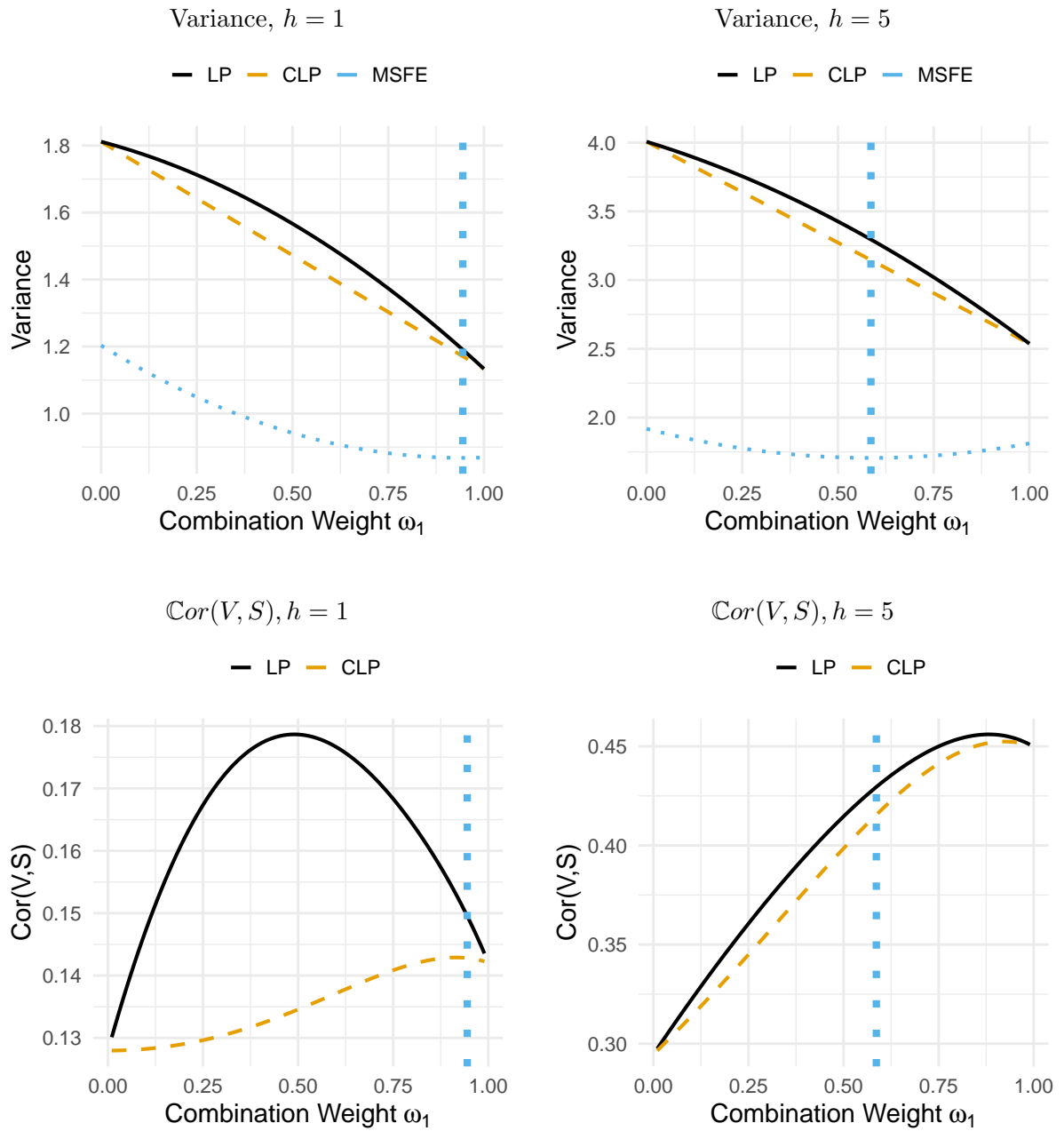
Figure 4: First row: Variance forecasts (averaged over time) and MSFEs, plotted against ω_1, the combination weight of the CMM model. Second row: Correlation between variance forecasts (V_{lp} or V_{clp}) and squared forecast errors S, again plotted against ω_1. Left and right column correspond to shortest and longest forecast horizon ($h = 1$ and $h = 5$). MSFE-optimal weight is marked by blue vertical line in each plot. Evaluation sample ranges from 1976:Q2 to 2018:Q3.
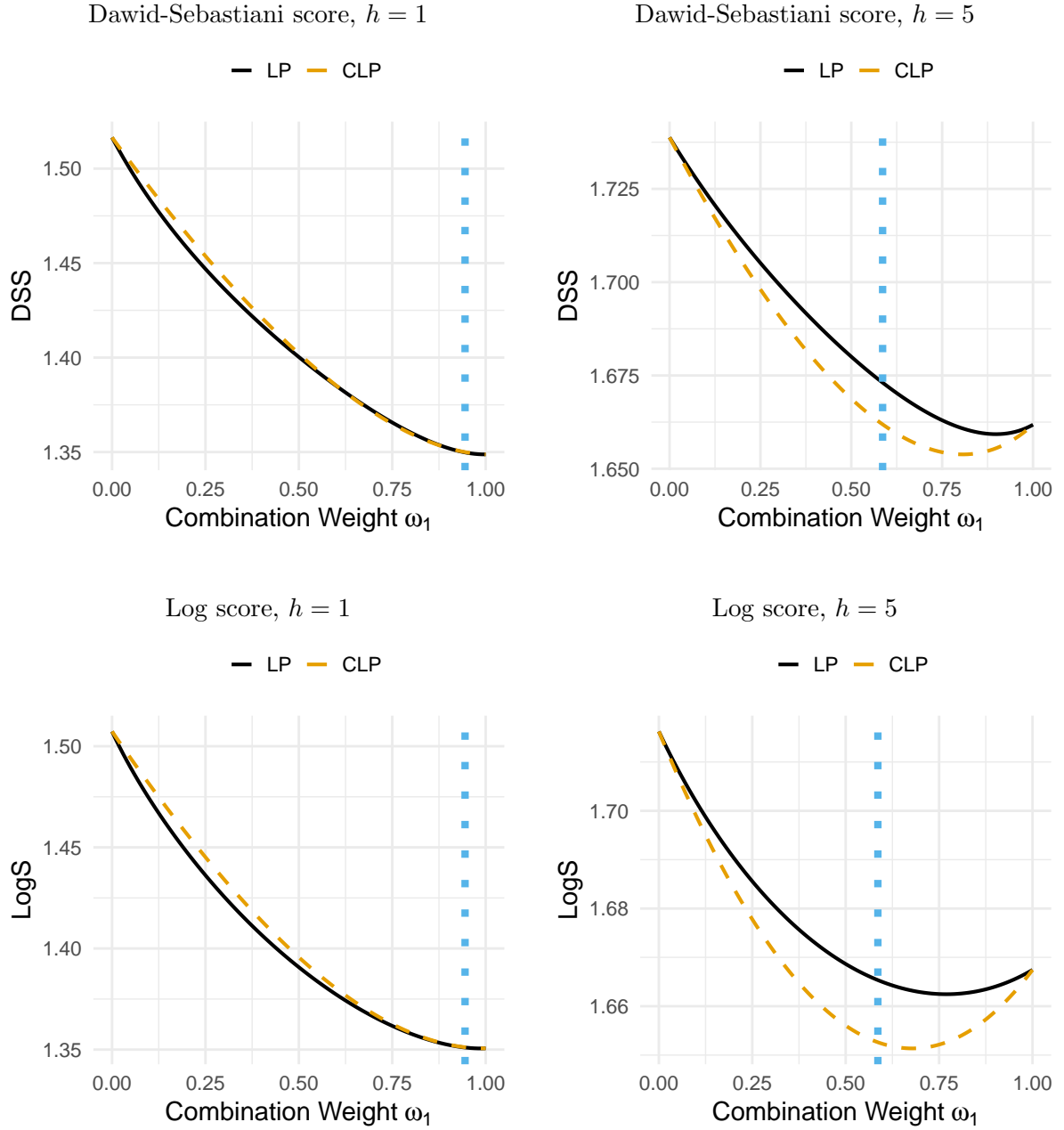
Figure 5: First row: Dawid-Sebastiani score, plotted against ω_1, the combination weight of the CMM model. Second row: Log score plotted against ω_1. Left and right column correspond to shortest and longest forecast horizon ($h = 1$ and $h = 5$). A lower score indicates a more accurate forecast. MSFE-optimal weight is marked by blue vertical line in each plot. Evaluation sample ranges from 1976:Q2 to 2018:Q3.

7 Conclusion

Incorporating various sources of information continues to be an important challenge in forecasting. Ideally, one would like to combine the underlying information sets, and then construct a single model that makes optimal use of the combined information set (e.g. Gneiting and Ranjan, 2013, p. 1750). This approach is hard or impossible to implement in practice. Forecast combinations, by contrast, are simple to implement. While they tend to make suboptimal use of available information (see e.g. Satopää 2017 on mean forecasts, and Gneiting and Ranjan 2013 on density forecasts), they often perform well in practice (e.g. Timmermann, 2006). In the context of forecast densities, the linear pool (LP) is the most popular combination technique.

We show that the LP's variance forecast is upward biased by twice the expected disagreement – a component of the LP's variance forecast that reflects differences between the individual mean forecasts – if the individual variance forecasts are unbiased. Moreover, we find that disagreement is uncorrelated with the squared mean forecast error of the LP under empirically relevant conditions. Motivated by these insights, we propose a simple modification of the LP that removes the disagreement component: the centered linear pool (CLP).

Roughly speaking, the CLP can be expected to improve on the LP under the following conditions: First, the forecast variances of the individual models must be large enough. Second, disagreement must not be overly informative about squared mean forecast errors. Third, for the CLP to differ from the LP in a relevant way, disagreement must not be too small. Absent structural breaks at the forecast origin, the first condition seems plausible if the individual forecasts are based on statistical models, where variance calibration is typically ensured in the model fitting process. The condition seems far more restrictive if the individual forecasts are judgmentally generated by humans who often tend to underestimate uncertainty (see e.g. Moore et al., 2015). The second and third conditions are more elusive, and are specific to the set of models being combined. The second condition is more likely to be fulfilled if the combination weights are close to the MSFE-optimal weights of Bates and Granger (1969). The third condition requires some degree of predictability of the conditional mean. In the absence of predictability, all individual forecasts are similar to the unconditional mean. This phenomenon is typical of equity return predictions, which are considered in

Knüppel and Krüger (2019). In this case, the CLP and LP are essentially equivalent.

Various alternatives to the LP have been proposed in the literature. While some of these methods are far more flexible than the CLP, their flexibility comes at the cost of additional parameters that must be estimated or fixed. In our empirical analysis of inflation data, two flexible methods based on the LP can mostly not significantly improve upon the CLP, perhaps due to the relatively short sample available for estimating their parameters, but possibly also reflecting noise in the disagreement component. Therefore, it may be helpful to consider the CLP instead of the LP as an input to more flexible combination methods.

References

Aastveit, K. A., Mitchell, J., Ravazzolo, F., and van Dijk, H. (2019). The evolution of forecast density combinations in economics. In *Oxford Research Encyclopaedia in Economics and Finance*. Oxford University Press.

Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109:1263–89.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.

Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20:451–468.

Busetti, F. (2017). Quantile aggregation of density forecasts. *Oxford Bulletin of Economics and Statistics*, 79:495–512.

Carriero, A., Clark, T. E., and Marcellino, M. (2020). Capturing macroeconomic tail risks with Bayesian vector autoregressions. Working Paper, Cleveland Federal Reserve Bank.

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

Clark, T. E., McCracken, M. W., and Mertens, E. (2020). Modeling time-varying uncertainty of multiple-horizon forecast errors. *Review of Economics and Statistics*, 102(1):17–33.

Clark, T. E. and Ravazzolo, F. (2015). The macroeconomic forecasting performance of autoregressive models with alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30:551–575.

Clements, M. P. (2014). Forecast uncertainty – ex ante and ex post: US inflation and output growth. *Journal of Business & Economic Statistics*, 32:206–216.

Clements, M. P. (2018). Are macroeconomic density forecasts informative? *International Journal of Forecasting*, 34:181–198.

Clements, M. P. and Galvao, A. B. (2020). Density Forecasting with BVAR Models under Macroeconomic Data Uncertainty. EMF Research Papers 36, Economic Modelling and Forecasting Group.

Dawid, A. P. and Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. *Annals of Statistics*, 27:65–81.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13:253–263.

Dovern, J., Fritsche, U., and Slacalek, J. (2012). Disagreement among forecasters in G7 countries. *Review of Economics and Statistics*, 94:1081–1096.

Engle, R. F. (1983). Estimates of the variance of US inflation based upon the ARCH model. *Journal of Money, Credit and Banking*, 15:286–301.

Faust, J. and Wright, J. H. (2013). Forecasting inflation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, pages 2–56. Elsevier.

Franta, M., Baruník, J., Horváth, R., and Šmídková, K. (2014). Are Bayesian fan charts useful? The effect of zero lower bound and evaluation of financial stability stress tests. *International Journal of Central Banking*, 10:159–188.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102:359–378.

Gneiting, T. and Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, 7:1747–1782.

Knüppel, M. and Krüger, F. (2019). Forecast uncertainty, disagreement, and the linear pool. Discussion Papers 28/2019, Deutsche Bundesbank.

Krüger, F., Clark, T. E., and Ravazzolo, F. (2017). Using entropic tilting to combine BVAR forecasts with external nowcasts. *Journal of Business & Economic Statistics*, 35:470–485.

Krüger, F., Lerch, S., Thorarinsdottir, T. L., and Gneiting, T. (2020). Predictive inference based on Markov Chain Monte Carlo output. *International Statistical Review*. Forthcoming.

Lahiri, K., Peng, H., and Sheng, X. (2015). Measuring uncertainty of a combined forecast and some tests for forecaster heterogeneity. CESifo Working Paper Series 5468, CESifo.

Lahiri, K. and Sheng, X. (2010). Measuring forecast uncertainty by disagreement: The missing link. *Journal of Applied Econometrics*, 25:514–538.

Lahiri, K., Teigland, C., and Zaporowski, M. (1988). Interest rates and the subjective probability distribution of inflation forecasts. *Journal of Money, Credit and Banking*, 20(2):233–248.

Lerch, S., Thorarinsdottir, T. L., Ravazzolo, F., Gneiting, T., et al. (2017). Forecaster's dilemma: Extreme events and forecast evaluation. *Statistical Science*, 32:106–127.

Lichtendahl Jr, K. C., Grushka-Cockayne, Y., and Winkler, R. L. (2013). Is it better to average probabilities or quantiles? *Management Science*, 59:1594–1611.

Manski, C. F. (2010). When consensus choice dominates individualism: Jensen's inequality and collective decisions under uncertainty. *Quantitative Economics*, 1:187–202.

Manski, C. F. (2015). Communicating uncertainty in official economic statistics: An appraisal fifty years after Morgenstern. *Journal of Economic Literature*, 53:631–53.

Matheson, J. E. and Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, 22:1087–1096.

McNees, S. K. (1992). The uses and abuses of consensus forecasts. *Journal of Forecasting*, 11:703–710.

Moore, D. A., Tenney, E. R., and Haran, U. (2015). Overprecision in judgment. *The Wiley Blackwell Handbook of Judgment and Decision Making*, pages 182–209.

Page, S. E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.

Page, S. E. (2018). *The Model Thinker: What You Need to Know to Make Data Work for You*. Basic Books.

Petersen, K. B. and Pedersen, M. S. (2012). The matrix cookbook. Available at `http://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf` (last accessed: October 18, 2020).

Satopää, V. A. (2017). Combining information from multiple forecasters: Inefficiency of central tendency. *Working Paper, arXiv:1706.06006*.

Stock, J. H. and Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39:3–33.

Stone, M. (1961). The opinion pool. *The Annals of Mathematical Statistics*, 32:1339–1342.

Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, pages 135 – 196. Elsevier.

Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics*, 67:983–994.

Wallis, K. F. (2011). Combining forecasts–forty years later. *Applied Financial Economics*, 21:33–41.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11:1–17.

A   Appendix: Proof of Proposition 2

Due to Assumption 2 (i) and (ii), i.e. due to normality and equal means of M and Y, we can write $Y = M'\gamma + U$, where

$$\begin{bmatrix} M \\ U \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_M \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_M & 0 \\ 0 & \Sigma_U \end{bmatrix} \right), \tag{13}$$

where γ is the vector of coefficients for the best linear predictor of Y given M. Without loss of generality, we assume that $\mu_M = 0$. Disagreement is given by

$$D = (M - \iota\ \omega' M)'\ G\ (M - \iota\ \omega' M),$$

where ι is an $n \times 1$ vector of ones, and G is an $n \times n$ matrix with ω sitting on the main diagonal, and all other elements equal to zero. Noting that $\iota' G = \omega'$ and $\iota' G \iota = 1$, this can be simplified to

$$D = M' A M, \tag{14}$$

where $A \equiv [G - \omega\omega']$. The squared error of the combined forecast is given by

$$\begin{aligned}
S &= \left[Y - \omega' M\right]' \left[Y - \omega' M\right] \\
&= \left[U' + M'(\gamma - \omega)\right] \left[U + (\gamma - \omega)' M\right] \\
&= U'U + M' B M + 2\ M'(\gamma - \omega)\ U, \tag{15}
\end{aligned}$$

where $B \equiv (\gamma - \omega)(\gamma - \omega)'$.

To compute the covariance between D and S, note that

$$DS = M'AM\ U'U + M'AM\ M'BM + 2M'AM\ M'(\gamma - \omega)U,$$

$$\mathbb{E}\left[DS\right] = \underbrace{\mathbb{E}\left[M'AM\right]}_{=\mathbb{E}[D]}\Sigma_U + \mathbb{E}\left[M'AM\ M'BM\right], \tag{16}$$

where we have used the independence of U and M. The second summand in (16) is a quartic form in a Gaussian random vector. The results in Section 8.2.4 of Petersen and Pedersen (2012) and the assumption that $\mathbb{E}[M_i] = 0, i = 1, \ldots, n$ imply that

$$\mathbb{E}\left[M'AM\ M'BM\right] = 2\ Tr\left[A\Sigma_M B\Sigma_M\right]$$

$$+ \underbrace{Tr\left[A\Sigma_M\right]}_{=\mathbb{E}[D]}\underbrace{Tr\left[B\Sigma_M\right]}_{=\mathbb{E}[S]-\Sigma_U},$$

where we have used the symmetry of A and B; the expectations of D and S follow from

equations (14) and (15) and the results in Section 8.2.2 of Petersen and Pedersen (2012). Substituting back into (16) and rearranging, we find that

$$Cov(D, S) = \mathbb{E}[DS] - \mathbb{E}[D]\mathbb{E}[S] = 2\, Tr\left[A\Sigma_M B\Sigma_M\right]. \tag{17}$$

Thus, the covariance between D and S is nonzero in general. However, in case of (constrained) optimal combination weights, Proposition 2 states that $Cov(D, S) = 0$. In order to prove this statement, we first derive an expression for $B = (\gamma - \omega^*)(\gamma - \omega^*)'$, where ω^* denotes the optimal combination weights. Since the weights ω^* are restricted to sum to 1,

$$\iota'\omega^* = 1,$$

they are given by the population value of a restricted least squares regression of Y on M. The probability limit of this estimator is given by

$$\omega^* = \gamma - \Sigma_M^{-1}\iota\left(\iota'\Sigma_M^{-1}\iota\right)^{-1}(\iota'\gamma - 1). \tag{18}$$

Hence,

$$
\begin{aligned}
B &\equiv (\gamma - \omega^*)(\gamma - \omega^*)' \\
&= \underbrace{\frac{(\iota'\gamma - 1)^2}{\left(\iota'\Sigma_M^{-1}\iota\right)^2}}_{\equiv\, \underline{c}\, \in\, \mathbb{R}_+}\left(\Sigma_M^{-1}\iota\iota'\Sigma_M^{-1}\right).
\end{aligned} \tag{19}
$$

We can now use these results to show that the term on the right-hand side of (17) equals zero. Equation (19) and the definition $A = [G - \omega^*\omega^{*\prime}]$ imply that

$$
\begin{aligned}
A\Sigma_M B\Sigma_M &= \underline{c}\left[G - \omega^*\omega^{*\prime}\right]\iota\iota', \\
&= \underline{c} \times \left(\underbrace{G\iota\iota'}_{=\omega^*\iota'} - \omega^*\underbrace{\omega^{*\prime}\iota}_{=1}\iota'\right), \\
&= \underline{c} \times 0 \times I,
\end{aligned}
$$

such that $Tr\left(A\Sigma_M B\Sigma_M\right) = 0$, completing the proof.

B Appendix: Details on models for Section 6

B.1 CMM model

The CMM model refers to the forecast error of the SPF point forecast at a given date and forecast horizon. To obtain a forecast distribution, one can simply shift the error distribution by the SPF mean forecast. As a simple example, suppose the SPF mean forecast is equal to three and the error distribution is standard normal. The forecast distribution is then given by a normal distribution with mean three and variance one.

The CMM model considers a five-dimensional vector η_t containing a nowcast error and four expectational updates; see Clark et al. (2020, Section III.B) for details. Notice that $\eta_{\mathbf{t}}$ is an auxiliary vector that encodes information on the SPF forecast errors. While the SPF forecast errors are highly persistent by construction, η_t can plausibly be modeled as a martingale difference sequence (MDS) with the property that $\mathbb{E}(\eta_t|\eta_{t-1}, \eta_{t-2}, \ldots) = 0$. Modeling η_t is hence preferable to modeling the forecast errors directly. We consider the following variant of the CMM model:

$$
\begin{aligned}
\eta_t &= A^{-1}\, \tilde{\eta}_t, \quad \tilde{\eta}_t = \mathbf{\Lambda_t}^{0.5}\varepsilon_{\mathbf{t}} \qquad (20)\\
\mathbf{\Lambda_t} &= \operatorname{diag}(\lambda_{1,t}, \ldots, \lambda_{5,t})\\
\varepsilon_{\mathbf{t}} &\sim \mathcal{N}(0, I_5),
\end{aligned}
$$

where A in (20) is a lower triangular matrix of dimension 5×5 and I_5 is the five-dimensional identity matrix. The variance-covariance matrix of η_t is hence given by $A^{-1}\mathbf{\Lambda_t}A^{-1'}$. This specification is almost identical to CMM's baseline variant using the MDS assumption, except for a minor difference in parametrization (CMM consider the inverse of the matrix A in equation 20). Stochastic volatility is captured by assuming a random walk process for the logarithmic value of $\lambda_{i,t}$, which is the ith element of $\mathbf{\Lambda_t}$:

$$
\begin{aligned}
\log \lambda_{i,t} &= \log \lambda_{i,t-1} + \nu_{i,t},\\
\left(\nu_{1,t}, \quad \ldots, \quad \nu_{5,t}\right)' &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Phi}).
\end{aligned}
$$

We use uninformative priors for the elements of A. Following CMM, we use an inverse Wishart prior for $\mathbf{\Phi}$, using 14 degrees of freedom and implying that the prior mean matrix

is given by $(0.2^2)\ I_5$. This prior specification is somewhat informative, and in particular discourages implausibly large values of $\mathbf{\Phi}$. Again following CMM, our prior for the initial log variances $\log \lambda_{i,0}$ is $\mathcal{N}(\log(0.25), 10)$, independently across $i = 1, \ldots, 5$.

Conditional on the model parameters drawn in MCMC iteration j, the CMM model's forecast distribution at a generic horizon is Gaussian with mean equal to the SPF forecast, and variance v_j. The final predictive distribution then obtains as a scale mixture of M Gaussian distributions, where M is the number of MCMC iterations. In our implementation we use $M = 10,000$ (without thinning), obtained after discarding $5,000$ burn-in draws. Our code for the CMM model is written in R and C++, and is partly based on the MAT-LAB code kindly made available by Elmar Mertens at `https://github.com/elmarmertens/CMMrestat-TimeVaryingUncertainty`.

B.2   UCSV model

The UCSV model by Stock and Watson (2007) assumes the following process for the inflation rate y_t:

$$
\begin{aligned}
y_t &= \tau_t + \varepsilon_t^y, & \varepsilon_t^y &\sim \mathcal{N}(0, \exp(h_t)) \\
\tau_t &= \tau_{t-1} + \varepsilon_t^\tau, & \varepsilon_t^\tau &\sim \mathcal{N}(0, \exp(g_t))) \\
h_t &= h_{t-1} + \varepsilon_t^h, & \varepsilon_t^h &\sim \mathcal{N}(0, \sigma_h^2) \\
g_t &= g_{t-1} + \varepsilon_t^g, & \varepsilon_t^g &\sim \mathcal{N}(0, \sigma_g^2)
\end{aligned}
$$

Here τ_t is the trend rate of inflation; ε_t^y and ε_t^τ denote the innovations to inflation itself and trend inflation respectively. Both innovations are assumed heteroscedastic, with time varying log variances h_t and g_t.

Our implementation and prior choices follow Chan (2013). For σ_h^2 and σ_g^2, we employ inverse gamma priors with shape parameter equal to 10 and rate parameter equal to 0.45.

We use the MCMC implementation by Chan (2013) to estimate the model, based on MATLAB code kindly made available by the author at `http://joshuachan.org/code/code_MASV.html`. We consider $10,000$ MCMC iterations, without thinning, and after discarding a burn-in sequence of $5,000$ draws. In each iteration, we compute the mean and variance of the Gaussian forecast distribution for inflation (conditional on model parame-

ters in the current iteration), at a given forecast horizon. As detailed below, the model's predictive distribution obtains as a mixture of $10,000$ normal distributions.

B.3   MCMC sampling and forecast distributions

The predictive distributions of the CMM and UCSV methods are both produced via Bayesian MCMC sampling, and both are mixtures of normals. We next describe the details of various forecast combination methods in this setup. To avoid clutter, we index the CMM and UCSV methods by 'C' and 'U', and we suppress the forecast origin date and forecast horizon. The forecast density of model $i \in \{C, U\}$ takes the form

$$f_i(y) = \frac{1}{M} \sum_{j=1}^{M} f_\mathcal{N}(y; m_{ij}, v_{ij}),$$

where M is the number of MCMC draws, $f_\mathcal{N}(y; m, v)$ is the density at $y \in \mathbb{R}$ of a Gaussian distribution with mean m and variance v, and m_{ij} and v_{ij} denote the predictive mean and variance in the jth MCMC draw for model i. Furthermore, let ω_1 and $1 - \omega_1$ denote the weights placed on models C and U respectively. The following formulas describe the relevant pooling methods in the empirical setup of Section 6.

- **Linear pool (LP)**

$$f_{lp}(y) = \frac{1}{M} \left(\omega_1 \sum_{j=1}^{M} f_\mathcal{N}(y; m_{Cj}, v_{Cj}) + (1-\omega_1) \sum_{j=1}^{M} f_\mathcal{N}(y; m_{Uj}, v_{Uj}) \right),$$

  where we consider a range of weights $\omega_1 \in [0, 1]$.

- **Centered linear pool (CLP)**

$$f_{clp}(y) = \frac{1}{M} \left(\omega_1 \sum_{j=1}^{M} f_\mathcal{N}(y; m_{Cj} + \bar{m} - \bar{m}_C, v_{Cj}) + (1-\omega_1) \sum_{j=1}^{M} f_\mathcal{N}(y; m_{Uj} + \bar{m} - \bar{m}_U, v_{Uj}) \right),$$

  where $\bar{m}_j = M^{-1} \sum_{j=1}^{M} m_{ij}$ denotes the mean forecast of method $j \in \{C, U\}$, and $\bar{m} = \omega_1 \bar{m}_C + (1 - \omega_1) \bar{m}_U$. Note that the forecast draws of methods C and U are recentered such that the mean forecast is equal to $\bar{m}$ in each case. We again consider a range of weights $\omega_1 \in [0, 1]$.

- **Spread-adjusted linear pool (sLP)**

$$f_{slp}(y) = \frac{1}{M}\left(\omega_1 \sum_{j=1}^{M} f_{\mathcal{N}}(y; m_{Cj}, \kappa^2 v_{Cj}) + (1-\omega_1)\sum_{j=1}^{M} f_{\mathcal{N}}(y; m_{Uj}, \kappa^2 v_{Uj})\right),$$

where $\kappa \in \mathbb{R}_+$ is a parameter to be estimated from a sample of past forecasts and realizations. As described in the text, we use an expanding window of at least 10 observations for estimation. We further fix $\omega_1 = 0.5$.

- **Spread-adjusted centered linear pool (sCLP)**

$$f_{sclp}(y) = \frac{1}{M}\left(\omega_1 \sum_{j=1}^{M} f_{\mathcal{N}}(y; m_{Cj} + \bar{m} - \bar{m}_C, \kappa^2 v_{Cj}) + (1-\omega_1)\sum_{j=1}^{M} f_{\mathcal{N}}(y; m_{Uj} + \bar{m} - \bar{m}_U, \kappa^2 v_{Uj})\right),$$

where the interpretation and handling of κ is the same as in the sLP, and we again fix $\omega_1 = 0.5$. The sCLP can be thought of as first centering the linear pool, and then applying spread adjustment by estimating κ.

- **Beta-transformed linear pool (bLP)**

$$f_{blp}(y) = f_{lp}(y)\, b_{\alpha,\beta}(F_{lp}(y)),$$

where $b_{\alpha,\beta}$ is the density of the beta distribution with parameters $\alpha > 0, \beta > 0$ to be estimated. We use an expanding window of at least 10 observations for estimation. Furthermore, we fix $\omega_1 = 0.5$ in the linear pool that enters the bLP.

- **Beta-transformed centered linear pool (bCLP)**

$$f_{bclp}(y) = f_{clp}(y)\, b_{\alpha,\beta}(F_{clp}(y)),$$

where the definition and implementation of $b_{\alpha,\beta}$ is as in the bLP.