Discussion Paper

Forecast uncertainty, disagreement, and the linear pool

Malte Knüppel
(Deutsche Bundesbank)

Fabian Krüger
(Heidelberg University)

Non-technical summary

Research Question

Point forecasts are often supplemented with information concerning the uncertainty surrounding these forecasts. Sometimes, a probability distribution for the target variable is provided instead, where its mean and its dispersion correspond to the point forecast and the associated uncertainty, respectively. If several of such probability forecasts for the same variable are available, a weighted sum of these forecasts often serves as a device to pool the information from all individual forecasts, yielding a single, aggregated probability distribution. This so-called linear pool is the concept underlying, for instance, the probability forecasts of the ECB's Survey of Professional Forecasts, as these forecasts are calculated as an (equally) weighted sum of the individual probability forecasts.

Contribution

Although the linear pool is used in many empirical applications, there are several open questions concerning the properties of this pool with respect to forecast uncertainty. However, it is well-known that the dispersion of the aggregated probability distribution is determined by the dispersions of the individual forecasts and the differences between the means of the individual forecasts. The latter component is often referred to as disagreement. We investigate the role of disagreement with respect to the forecast uncertainty of the linear pool.

Results

If the dispersions of the individual forecasts are unbiased, i.e. if the dispersion of each forecast (ex-ante uncertainty of the individual forecast) is equal to the dispersion of the forecast errors of the corresponding mean forecast (ex-post uncertainty of the individual forecast) on average, then the ex-ante uncertainty of the linear pool turns out to be biased. To be more precise, in this case the linear pool's dispersion exceeds the linear pool's ex-post uncertainty, i.e. the dispersion of the forecast errors of the linear pool's mean forecast, on average. The size of this bias is at least as large as the average disagreement. It also turns out that, under empirical relevant conditions, disagreement does not contain any useful information concerning the linear pool's ex-post uncertainty beyond the information already contained in the ex-ante uncertainties of the individual forecasts. Since disagreement then does not help to predict ex-post uncertainty and contributes to the bias

of the linear pool, we propose a simple alternative for pooling the information from all individual forecasts which does not include a disagreement component. This alternative pool yields better forecasts than the linear pool in simulations and empirical applications.

Nichttechnische Zusammenfassung

Fragestellung

Punktprognosen werden häufig mit Angaben darüber versehen, von wieviel Unsicherheit diese Prognosen umgeben sind. Manchmal wird stattdessen auch direkt eine Wahrscheinlichkeitsverteilung für die prognostizierte Variable angeführt, deren Mittelwert und Streuung dann der Punktprognose und der sie umgebenden Unsicherheit entsprechen. Wenn verschiedene dieser Wahrscheinlichkeitsprognosen für dieselbe Variable verfügbar sind, wird oft die gewichtete Summe dieser Prognosen betrachtet, um die Informationen aus allen Einzelprognosen in einer einzigen, aggregierten Wahrscheinlichkeitsprognose zusammenzufassen. Auf diesem Prinzip beruhen zum Beispiel die Wahrscheinlichkeitsprognosen des Survey of Professional Forecasters der EZB, die sich als (gleich)gewichtete Summe der Wahrscheinlichkeitsprognosen aller Umfrageteilnehmer ergeben.

Beitrag

Obwohl in vielen empirischen Studien die gewichtete Summe von einzelnen Wahrscheinlichkeitsprognosen verwendet wird, sind die theoretischen Eigenschaften dieser linearen Aggregation in Bezug auf die Prognoseunsicherheit kaum erforscht worden. Allerdings ist bekannt, dass sich die Streuung der aggregierten Wahrscheinlichkeitsprognose aus den Streuungen der Einzelprognosen und den Unterschieden zwischen den Mittelwerten der Einzelprognosen, der sogenannten Uneinigkeit, ergibt. Wir untersuchen, welche Rolle die Uneinigkeit in Bezug auf die Prognoseunsicherheit der aggregierten Wahrscheinlichkeitsprognose spielt.

Ergebnisse

Wenn die Streuung der Einzelprognosen unverzerrt ist, wenn also die Streuung der jeweiligen einzelnen Wahrscheinlichkeitsprognosen (ex-ante Unsicherheit der Einzelprognosen) im Mittel der Streuung der Fehler ihrer Punktprognosen entspricht (ex-post Unsicherheit der Einzelprognosen), dann ist die ex-ante Unsicherheit der aggregierten Prognose verzerrt. Die Streuung der aggregierten Wahrscheinlichkeitsprognose übersteigt dann nämlich im Mittel die ex-post Unsicherheit der aggregierten Prognose, also die Streuung der Fehler jener Punktprognose, die sich aus der aggregierten Wahrscheinlichkeitsprognose ergibt. Die Höhe der Verzerrung entspricht dabei mindestens dem Betrag der durchschnittlichen Uneinigkeit. Zudem sagt eine Veränderung der Uneinigkeit unter empirisch relevanten

Bedingungen nichts über die zukünftige Veränderung der ex-post Unsicherheit der aggregierten Prognose aus, was nicht schon aus der Veränderung der ex-ante Unsicherheit der Einzelprognosen ersichtlich ist. Da die Uneinigkeit im Rahmen der aggregierten Wahrscheinlichkeitsprognose dann also keinen prognostischen Wert besitzt und außerdem zu Verzerrungen führt, schlagen wir eine einfache, alternative Kombination von einzelnen Wahrscheinlichkeitsprognosen vor, in der die Uneinigkeit keine Rolle spielt. Diese alternative Kombination liefert in Simulationen und verschiedenen empirischen Anwendungen teils deutlich bessere Prognosen als die übliche gewichtete Summe.

Forecast Uncertainty, Disagreement, and the Linear Pool*

Malte Knüppel
Deutsche Bundesbank

Fabian Krüger
Heidelberg University

July 9, 2019

Abstract

The linear pool is the most popular method for combining density forecasts. We analyze the linear pool's implications concerning forecast uncertainty in a new theoretical framework that focuses on the mean and variance of each density forecast to be combined. 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 the linear pool's variance prediction. Moreover, we find that disagreement has no predictive content for ex-post forecast uncertainty under conditions which can be empirically relevant. These findings suggest the removal of the disagreement component from the linear pool. The resulting centered linear pool outperforms the linear pool in simulations and in empirical applications to inflation and stock returns.

JEL classification: C32, C53.

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, Baruník, Horváth, and Šmídková (2014, Table 1). Similarly, Manski (2015) calls for systematic measurement and communication of uncertainty in official statistics. In statistical terms, confronting uncertainty about future developments means 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, in a recent comprehensive survey Aastveit, Mitchell, Ravazzolo, and van Dijk (2018, p. 20) argue 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 calculated 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 Surveys of Professional Forecasters (SPF) conducted by the European Central Bank and the Federal Reserve Bank of Philadelphia.

In the present paper, we analyze the LP's implications concerning forecast uncertainty. For this purpose, we develop a novel mean-variance prediction space framework for the joint distribution of mean forecasts, variance forecasts, and the target variable in terms of their first two moments. This setup allows us to derive several new results. We focus 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, Fritsche, and Slacalek, 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, the LP's forecast is underconfident (i.e. the LP's variance is too large), and the expected disagreement is one component of this upward bias. This result sharpens the findings by Gneiting and Ranjan (2013) who characterize the LP's underconfidence in terms of its Probability Integral Transform (PIT). 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 starkly violating a desideratum of good uncertainty forecasts. Third, we show 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 usually 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 trivial 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 empirical examples of inflation and stock return forecasts. In both

empirical examples, the CLP outperforms the LP. We conjecture that the CLP may be a better starting point than the LP for considering sophisticated specifications of the combination weights as proposed, e.g. by Billio, Casarin, Ravazzolo, and van Dijk (2013) and Del Negro, Hasegawa, and Schorfheide (2016).

The remainder of this paper is structured as follows: Section 2 derives simple yet important 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 a general result on bias in the LP's variance forecast, which is based on the sole assumption that the individual variance forecasts are unbiased. Section 5 turns to a prediction space framework which prescribes a joint model for mean forecasts, variance forecasts, and the variable to be predicted. We introduce this more specific setup in order to derive more specific results, and to identify determinants of the LP's performance. Sections 6 and 7 contain results of Monte Carlo simulations and empirical applications, and Section 8 concludes.

2 Properties of an optimal variance forecast

As a first step in our analysis, we derive two 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 analyses. Specifically, the Dawid-Sebastiani score (DSS) equals the negative logarithmic score of a Gaussian forecast density $f_\mathcal{N}$ with mean m and variance v, i.e.

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

where y denotes the realization of the target variable.[1] Note that a smaller score corresponds to a better forecast. Consider forecasting the parameters m and v of a random variable Y, conditional on some information set $\mathcal{I}$. Then $\mathbb{E}\left[DSS(m, v, Y)|\mathcal{I}\right]$ 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 Dawid-Sebastiani score 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. Note that the Dawid-Sebastiani score focuses on the first two moments of f exclusively. As a consequence, a density forecast with misspecified higher order moments

---

[1]In the literature, the equivalent score $\log(v) + \frac{(y-m)^2}{v}$ is usually used. However, we stick to the variant in Equation (1) for better comparability with the logarithmic score in our Monte Carlo simulations.

may perform equally well as the correct density forecast, but not strictly better. In the terminology of Gneiting and Raftery (2007), the Dawid-Sebastiani score is a proper but not a strictly proper scoring rule.

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 important implications in this context: First, V and S should be equal on average, i.e. $\mathbb{E}[S] = \mathbb{E}[V]$. Throughout the paper, we refer to the unconditional expectation $\mathbb{E}[S]$ as the mean squared forecast error (MSFE). Second, from the law of total covariance, the requirement that $V = \mathbb{E}[S|\mathcal{I}]$ implies that $\mathbb{C}ov[V, S] = \mathbb{V}\left[\mathbb{E}[S|\mathcal{I}]\right] \geq 0$. The inequality is strict only in the presence of predictable heteroskedasticity, because in this case $\mathbb{E}[S|\mathcal{I}]$ varies with $\mathcal{I}$.

3 The linear pool's variance forecast: Baseline example

We next provide a simple example which shows that the LP is likely to violate at least one of the above-mentioned desiderata of an optimal variance forecast, and can easily violate both. Consider a variable Y determined 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 $0 < \omega_1 < 1$ 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 and squared forecast errors are identical in expectation. Second, the covariance between each variance forecast and the corresponding squared error is equal to the variance of the expected (conditional) squared forecast error. Due to the homoskedasticity in this simple example, the latter two terms are in fact equal to zero.

The LP's variance forecast is of the form $V_{lp} = a + D$, where a is a constant and

3

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	$\mathbb{E}[S_i]$	$= (1-\rho^2)\sigma_X^2 + \sigma_U^2$
Linear pool		
Mean forecast	M_{lp}	$= \omega_1 M_1 + (1-\omega_1)M_2$
Disagreement	D	$= \omega_1(M_{lp} - M_1)^2 + (1-\omega_1)(M_{lp} - M_2)^2$
Variance forecast	V_{lp}	$= \omega_1\left(V_1 + (M_{lp}-M_1)^2\right) + (1-\omega_1)\left(V_2 + (M_{lp}-M_2)^2\right)$
		$= V + D$
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	$\mathbb{C}ov[D,S]$	$= 2\omega_1(1-\omega_1)(2\omega_1-1)^2(1-\rho)^2(1+\rho)^4\sigma_X^4$

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.

$D = \omega_1(M_1 - M_{lp})^2 + (1-\omega_1)(M_2 - M_{lp})^2$ is the well-known measure of disagreement between the two point forecasts.[2] Strikingly, the LP's variance V_{lp} fails at least one of the requirements mentioned in Section 2, and it fails both in a special but important case. First, the LP's expected variance, $\mathbb{E}[V_{lp}]$ exceeds its MSFE, $\mathbb{E}[S]$, for all admissible values of ω_1. The LP can therefore be labeled underconfident. The disagreement term D, which is positive-valued, contributes to the LP's underconfidence. Second, the LP's variance V_{lp} has no predictive content for its squared forecast error in the important case $\omega_1 = 0.5$. With this value of ω_1, the covariance between both quantities equals zero, and it can also be shown that D is independent of S (see Appendix A). Note that $\omega_1 = 0.5$ is a popular default 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 thus be regarded as a noise term which deteriorates the LP's variance forecast.

---

[2]See, for instance, Wallis (2005) for the decomposition of the LP's variance into the weighted average variance component and the disagreement component.

4 Bias in the linear pool's variance forecast: General case

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

$$m_{lp} = \sum_{i=1}^n \omega_i m_i \tag{4}$$

$$v_{lp} = \sum_{i=1}^n \omega_i v_i + d \tag{5}$$

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. In what follows, we will mostly take the mean specification in (4) as given, and investigate the properties of the variance specification in (5) conditional on (4). Moreover, we will consider a simple modification, the centered linear pool (CLP), with $m_{clp} = m_{lp}$ and

$$v_{clp} = \sum_{i=1}^n \omega_i v_i \tag{6}$$

$$= v_{lp} - d.$$

Hence, the CLP has the same mean forecast as the LP, but its variance forecast does not contain the disagreement term. 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 simply relocated such that its mean equals $m_{clp} = m_{lp}$ instead of m_i before being combined.

Equations (4) to (6) 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 $\{M_i, V_i\}_{i=1}^n$. Randomness in the forecasts may reflect both conditioning information and noise. To evaluate the performance of a combination method, it is necessary to make assumptions about the joint distribution of the underlying individual forecasts and the target variable Y. We first consider a situation in which the individual forecasts imply a correct assessment of their own uncertainty, as formulated in the following assumption.

A1 The individual variance forecasts $\{V_i\}_{i=1}^n$ are unconditionally unbiased, i.e. they satisfy $\mathbb{E}[V_i] = \mathbb{E}\left[(Y - M_i)^2\right], i = 1, \ldots, n$.

As mentioned earlier, assumption A1 is implied by the optimality condition of the

Dawid-Sebastiani scoring rule in Equation (3). However, A1 imposes only an unconditional notion of unbiasedness, and is hence much weaker than Equation (3). Denoting the squared error of the combined mean forecast $\sum_{i=1}^{n} \omega_i M_i$ by S, we have the following first result.

Proposition 4.1. *Consider any joint distribution of M, V and Y. Under A1 and assuming that the combination weights ω_i are positive and sum to one, it holds that*

$$\mathbb{E}[V_{lp}] > \mathbb{E}[V_{clp}] > \mathbb{E}[S],$$

i.e. the variance of both the LP and the CLP is upward biased, with the bias being larger for the LP.

Proof. Denote by $M_{lp} = \omega' M$ and $V_{lp} = \omega' V + D$ the mean and variance of the LP, where $D = \sum_{i=1}^{n} \omega_i (M_i - M_{lp})^2$ is the disagreement term and ω is a vector containing the combination weights ω_i for $i = 1, \ldots, n$. Similarly, let $M_{clp} = M_{lp}$ and $V_{clp} = \omega' V$ denote the mean and variance of the CLP, and define $S = (Y - M_{lp})^2 = (Y - M_{clp})^2$. We have that

$$\begin{aligned}
\mathbb{E}\left[(Y - \omega' M)^2\right] &< \mathbb{E}\left[\sum_{i=1}^{n} \omega_i (Y - M_i)^2\right] \\
&= \sum_{i=1}^{n} \omega_i \, \mathbb{E}\left[(Y - M_i)^2\right] \\
&= \sum_{i=1}^{n} \omega_i \, \mathbb{E}[V_i], \\
&= \mathbb{E}[V_{clp}], \\
&< \mathbb{E}[V_{lp}] = \mathbb{E}[V_{clp}] + \mathbb{E}[D],
\end{aligned}$$

where the key inequality in the first line follows from the convexity of the square function. ∎

The result shows that the variance forecasts of the LP and the CLP are upward biased, even though the individual variance forecasts are unbiased. This finding is similar in spirit to Gneiting and Ranjan (2013, Theorem 3.1(c)) who 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 PIT which depends on the entire forecast density. Of course, considering the entire density is appealing in principle. Without further restrictions, however, it can be hard to identify what drives a density's dispersion as measured by the PIT. Our approach of defining a density's uncertainty in terms of its variance allows us to sharpen the result of Gneiting and Ranjan (2013) by identifying disagreement as a variance-bias augmenting term.[3]

---

[3] Focusing only on the first two moments can be justified empirically at least in macroeconomic applications. For example, Clark and Ravazzolo (2015) find that using non-Gaussian error term distributions does not improve the density forecasting performance of autoregressive models with time-varying volatility.

Finally, note that the statement of Proposition 4.1 remains true if A1 is replaced by the following, weaker assumption:

A1' The individual variance forecasts $\{V_i\}_{i=1}^n$ satisfy $\mathbb{E}[V_i] \geq \mathbb{E}\left[(Y - M_i)^2\right], i = 1, \ldots, n$.

Hence if the individual densities are underconfident, the LP is underconfident as well.

5 Properties of the linear pool in a prediction space model

We next impose more structure on the joint distribution of the forecasts and realizations in order to derive more specific results and identify drivers of the LP's forecast performance.

5.1 Prediction space model for forecasts and realizations

The joint distribution of forecasts and realizations has already been considered as an important analytical tool in the contributions by Bates and Granger (1969) and Murphy and Winkler (1987). Gneiting and Ranjan (2013) provide a formalization as a 'prediction space' for full density forecasts. Here we consider a simplified variant where each forecast is characterized by a mean and variance only; see Ehm, Gneiting, Jordan, and Krüger (2016, Section 3.1) for further discussion.

Consider a vector of mean forecasts $M \in \mathbb{R}^n$, a vector of variance forecasts $V \in \mathbb{R}_+^n$, an error term $U \in \mathbb{R}$ (see below) and a positive-valued common factor $\eta \in \mathbb{R}_+$ such that

$$\eta^{-1/2} \begin{bmatrix} M \\ U \\ \eta^{-1/2}V \end{bmatrix} \sim \left(\begin{bmatrix} \mu_M \\ 0 \\ \mu_V \end{bmatrix}, \begin{bmatrix} \Sigma_M & 0 & 0 \\ 0 & \sigma_U^2 & 0 \\ 0 & 0 & \Sigma_V \end{bmatrix} \right), \tag{7}$$

where the column on the right-hand side denotes the expectations of the variables, and the matrix denotes the variance-covariance matrix. The random variable η is assumed to be independent of the variables on the left-hand side, i.e. of $\eta^{-1/2}M, \eta^{-1/2}U, \eta^{-1}V$.[4] Without loss of generality, we assume that the expectation of the target variable Y conditional on η is zero. The error term U is defined as the residual of the projection of the target variable Y on the bias-corrected forecasts $M - \sqrt{\eta}\mu_M$, i.e. by

$$Y = (M - \sqrt{\eta}\mu_M)' \gamma + U, \tag{8}$$

where γ is the vector of coefficients resulting from this projection. The conditional distri-

---

[4]This assumption could equivalently be expressed by writing M_i and U as products of η and a random variable that is independent of η, and by writing V_i as a product of η^2 and an independent random variable.

bution of the original variables thus is given by

$$
\left[\begin{array}{c} Y \\ M \\ U \\ V \end{array}\,\Bigg|\,\eta\right] = \left(\left[\begin{array}{c} 0 \\ \sqrt{\eta}\mu_M \\ 0 \\ \eta\mu_V \end{array}\right], \left[\begin{array}{cccc} \eta\sigma_Y^2 & \eta\gamma'\Sigma_M & \eta\sigma_U^2 & 0 \\ \eta\Sigma_M\gamma & \eta\Sigma_M & 0 & 0 \\ \eta\sigma_U^2 & 0 & \eta\sigma_U^2 & 0 \\ 0 & 0 & 0 & \eta^2\Sigma_V \end{array}\right]\right). \tag{9}
$$

Note that η acts as a common factor which scales the mean vectors of M and V, as well as the model's covariance and variance terms (except the ones that are restricted to zero). Albeit simple, such a common factor specification of time-varying uncertainty is in line with the empirical results of Carriero, Clark, and Marcellino (2016) for many macroeconomic and financial variables. Without loss of generality, we assume that $\mathbb{E}[\eta] = 1$. Note that we impose few restrictions on the process that generates η. For example, η could be a discrete variable that takes values of 0.5 and 1.5 with probability one half each. As another example, η could follow an autocorrelated time series process. The only requirement we impose is that $\mathbb{E}[\eta] = 1$ and $\mathbb{V}[\eta] \equiv \sigma_\eta^2$ exist; hence, the process generating η must be stationary. That said, none of our subsequent results depends on any properties of this process other than σ_η^2.

By varying the biases in μ_M, the weight vector γ and the variance of U, the framework can accommodate a wide range of scenarios, with mean forecasts M ranging from poor to precise. Finally, we assume that the variance forecasts V are conditionally uncorrelated with M and U. Hence, our setup does not cover situations where changes in uncertainty affect the mean of the target variable, like, for instance, in GARCH-in-mean models.

5.2 Bias in the linear pool's variance forecast

The following result uses the prediction space framework to derive a specific characterization of the bias in the LP's variance forecast.

Proposition 5.1. *Under A1, it holds that* $\mathbb{E}[S] = \mathbb{E}[\omega'V] - \mathbb{E}[D]$. *Hence the LP and CLP systematically deviate from S according to*

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

Proof. See Appendix C. ∎

Proposition 5.1 quantifies the relationship between the LP's expected squared forecast error, $\mathbb{E}[S]$, and the LP's variance forecast, $\mathbb{E}[V_{lp}]$. As shown in Section 2, both quantities should be equal in optimum. The proposition thus provides a precise assessment of the LP's underconfidence.[5]

---

[5]Proposition 5.1 is related to a result by Lahiri and Sheng (2010) who, however, work under a different set of assumptions. Their Equation (8) yields the same statement as our Proposition 5.1 if the number of

Propositions 4.1 and 5.1 have the same qualitative interpretation: If the individual variance forecasts are unbiased, then both the LP and the CLP are underconfident. However, Proposition 5.1 exploits the prediction space framework to make a precise quantitative statement on the pools' underconfidence.

5.3 Disagreement encompassed by weighted average variance forecasts

We next derive a result that speaks to the desideratum that a variance forecast should be highly correlated with the squared errors of the mean forecast (see Section 2). Specifically, consider the following linear regressions:

$$S = a_0 + a_1 D + a_2 \; \omega'V + \text{error} \tag{10}$$

and

$$S = b_1 D + b_2 \; \omega'V + \text{error} \tag{11}$$

where (11) is aimed at the case of constant variance forecasts $\omega'V$, such that (10) cannot be employed. The LP entails the assumptions that $a_0 = 0, a_1 = a_2 = 1$ or, alternatively, that $b_1 = b_2 = 1$. We are particularly interested in the coefficients a_1 and b_1 which quantify the contribution of disagreement to the forecast of the squared error. The population coefficient of a_1 equals

$$
\begin{aligned}
a_1 &= \frac{\mathbb{C}ov[S, D - \tilde{\alpha} - \tilde{\beta} \; \omega'V]}{\mathbb{V}[D - \tilde{\alpha} - \tilde{\beta} \; \omega'V]} \\
&= \frac{\mathbb{C}ov[S, D] - \tilde{\beta} \; \mathbb{C}ov[S, \omega'V]}{\mathbb{V}[D] + \tilde{\beta}^2 \omega'\mathbb{V}[V]\omega - 2\tilde{\beta} \; \mathbb{C}ov[D, \omega'V]},
\end{aligned}
$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are the population coefficients from a linear regression of D on a constant and $\omega'V$.[6] A similar expression holds for b_1.

The following result states conditions under which $a_1 = b_1 = 0$ holds, corresponding to a particularly stark violation of the LP's implicit assumption.

Proposition 5.2. *Assume that*

A2 The forecast biases are all identical, i.e. $\mu_M = \begin{bmatrix} \kappa, & \dots, & \kappa \end{bmatrix}'$ for some $\kappa \in \mathbb{R}$.

A3 The combination weights minimize $\mathbb{E}[S]$ subject to the constraint of adding up to one (Bates and Granger, 1969).

A4 The joint distribution of M and U conditional on η is normal.

A5 It holds that $\Sigma_V = 0$.

---

forecasters (N in their notation) diverges to infinity; in that case, their expression $\sigma^2_{\lambda|th}$ corresponds to our $\mathbb{E}[S]$. However, their statement is different from ours for any fixed value of N. Furthermore, Lahiri and Sheng (2010) do not consider implications for the combination of density forecasts.

[6]The formula remains the same, except that $\mathbb{C}ov[A, B]$ is replaced by $\mathbb{E}[AB]$ and $\mathbb{V}[A]$ is replaced by $\mathbb{E}[A^2]$).

Then, $a_1 = b_1 = 0$, i.e. $\omega'V$ encompasses disagreement in the prediction of S.

Proof. See Appendix C. ∎

    Assumption A2 restricts the biases in the mean forecasts M to be equal, with unbiased mean forecasts ($\kappa = 0$ in A2) being an important special case. From an empirical perspective, A2 can be motivated by the fact that different model-based forecasting methods often rely on a similar sample of data for parameter estimation. Assumption A3 restricts the way in which a researcher derives the combination weights ω. In the context of density forecasting, the restriction of adding up to one is necessary to guarantee that the combined object is a density.[7] Since these weights minimize the MSFE, we refer to them as the MSFE-optimal weights ω^*. Note that ω^* can have negative elements which could, in principle, lead to a negative variance prediction by the LP. We disregard the latter case which seems of minor applied relevance. We also note that ω^* is optimal in terms of the Dawid-Sebastiani score (given the prior adding-up-constraint), and is hence consistent with our setup.[8] Assumptions A4 and A5 impose restrictions on the prediction space, conditional on the common factor η. However, the prediction space retains considerable flexibility by specifying alternative distributions of η.[9]

    While the assumptions underlying Proposition 5.2 are restrictive, they do not appear implausible from an empirical point of view. Furthermore, note that the case covered by the proposition is a fairly drastic one, in that disagreement plays no role whatsoever, which is a particularly stark contrast to the LP's assumption. Below we present numerical results which show that similar results (disagreement having limited but nonzero predictive content) hold under a broader set of conditions.

5.4   Choice of combination weights

The magnitude of the LP's variance bias (discussed in Propositions 4.1 and 5.1) and the correlation between D and S (discussed in Proposition 5.2) depend on ω, the vector of combination weights. Hence it seems tempting to choose ω in a way that limits the drawbacks of the LP mentioned above. Unfortunately, this approach is not feasible in general: ω also affects the LP's mean forecast, and a given choice of ω can generate good mean forecasts but poor variance forecasts or vice versa. The following result describes such a situation.

Proposition 5.3. *Assume that A1 holds, and assume that*

    A6 All individual mean forecasts have the same MSFE, i.e. $\mathbb{E}[(Y - M_i)^2] = c$ for some $c > 0$ and all $i = 1, \dots, n$.

---

[7]In combinations of point forecasts, the same restriction is also commonly employed.

[8]As noted in Section 2, minimizing the expected Dawid-Sebastiani score requires the researcher to first select an MSFE-optimal point forecast, and then select the corresponding variance. The Bates and Granger (1969) minimization problem solves the first step, given the prior constraint that the weights add up to one.

[9]Moreover, A4 could be relaxed slightly at the cost of a more demanding exposition. In fact, U can be non-normal if certain moments of U are uncorrelated with certain moments of M conditional on η. Normality of M conditional on η continues to be required.

Then the MSFE-optimal combination weights ω^* maximize the upward bias of the LP's variance.

Proof. See Appendix C. ∎

Proposition 5.3 presents a simple but empirically relevant scenario in which the weights that minimize the MSFE of the LP's mean forecast also maximize the upward bias of the LP's variance forecast. Assumption A6 requires that the mean forecasts to be combined have the same MSFE, which holds approximately in many applications. An interesting special case of A6 arises when all forecasts are unbiased and all pairwise correlations between the forecast errors are identical, i.e. $\mathbb{C}or(Y - M_i, Y - M_j) = \rho$ for all pairs of forecasts i, j with $i \neq j$. In this case, equal weights are optimal in terms of MSFE (Timmermann, 2006, Section 2.4). Motivated by the good performance of equal combination weights in applications, the latter case has received much attention in the literature (e.g. Smith and Wallis, 2009; Elliott, 2011; Claeskens, Magnus, Vasnev, and Wang, 2016).

If the assumptions of Propositions 5.1 and 5.2 apply, it is clear that the MSFE-optimal weights ω^* are not optimal for the LP's variance because the disagreement component is a bias-augmenting noise term. Since disagreement and, consequently, the bias equals zero 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 $i \in \{1, \ldots, n\}$, 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 $\iota_{[i]}$ exceeds the corresponding loss in mean forecast accuracy for any forecast i. 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 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.

5.5 Disagreement as a proxy for weighted average variance

Our results up to this point present conditions under which disagreement is detrimental to the LP's performance. These results refer to the situation where both disagreement D and the weighted average variance $\omega'V$ are available, as is clearly the case when combining various density forecasts. In the distinct situation where only mean forecasts are available, however, disagreement is sometimes considered as a simple proxy for predictive uncertainty. The economic policy uncertainty index provided by Baker, Bloom, and Davis (2019), which considers disagreement among participants of the Federal Reserve Bank of Philadelphia's SPF, is a prominent example. The following result states conditions under which disagreement is a viable proxy of the weighted average variance in our theoretical framework.

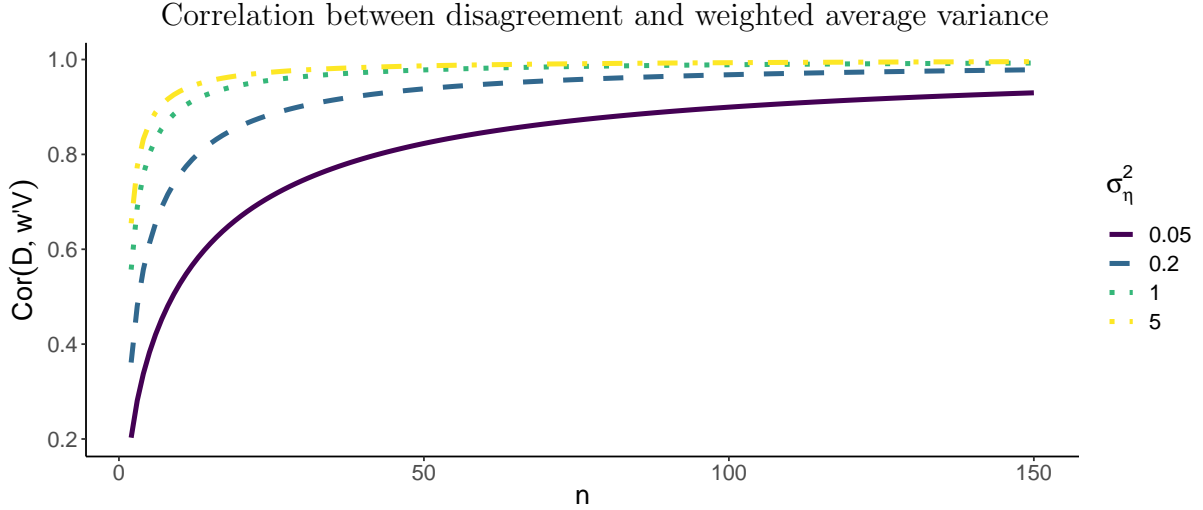**Proposition 5.4.** *Assume that A5 holds, and that*

Figure 1: Correlations of disagreement D and weighted average variance $\omega'V$ for different values of σ_η^2 and n. Combination weights and parameters of prediction space are given by $\omega = n^{-1}\iota$, $\gamma = \iota$, $\Sigma_M = I_n$, where ι is an $n \times 1$ vector of ones and I_n is the identity matrix of size n. η follows a gamma distribution with $\mathbb{E}[\eta] = 1$ and $\mathbb{V}[\eta] = \sigma_\eta^2$.

A2' *All mean forecasts are unbiased, i.e. $\mu_M = 0\,\iota$.*

A7 *$\omega = n^{-1}\iota$ and $\gamma = \tilde{\gamma}\,\iota$, where ι is an $n \times 1$ vector of ones and $\tilde{\gamma}$ is a scalar.*

A8 *Σ_M is such that its diagonal elements are all equal to σ^2, and its off-diagonal elements are all equal to θ.*

Then if the number of forecasters n goes to infinity, it holds that $\mathbb{C}or(D, S)/\mathbb{C}or(\omega'V, S) \to 1$ and that $\mathbb{C}or(D, \omega'V) \to 1$.

Proof. See Appendix C. ∎

The result shows that under certain conditions, disagreement and weighted average variance are equivalent for large n, in the sense that their correlation converges to one and they become equally correlated with S, the squared forecast error of the combination. From a practical perspective, the proposition thus provides a justification for using D as a proxy for $\omega'V$ when the latter is not available and the number of forecasters is large enough. Figure 1 illustrates the proposition by plotting the correlation between D and $\omega'V$. The correlation increases in n and σ_η^2, the variance of the common factor η. Conversely, the figure shows that the correlation of D and $\omega'V$ tends to be low if only few forecasts are available and if σ_η^2 is small, such that fluctuations in disagreement reflect noise rather than fluctuations in η. The result that σ_η^2 needs to be large enough in order to render disagreement a good proxy for uncertainty is supported by the empirical findings of Boero, Smith, and Wallis (2015, p. 1044) who state that "[..] the joint results from the US and UK surveys suggest the encompassing conclusion that disagreement is a useful proxy for uncertainty when it exhibits large fluctuations [..]".

While certainly restrictive, A2', A7 and A8 may be useful assumptions to make when using a survey of forecasters like the SPF. In such a survey, n may be relatively large,

and making simplifying assumptions along the lines of A2', A7 and A8 is often a sensible strategy in order to reduce estimation noise. In particular, Capistrán and Timmermann (2009) show that entry and exit of forecasters makes it challenging to identify more refined correlation structures among the individual point forecasts.

6   Monte Carlo simulations

Here we illustrate our results via simulation examples. For simplicity, we fix the factor η in Equation (9) at a value of one, which corresponds to the limiting case that $\mathbb{V}[\eta] \to 0$.[10] Specifically, we simulate variants of the baseline example in Section 3, where the target variable is given by

$$Y = X_1 + X_2 + U,$$

with

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

In all simulations, U is normally distributed. Using the notation of Section 5, the mean forecast of forecaster i is given by

$$M_i = X_i$$

such that equation (8) holds with $\gamma = \begin{bmatrix} 1, & 1 \end{bmatrix}'$ and $\mu_M = \begin{bmatrix} 0, & 0 \end{bmatrix}'$. Moreover, Σ_M equals

$$\Sigma_M = \begin{bmatrix} 1 & 0 \\ 0 & \sigma^2_{X_2} \end{bmatrix}.$$

The corresponding variance forecasts equal

$$\begin{aligned} V_1 &= \sigma^2_{X_2} + \sigma^2_U = \sigma^2_{X_2} + 1 \\ V_2 &= \sigma^2_{X_1} + \sigma^2_U = 2. \end{aligned}$$

Thus, the conditional moments of V are described by (9) with $\mu_V = \begin{bmatrix} V_1, & V_2 \end{bmatrix}'$, $\Sigma_V = 0$, and $\eta = 1$. Hence, each forecaster produces mean and variance forecasts that are ideal conditional on her information set, thereby fulfilling assumption A1 of unconditional unbiasedness. The combined mean forecast of each combination scheme to be considered equals

$$M_c = \omega' M = \omega_1 M_1 + (1 - \omega_1) M_2,$$

where ω_1 is the weight for the first forecast.

In the Monte Carlo simulations, 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

---

[10]We also experimented with more general scenarios, using various data generating processes for η. However, these scenarios offer little additional insight in the present context, motivating our simple choice.

the theoretical results. We denote the ith density forecast by

$$f_i\left(M_i, V_i\right)$$

where M_i denotes the mean and V_i the variance of f_i. The density f_i does not need to be normally distributed, but we suppress the potential dependence on additional parameters in our notation. The density of the LP is given by

$$f_{lp} = \omega_1 f_1\left(M_1, V_1\right) + \left(1 - \omega_1\right) f_2\left(M_2, V_2\right)$$

whereas the density of the CLP equals

$$f_{clp} = \omega_1 f_1\left(M_c, V_1\right) + \left(1 - \omega_1\right) f_2\left(M_c, V_2\right).$$

Thus, the CLP relocates both individual density forecasts at M_c, and then combines the relocated densities linearly. Finally, the density of the VULP is

$$f_{vulp} = \omega_1 f_1\left(M_c, V_1 - \mathbb{E}\left[D\right]\right) + \left(1 - \omega_1\right) f_2\left(M_c, V_2 - \mathbb{E}\left[D\right]\right).$$

Hence, in addition to relocating, the VULP rescales both densities such that the individual variance forecasts are reduced by $\mathbb{E}\left[D\right]$. Each of these pools implies a mean and a variance forecast. While all three combined densities have the same mean forecast (i.e. $M_{lp} = M_{clp} = M_{vulp} = M_c$), the respective variance forecasts differ. The LP produces the variance forecast

$$
\begin{aligned}
V_{lp} &= \omega'V + D \\
&= \omega_1\left(\sigma_{X_2}^2 + 1\right) + \left(1 - \omega_1\right)2 + D
\end{aligned}
$$

with

$$D = \omega_1\left(M_1 - M_c\right)^2 + \left(1 - \omega_1\right)\left(M_2 - M_c\right)^2.$$

The CLP yields

$$
\begin{aligned}
V_{clp} &= \omega'V \\
&= \omega_1\left(\sigma_{X_2}^2 + 1\right) + \left(1 - \omega_1\right)2.
\end{aligned}
$$

Using the VULP results in

$$V_{vulp} = \omega'V - \mathbb{E}\left[D\right]$$

with

$$\mathbb{E}\left[D\right] = \omega_1\left(1 - \omega_1\right)\left(\sigma_{X_2}^2 + 1\right).$$

Note that V_{clp} and V_{vulp} are constant, whereas V_{lp} contains a stochastic component.

In the first case considered, we employ $\sigma_{X_2}^2 = 1$.[11] The weight ω_1 for the first forecast ranges from 0 to 1, and the MSFE-optimal weight is $\omega_1^* = 0.5$. As shown in the lower left panel of Figure 2, the variance forecasts have the property described in Proposition 5.1. The MSFE of the combined forecast equals $\mathbb{E}[S] = V_{vulp}$. The CLP's constant variance,

---

[11]This setup corresponds to the baseline example in Section 3 with $\rho = 0$.

V_{clp}, lies halfway between the constant optimal variance V_{vulp} and the LP's expected variance $\mathbb{E}[V_{lp}]$. The difference between the CLP's variance and the LP's variance is given by $V_{clp} - \mathbb{E}[V_{lp}] = \mathbb{E}[D]$. In line with Proposition 5.3, $\mathbb{E}[D]$ reaches its maximum at ω_1^*, implying that the largest bias in the LP's variance forecasts occurs when MSFE-optimal weights are used.

The regression coefficients displayed in the lower right panel at ω_1^* illustrate the result of Proposition 5.2. The coefficient for disagreement equals zero, because disagreement D has no explanatory power for the squared error S beyond what is contained in the combined variance forecast $\omega'V$ which is constant here. Since $\omega'V$ is larger than $\mathbb{E}[S]$, its coefficient is smaller than one. A coefficient equal to one for D, as is implicitly used by the LP, is only observed for weights which differ considerably from ω_1^*. The regression coefficient for $\omega'V$ is almost constant across all values of ω_1.

The scoring rules displayed in the two upper panels of Figure 2 show that all pools perform best when using the weight ω_1^*. The LP performs worse than the CLP and the VULP at ω_1^* and for a wide range of weights around ω_1^*. The VULP outperforms the two other pools for most values of ω_1. Concerning the Dawid-Sebastiani score and the weight choice $\omega_1 = \omega_1^*$, the superiority of the VULP follows from the fact that its mean and variance forecasts satisfy the optimality restrictions in Equations (2) and (3). For values of ω_1 close to zero or one, the differences in the scores tend to be small. With the current setting, the CLP and the VULP yield normal densities, leading to the equality of the Dawid-Sebastiani score and the logarithmic score. The LP produces non-normal densities.

The second case we consider is equal to the first case except for the value of $\sigma_{X_2}^2$ which now equals 1.5. Thus, f_2 has a lower variance than f_1, and M_2 produces a lower MSFE than M_1. Therefore, the weighted average variance forecast $\omega'V$ of the CLP displayed in Figure 3 increases with ω_1. The variance forecast of the VULP is minimal at $\omega_1^* = 0.4$ which is the MSFE-optimal weight. The bias term $\mathbb{E}[D]$ continues to have its maximum at $\omega_1 = 0.5$.[12] As implied by Proposition 5.2, the regression coefficient for D equals zero at ω_1^*. The regression coefficient for $\omega'V$ slowly decreases with ω.

The optimal weights differ for each pool. For the VULP, the minimal Dawid-Sebastiani 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 Dawid-Sebastiani score 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 logarithmic score, where the optimal weights for the VULP and the CLP are virtually the same as for the Dawid-Sebastiani score, although the 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 empirical applications. Like in the first case, with respect to both scores the LP performs worst for a wide range of weights around ω_1^*, and the VULP performs best.

---

[12]Note that Assumption A6 underlying Proposition 5.3 does not hold in this case, explaining why the bias in the LP's variance is not maximized at ω_1^*.

In the third case, 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 misspecified in terms of their functional form since they are rescaled t-distributions, 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. Concerning the weighted average variance forecasts displayed in Figure 4, the results are identical to those from the first case. Yet the regression coefficients displayed in the lower right panel differ because the assumption of joint normality of $\begin{bmatrix} M_1, & M_2, & U \end{bmatrix}'$ required by Proposition 5.2 is violated. As a consequence, at ω_1^*, D contains information beyond that contained in $\omega'V$. However, the coefficient for D still has its minimum at ω_1^*. Again, the coefficient for $\omega'V$ is relatively stable across all values of ω.

For each pool, the Dawid-Sebastiani score and the logarithmic score differ because all pooled 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 is narrower than in the first case, but still covers more than the central 50% of all weights considered. The VULP outperforms the CLP with respect to the Dawid-Sebastiani score. For the logarithmic score, the VULP and CLP attain similar values, with the CLP performing marginally better.

7 Empirical case studies

We next investigate the properties of the LP and CLP in two case studies from macroeconomics and finance. In both case studies, we construct measures of forecast uncertainty from time series models with stochastic volatility.

7.1 Forecasting inflation

Here we employ inflation forecasts from the univariate unobserved component model with stochastic volatility (UCSV) by Chan (2013) and from the bivariate unobserved component model with trends and cycles (biUC) by Chan, Koop, and Potter (2016). The UCSV model is closely related to the model in Stock and Watson (2007), but the stochastic volatilities evolve as AR(1)-processes instead of random walks. Trend inflation in the biUC model is bounded between 0% and 5%, and the non-accelerating inflation rate of unemployment (NAIRU) is bounded between 4% and 7%. See Chan (2013) and Chan et al. (2016) for all details on the Bayesian estimation methodology and prior choices.[13] The models are estimated recursively using the annualized quarterly growth rate of the GDP deflator and the unemployment rate of the US from 1948Q1 to 2018Q2. The evaluation sample starts in 1964Q4. We investigate forecasts for one and two quarters ahead. We denote the weight for the UCSV model by ω_1.

---

[13]Our codes are based on the MATLAB codes kindly made available by Joshua Chan at `http://joshuachan.org/code/code_MASV.html` and `http://joshuachan.org/code/code_biUC_bounded.html`.
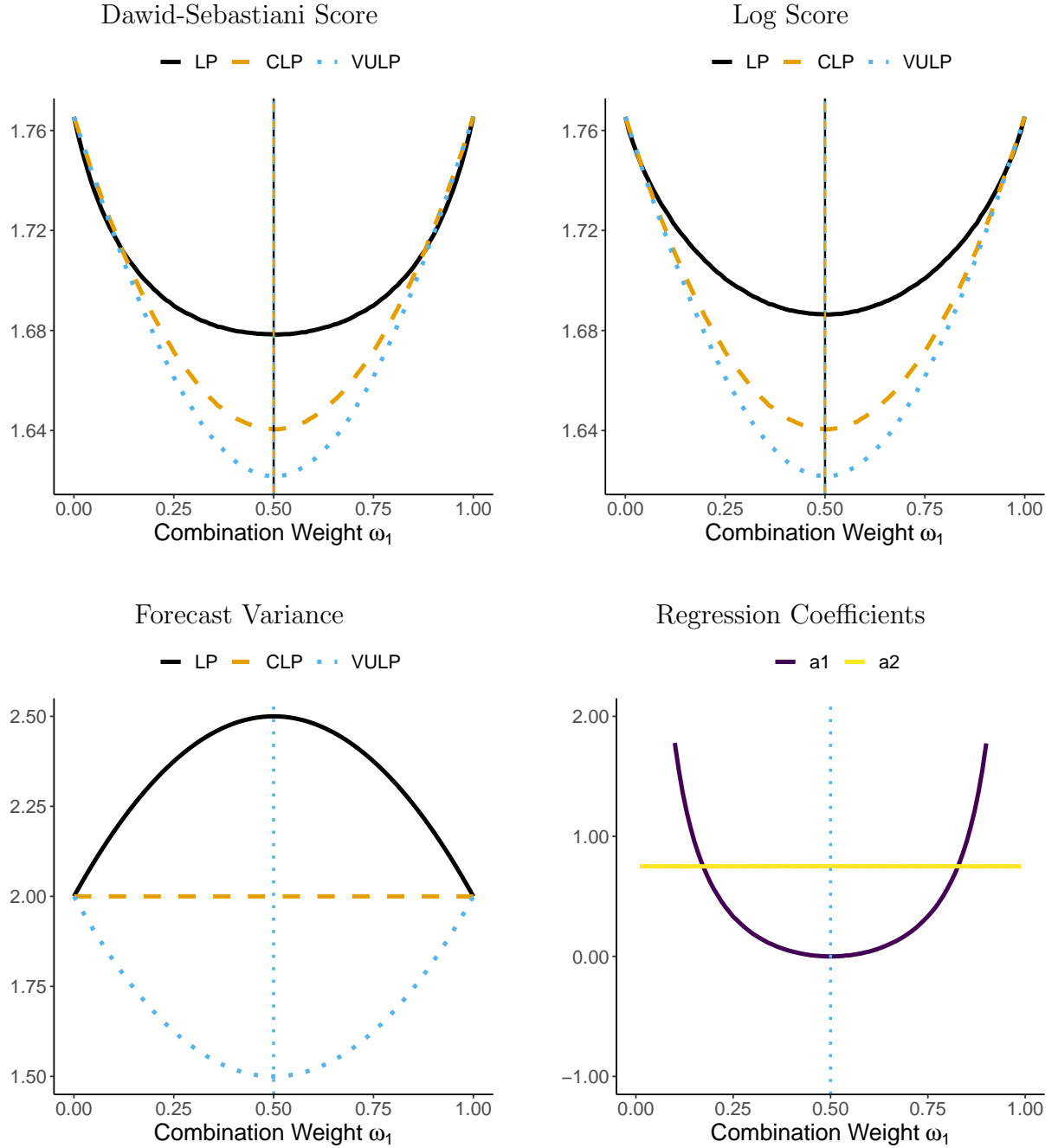
Figure 2: Monte Carlo simulations, case 1 (Gaussian forecasts, $\sigma^2_{X_2} = 1$). The top row shows the Dawid-Sebastiani and logarithmic scores (log scores) for the linear pool (LP), the centered linear pool (CLP) and the variance-unbiased linear pool (VULP), plotted against the combination weight ω_1. A lower score indicates a more accurate forecast. The vertical lines indicate the optimal weights for each pool considered. The optimal weight for the VULP (shown in blue) is also optimal in terms of MSFE. The bottom row shows the forecast variance and the regression coefficients in Equation (10), again plotted against ω_1. The vertical lines indicates the VULP- and MSFE-optimal weight. All results are based on 10,000 simulations and 10,000 observations for Y in each simulation.
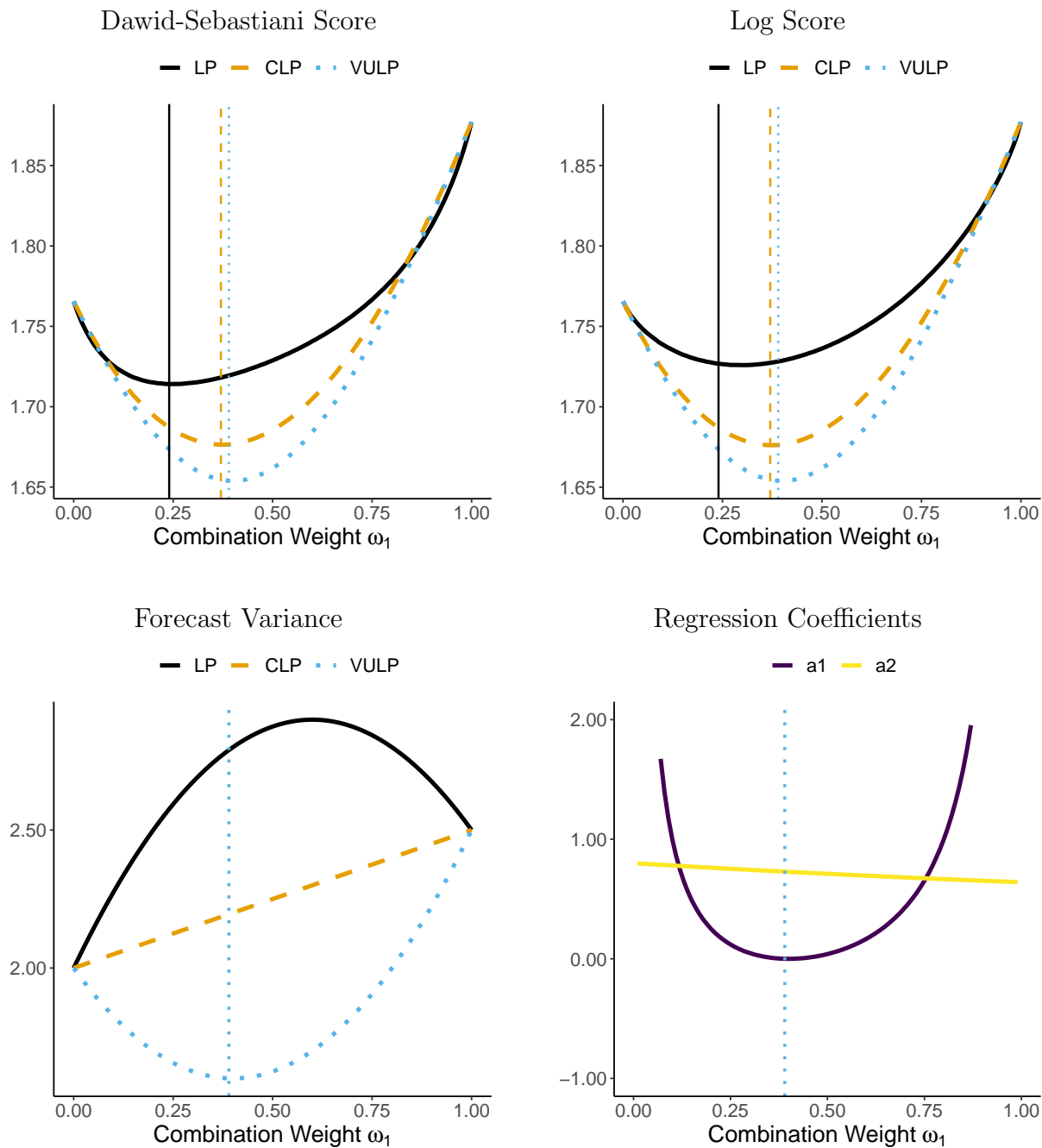
Figure 3: Monte Carlo simulations, case 2 (Gaussian forecasts, $\sigma^2_{X_2} = 1.5$). Figure design analogous to Figure 2.
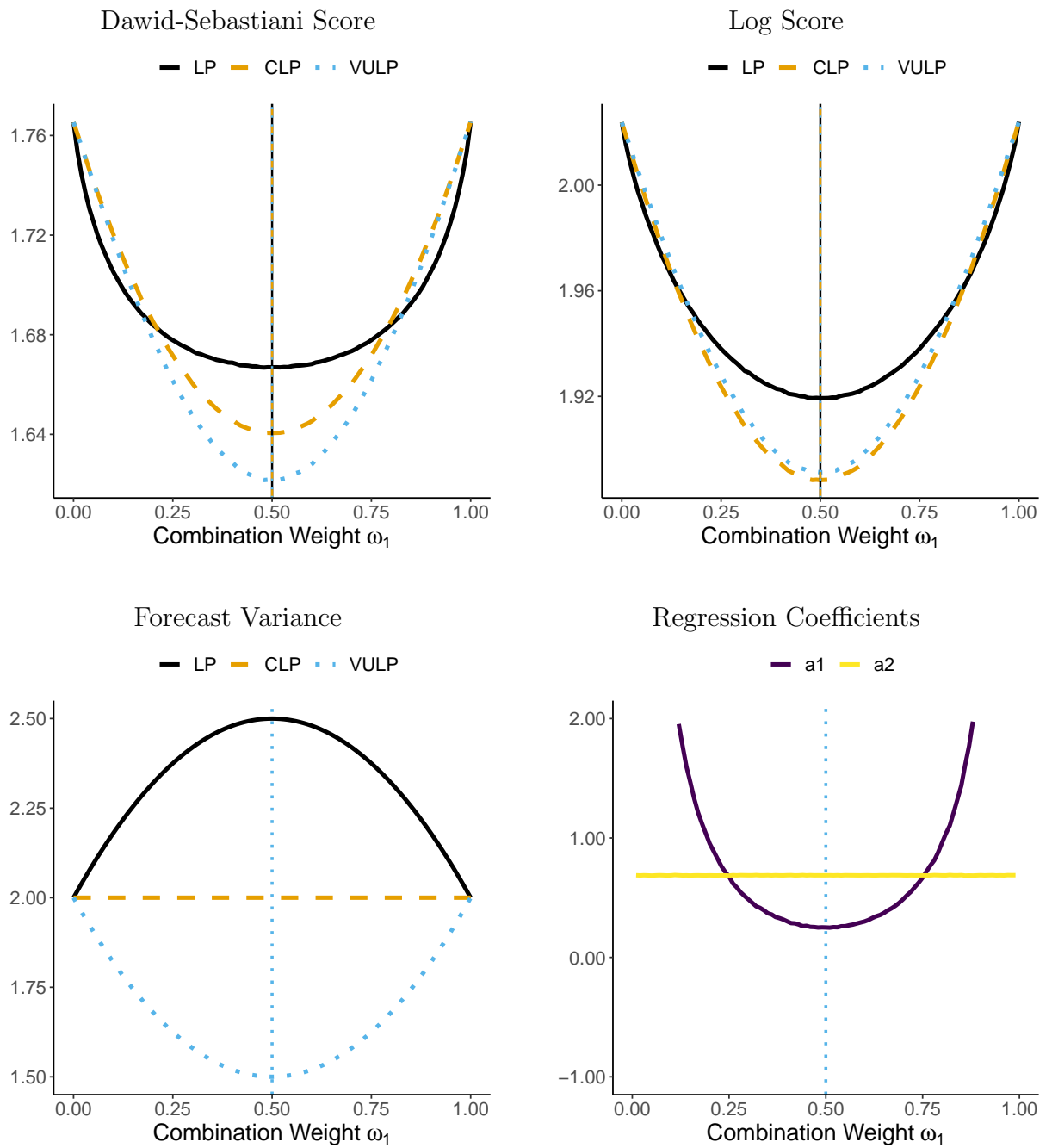
Figure 4: Monte Carlo simulations, case 3 (rescaled t-distributed density forecasts, $\sigma^2_{X_2} = 1$). Figure design analogous to Figure 2.

Figure 5 presents the forecasts for the mean and the variance of inflation. The forecasts of both models are more strongly correlated for $h = 1$ than for $h = 2$, and the biUC model tends to forecast larger variances especially around 1980. Figure 6 displays results of the LP and the CLP for all positive weights.[14] The middle row of the figure summarizes the pools' variance forecasts and their MSFE. The results for the corner 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 biUC model. Furthermore, the MSFEs of the pools' mean forecasts attain their smallest values at $\omega_1^* = 0.48$ for $h = 1$ and at $\omega_1^* = 0.32$ for $h = 2$.

The regression coefficients of Equation (10) displayed in the lower row of Figure 6 exhibit a dependence on the weights which is similar to the Monte Carlo simulations. The coefficient of $\omega'V$ is positive, smaller than 1, and relatively stable. It decreases with ω_1 due to the stronger variance bias of the biUC model. The coefficient of D reaches values around 1 only for relatively extreme values of ω_1. For weights around ω_1^*, the coefficient is comparatively small and stable. In contrast to the Monte Carlo simulations considered, the coefficient of D becomes negative for a wide range of weights around ω_1^*.

For both horizons, in general, the Dawid-Sebastiani score of the CLP shown in the top row of Figure 6 is lower than the Dawid-Sebastiani score of the LP. The opposite holds for extreme weights only, where the scores of both pools are almost identical. As in the Monte Carlo simulations with unequal MSFEs, the optimal weights of both pools result from the implicit trade-off between mean and variance forecast accuracy. Relative to ω_1^* both pools prefer to put more weight on the model with the lower MSFE, leading to a lower (and thus, less biased) variance forecast. The trade-off between mean and variance accuracy is particularly daunting for the LP, for which disagreement enters the variance equation. As a consequence, the optimal weight for the LP implies a more pronounced deterioration of its mean forecast accuracy. While the optimal weights of the CLP for $h = 1$ and $h = 2$ equal 0.45 and 0.23, respectively, the corresponding optimal weights of the LP are given by 0.40 and 0.12.

Since equal weights are ubiquitously used in practice, we report more detailed results for this case. As shown in Tables 2, the CLP's weighted average variance forecasts exceed the respective MSFEs by roughly 50% for both horizons. Disagreement further aggravates this bias for the LP, accounting for 8% of the LP's variance at $h = 1$ and for 11% at $h = 2$. The superior predictive accuracy of the CLP over the LP is not statistically significant at $h = 1$, but it is for $h = 2$.[15] Table 3 reports on regressions of S on a constant, D and $\omega'V$. For both horizons, neither the positive constants nor the negative coefficients of D are statistically different from 0, whereas the positive coefficients of $\omega'V$ are. These results indicate that D is encompassed by $\omega'V$ in the prediction of the squared forecast errors S. The pairwise correlations of D, S, and $\omega'V$ are all positive, but the correlations of D and S are clearly smaller than those of $\omega'V$ and S.

---

[14]In contrast to the Monte Carlo simulations, there are no results for the variance-unbiased linear pool, since its implementation would require knowledge about $\mathbb{E}[D|\eta]$. Moreover, we focus on the Dawid-Sebastiani score for brevity.

[15]Here and in the following, we refer to two-sided tests at the five percent level when reporting on statistical significance.

Inflation case study

	V_{lp}	$\omega'V$	D	S	DSS_{lp}	DSS_{clp}	DM t-stat
$h = 1$	1.618	1.494	0.124	1.031	1.334	1.327	1.902
$h = 2$	2.485	2.210	0.275	1.502	1.523	1.510	2.652

Return case study

	V_{lp}	$\omega'V$	D	S	DSS_{lp}	DSS_{clp}	DM t-stat
	23.255	23.192	0.063	18.101	3.776	3.776	0.218

Table 2: Sample averages for several variables and Diebold-Mariano (DM) t-statistics for equal predictive accuracy of LP and CLP. A positive sign of the t-statistic indicates a better score of CLP. In both case studies, equal weights are used. The test statistic is based on Newey and West (1987) standard errors, with truncation lag chosen according to Andrews (1991) and Zeileis (2004).

Inflation case study

	(Intercept)	D	$\omega'V$	$\mathbb{C}or(D, S)$	$\mathbb{C}or(\omega'V, S)$	$\mathbb{C}or(D, \omega'V)$
$h = 1$	0.229	-0.397	0.570	0.107	0.386	0.457
	(0.144)	(0.358)	(0.160)			
$h = 2$	0.313	-0.280	0.573	0.105	0.367	0.466
	(0.198)	(0.285)	(0.214)			

Return case study

	(Intercept)	D	$\omega'V$	$\mathbb{C}or(D, S)$	$\mathbb{C}or(\omega'V, S)$	$\mathbb{C}or(D, \omega'V)$
	6.478	14.364	0.462	0.093	0.258	0.279
	(3.462)	(40.738)	(0.110)			

Table 3: On the left, results from regression of squared forecast errors on disagreement and equally weighted average variance forecasts. Robust standard errors in parentheses, see below Table 2 for details. On the right, correlation coefficients between disagreement, equally weighted average variance forecasts, and squared forecast errors.

7.2 Forecasting monthly excess stock returns

As a second case study, we consider predicting monthly excess returns of the S&P 500 via models of the form

$$R_t = \alpha + \beta X_{t-1} + \varepsilon_t, \tag{12}$$

where R_t is the monthly excess return, X_t is one out of 15 predictors considered by Welch and Goyal (2008), and ε_t is an error term with stochastic volatility. We follow Rapach, Strauss, and Zhou (2010) in combining 15 univariate models of the form (12). Each model is estimated via Bayesian methods; see Appendix D for details and prior choices. The models are estimated recursively, with observations ranging back to January 1970. The evaluation period is from January 1990 to December 2015. We only consider the case of equal combination weights due to the larger number of models involved.

Figure 7 presents the return forecasts. In contrast to the previous case study, we find that disagreement D is dwarfed by the average variance forecast $\omega'V$, with the latter exceeding disagreement by a factor of around four hundred on average (see Table 2). This result can be explained by low predictability of the mean excess return, which corresponds to similar mean predictions of the 15 individual models. Since D is very small in magnitude, it hardly hampers the predictive accuracy of the LP, which is statistically indistinguishable from its CLP counterpart. Yet, Table 3 shows that, like in the previous case study, $\omega'V$ encompasses D in the prediction of the squared forecast errors S. Relative to the correlation of $\omega'V$ and S, the correlation of D and S is larger than in the inflation case study. Based on Proposition 5.4, this result might be explained by the fact that more forecasts are pooled here, rendering D a less noisy predictor of S relative to $\omega'V$.

8 Conclusion

In this paper, we have established a general setup for modeling the joint distribution of mean forecasts, variance forecasts, and the variable to be predicted. This setup has allowed us to derive several new results concerning the linear pool's variance forecast. In particular, we have described several scenarios in which disagreement – a component of the linear pool's variance forecast that reflects differences between the individual mean forecasts – is harmful to the linear pool's variance forecast. This has motivated us to propose the centered linear pool, a trivial modification of the linear pool that removes the disagreement component. This proposal is well aligned with macroeconomic applications where available time series are short and prone to structural breaks, such that simple approaches seem promising to avoid overfitting. In more data-rich situations, a natural reaction to the shortcomings of the linear pool would be to consider more flexible combination schemes like the approach of Kapetanios, Mitchell, Price, and Fawcett (2015) or the Bayesian alternatives surveyed in Aastveit et al. (2018, Section 3.4). While such approaches do not target disagreement specifically, they provide general re-calibration schemes at the cost of additional parameters to be estimated.

If the concept of the linear pool is to be used, our results suggest that performance-improving refinements should target the individual densities before they are combined.

Procedures which aim at modifying the aggregated density provided by the linear pool are unable to target the noisy disagreement term only. Therefore, the consideration of optimized time-varying weights as surveyed in Aastveit et al. (2018, Section 3), or the spread-adjusted or beta-transformed linear pools proposed in Gneiting and Ranjan (2013, Sections 3.2 and 3.3) can be expected to yield only limited gains in accuracy unless the disagreement term is relatively small. However, the same methods can be applied to the centered linear pool, and the corresponding gains could actually be larger in that case, for instance because the likelihood function, like the scores considered, becomes less flat with respect to the weights.

Finally, our analysis rests on the assumption that a better proxy of forecast uncertainty (based on the individual models' variance forecasts) than disagreement is available. This assumption is plausible in the context of forecast density combinations. In situations where only point forecasts are available, our Proposition 5.4 formalizes the notion that disagreement may be a reasonable 'second-best', provided that it is calculated from a sufficiently large cross-section of point forecasts. The latter assessment is in line with the approach used in many empirical studies based on macroeconomic survey forecasts.

References

Aastveit, K. A., J. Mitchell, F. Ravazzolo, and H. van Dijk (2018). The Evolution of Forecast Density Combinations in Economics. In *Oxford Research Encyclopaedia in Economics and Finance*. Forthcoming; working paper version available at https://www.econstor.eu/bitstream/10419/185588/1/18069.pdf.

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

Baker, S. R., N. Bloom, and S. J. Davis (2019). Economic policy uncertainty index. Online resource; methodological information available at http://www.policyuncertainty.com/methodology.html (last accessed on January 26, 2019).

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

Billio, M., R. Casarin, F. Ravazzolo, and H. K. van Dijk (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics 177*, 213 – 232.

Boero, G., J. Smith, and K. F. Wallis (2015). The measurement and characteristics of professional forecasters' uncertainty. *Journal of Applied Econometrics 30*, 1029–1046.

Capistrán, C. and A. Timmermann (2009). Forecast combination with entry and exit of experts. *Journal of Business & Economic Statistics 27*, 428–440.

Carriero, A., T. E. Clark, and M. Marcellino (2016). Common drifting volatility in large Bayesian VARs. *Journal of Business & Economic Statistics 34*, 375–390.

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

Chan, J. C., G. Koop, and S. M. Potter (2016). A bounded model of time variation in trend inflation, NAIRU and the Phillips curve. *Journal of Applied Econometrics 3*, 551–565.

Claeskens, G., J. R. Magnus, A. L. Vasnev, and W. Wang (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting 32*, 754–762.

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

Craig, A. T. (1943). Note on the independence of certain quadratic forms. *The Annals of Mathematical Statistics 14*, 195–197.

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

Del Negro, M., R. B. Hasegawa, and F. Schorfheide (2016). Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics 192*, 391–405.

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

Ehm, W., T. Gneiting, A. Jordan, and F. Krüger (2016). Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*, 505–562.

Elliott, G. (2011). Averaging and the optimal combination of forecasts. *Working Paper, University of California at San Diego*.

Franta, M., J. Baruník, R. Horváth, and K. Šmídková (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.

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

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

Kapetanios, G., J. Mitchell, S. Price, and N. Fawcett (2015). Generalised density forecast combinations. *Journal of Econometrics 188*, 150–165.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies 65*, 361–393.

Koop, G. and D. Korobilis (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics 3*, 267–358.

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

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

Murphy, A. H. and R. L. Winkler (1987). A general framework for forecast verification. *Monthly Weather Review 115*, 1330–1338.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

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

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies 23*, 821–862.

Schmidt, V. (2013). Stochastics iii. Lecture notes, Ulm University, available at `https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/lehre/ws12/Stochastik_III/skript_02.pdf`.

Smith, J. and K. F. Wallis (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics 71*, 331–355.

Stock, J. H. and M. W. Watson (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. Volume 1 of *Handbook of Economic Forecasting*, pp. 135 – 196. Elsevier.

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

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*, 1455–1508.

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

Appendix

A   Derivations for baseline example

Here we prove that D and S are independent for $\omega = 0.5$. Consider the zero mean vector $W = \begin{bmatrix} X_1, & X_2, & U \end{bmatrix}'$ with variance-covariance-matrix

$$\Omega = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X^2 & 0 \\ \rho\sigma_X^2 & \sigma_X^2 & 0 \\ 0 & 0 & \sigma_X^2 \end{bmatrix}.$$

We can write W as $W = C\ Z$, where C is the lower-diagonal Cholesky matrix of Ω, and Z is a trivariate vector of independent standard normals. Simple algebra yields that

$$C = \begin{bmatrix} \sigma_X & 0 & 0 \\ \rho\sigma_X & \sqrt{1-\rho^2}\sigma_X & 0 \\ 0 & 0 & \sigma_U \end{bmatrix}.$$

D is proportional to $(X_1 - X_2)^2$. We can write

$$(X_1 - X_2)^2 = (\underbrace{\begin{bmatrix} 1 & -1 & 0 \end{bmatrix}}_{=a'} \times W)^2 = Z'\ C'aa'C\ Z;$$

furthermore,

$$S = (\underbrace{\begin{bmatrix} \omega\ (1+\rho) & (1-\omega)\ (1+\rho) & 1 \end{bmatrix}}_{=b'} W)^2 = Z'\ C'bb'C\ Z.$$

A standard result (Craig, 1943) states that D and S are independent if $C'\ aa'\ CC'\ bb'\ C = 0$. Simple but tedious algebra shows that this condition is satisfied for $\omega = 0.5$.

B   Basic formulas

The following formulas are used repeatedly in the proofs of Appendix C, and are thus listed for easier reference. Formulas involving third and higher moments of Gaussian quadratic forms involve cumbersome yet well-known calculations available in Theorems 1.6 and 1.7 of Schmidt (2013).

Object	Expression	Assumptions
$\mathbb{E}[D]$	$Tr\left(A\left(\Sigma_M + \mu_M\mu_M'\right)\right)$	
$\mathbb{E}[\omega'V]$	$\omega'\mu_V$	
$\mathbb{E}[\omega'V]$	$Tr\left(\left(\gamma\gamma' - 2\omega\gamma'\right)\Sigma_M + diag\left(\omega\right)\left(\Sigma_M + \mu_M\mu_M'\right)\right) + \sigma_U^2$	A1
$\mathbb{E}[S]$	$Tr\left(B\Sigma_M\right) + Tr\left(\omega\omega'\,\mu_M\mu_M'\right) + \sigma_U^2$	
$\mathbb{C}ov[D,S]$	$2\,\mathbb{E}\left[\eta^2\right]\,Tr\left(A\Sigma_M\left(B\Sigma_M - 2\left(\gamma - \omega\right)\omega'\,\mu_M\mu_M'\right)\right)$ $+\,\sigma_\eta^2\,\mathbb{E}[S]\,\mathbb{E}[D]$	A4
$\mathbb{C}ov[\omega'V,S]$	$\sigma_\eta^2\,\mathbb{E}[\omega'V]\,\mathbb{E}[S]$	
$\mathbb{C}ov[D,\omega'V]$	$\sigma_\eta^2\,\mathbb{E}[\omega'V]\,\mathbb{E}[D]$	
$\mathbb{V}[D]$	$2\,\mathbb{E}\left[\eta^2\right]\,Tr\left(A\Sigma_M A\left(\Sigma_M + 2\mu_M\mu_M'\right)\right) + \sigma_\eta^2\,\left(\mathbb{E}[D]\right)^2$	
$\mathbb{V}[\omega'V]$	$\mathbb{E}\left[\eta^2\right]\,\omega'\Sigma_V\omega + \sigma_\eta^2\,\left(\mathbb{E}[\omega'V]\right)^2$	

Table 4: Formulas and underlying assumptions. Matrices A and B are defined as follows: $A = diag(\omega) - \omega\omega'$, $B = (\gamma - \omega)(\gamma - \omega)'$.

C   Proofs

Proposition 4.2

Assumption A1 implies that $\mu_{V,}$, the ith element of vector μ_V, equals $\mathbb{E}[S_i]$, the MSFE of forecast i. An expression for the latter can be recovered from the expression for $\mathbb{E}[S]$ in Table 4, noting that the ith forecast corresponds to the special case $\omega = \iota_{[i]}$, an $n \times 1$ vector with all elements equal to zero except the ith element being equal to one. The result then follows immediately.

Proposition 4.3

Proof. A4 allows to derive explicit results for the relevant variance and covariance terms in a_1 and $b1$ defined in Section 5.3; see Table 4 for the corresponding formulas. The result of Proposition 4.3 then follows from Lemma C.1 below and A5.   □

Lemma C.1. *Under the prediction space framework together with A2 and A3, the following relations hold:*

- $Tr\left(A\Sigma_M B\Sigma_M\right) = 0$

- $\mu'_M A = 0 \; \iota'$, *where ι is an $n \times 1$ vector of ones.*

Proof. We first derive an expression for $B = (\gamma - \omega^*)(\gamma - \omega^*)'$, where ω^* denotes the MSFE-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. A2 implies that we can write $\mathbb{E}\left[MM'\right] = \kappa^2 \iota\iota' + \Sigma_M$ for some $\kappa \in \mathbb{R}$. From the inverse formula in Equation (160) of Petersen and Petersen (2012), we then find that

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

Hence,

$$
\begin{aligned}
B &\equiv (\gamma - \omega^*)(\gamma - \omega^*)' \\
&= \underbrace{\frac{(\iota'\gamma - 1)^2}{\left(\iota'\Sigma_M^{-1}\iota\right)^2}}_{\equiv \, \underline{c} \, \in \, \mathbb{R}_+} \left(\Sigma_M^{-1}\iota\iota'\Sigma_M^{-1}\right).
\end{aligned}
\tag{13}
$$

28

Equation (13) and the definition $A = [diag\,(\omega^*) - \omega^*\omega^{*\prime}]$ imply that

$$A\Sigma_M B\Sigma_M = \underline{c}\,[diag\,(\omega^*) - \omega^*\omega^{*\prime}]\,\iota\iota',$$

$$= \underline{c} \times \left(\underbrace{diag\,(\omega^*)\,\iota\iota'}_{=\omega^*\iota'} - \omega^* \underbrace{\omega^{*\prime}\iota\iota'}_{=1} \right),$$

$$= \underline{c} \times 0 \times I,$$

such that $Tr\,(A\Sigma_M B\Sigma_M) = 0$. We further have that $\mu'_M A = \kappa\ \iota' diag(\omega^*) - \kappa\ \iota'\omega^*\omega^{*\prime} = \kappa\ \omega^{*\prime} - \kappa\ \omega^{*\prime} = 0\ \iota'$. ∎

Proposition 5.3

Proof. We have that $\mathbb{E}[S] = \mathbb{E}[\omega'V] - \mathbb{E}[D] = c - \mathbb{E}[D]$, where the first equality follows from A1 and Proposition 5.1, and the second equality follows from A7. Note that $\mathbb{E}[D]$ depends on ω. Hence minimizing $\mathbb{E}[S]$ is equivalent to maximizing $\mathbb{E}[D]$, which in turn is equivalent to maximizing the upward bias of the LP, given by $2\ \mathbb{E}[D]$. □

Proposition 5.4

Proof: Under the stated assumptions, we can write

$$\begin{aligned} A &= n^{-1}I_n - n^{-2}\iota\iota' \\ B &= (n^{-1} - \tilde{\gamma})^2\iota\iota' \\ \Sigma_M &= (\sigma^2 - \theta)I_n + \theta\ \iota\iota' \end{aligned}$$

Using the fact that $\iota'\iota\iota' = n\ \iota'$, we find that

$$A\Sigma_M B\Sigma_M = 0\ \iota\iota' \tag{14}$$
$$Tr(A\Sigma_M A\Sigma_M) = (n-1)\ n^{-2}\ (\sigma^2 - \theta) \tag{15}$$

For the first statement, we consider the ratio

$$R_n = \frac{\mathbb{C}or(D, S)}{\mathbb{C}or(\omega'V, S)} = \frac{\mathbb{C}ov(D, S)}{\mathbb{C}ov(\omega'V, S)} \frac{\sqrt{V(\omega'V)}}{\sqrt{V(D)}}$$

Using A2', Equations (14) and (15) and the basic formulas, we get

$$R_n = \frac{\sigma_\eta\ E(D)}{\sqrt{2\ E(\eta^2)(n-1)n^{-2}(\sigma^2 - \theta) + \sigma_\eta^2\ (E(D))^2}}.$$

The result that $R_n \to 1$ as $n \to \infty$ then follows from noting that the first summand in the denominator vanishes as $n \to \infty$. The proof for the second statement is very similar and thus omitted.

29

D Details on the SV model

This appendix provides details on the model used in Section 7.2, which is of the form

$$Y_t = X'_{t-1}\beta + \varepsilon_t, \tag{16}$$
$$\varepsilon_t \sim \mathcal{N}(0,\ \exp(h_t)), \tag{17}$$
$$h_t = h_{t-1} + \nu_t, \tag{18}$$
$$\nu_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\nu^2), \tag{19}$$

where the vector $X_t = \begin{bmatrix} 1, & Z_t \end{bmatrix}'$ contains an intercept and a scalar regressor Z_t. We denote the sample period by $t = 1, \ldots, T$, and let $X = \begin{bmatrix} X_0 & X_1 & \ldots & X_{T-1} \end{bmatrix}'$ and $Y = \begin{bmatrix} Y_1, & \ldots, & Y_T \end{bmatrix}$.

Sketch of Gibbs sampler

We first initialize all parameters at arbitrary initial values (indicated by subindex $^{(0)}$). Then, in Gibbs sampler iteration $i = 1, 2, \ldots, n_G$,

1. Draw $\beta^{(i)}$ from a multivariate normal distribution with the following variance and mean:

$$\overline{V}_\beta = \left(\underline{V}_\beta^{-1} + X'^{(i-1)}X \right)^{-1}, \tag{20}$$
$$\overline{m}_\beta = \overline{V}_\beta \times \left(\underline{V}_\beta^{-1}\underline{m}_\beta + X'^{(i-1)}Y \right), \tag{21}$$

   where $W^{(i-1)} = \mathrm{diag}(1/\exp(h_1^{(i-1)}), \ldots, 1/\exp(h_T^{(i-1)}))$ and $\underline{V}_\beta$ and $\underline{m}_\beta$ denote the variance and mean of the Gaussian prior distribution for β.

2. Draw $h_1^{(i)}, \ldots, h_T^{(i)}$ using the procedure of Kim, Shephard, and Chib (1998); see Koop and Korobilis (2010, pp. 309-310) for a concise description.

3. Draw $\left(\sigma_\nu^{2(i)} \right)^{-1}$ from a Wishart distribution with $T + \underline{b}_\nu$ degrees of freedom and scale parameter $\left(\sum_{t=2}^{T}(h_t^{(i)} - h_{t-1}^{(i)})^2 + \underline{a}_\nu \right)^{-1}$, where $\underline{a}_\nu$ and $\underline{b}_\nu$ are prior parameters (see Table 5).

4. Simulate next period's log volatility as $h_{T+1}^{(i)} = h_T^{(i)} + \nu_{T+1}^{(i)}$, where $\nu_{T+1}^{(i)} \sim \mathcal{N}(0, \sigma_\nu^{2,(i)})$. The forecast mean and variance for iteration i are then given by $m_{T+1}^{(i)} = X_T'^{(i)}$ and $v_{T+1}^{(i)} = \exp(h_{T+1}^{(i)})$, respectively.

We use $n_G = 20,000$ Gibbs sampler draws, which are preceded by a burn-in period of $5,000$ draws. We compute the forecast mean and variance as

$$m_{T+1} = \frac{1}{n_G} \sum_{i=1}^{n_G} m_{T+1}^{(i)},$$

$$v_{T+1} = \frac{1}{n_G} \sum_{i=1}^{n_G} v_{T+1}^{(i)} + \frac{1}{n_G} (m_{T+1}^{(i)} - m_{T+1})^2;$$

these formulas follow from the usual view that the forecast distribution is an equally weighted mixture of the n_G forecast distributions obtained at the individual Gibbs iterations.

Priors

Table 5 summarizes the prior parameters. We use loose priors for β and a diffuse prior for (the initial log variance) h_0. Our choices of $\underline{a}_\nu$ and $\underline{b}_\nu$ allow for a considerable amount of stochastic volatility, as shown in the right panel of Figure 7.

Parameter	Value
$\underline{m}_\beta$	$\begin{bmatrix} 0, & 0 \end{bmatrix}'$
$\underline{V}_\beta$	$1000 \times I_2$
$\underline{m}_{h_0}$	0
$\underline{V}_{h_0}$	100
$\underline{a}_\nu$	1
$\underline{b}_\nu$	5

Table 5: Prior parameters for the two time series models. $\underline{m}_{h_0}$ and $\underline{V}_{h_0}$ denote the prior mean and variance of the initial log variance, h_0. All other parameters have been introduced in the text.

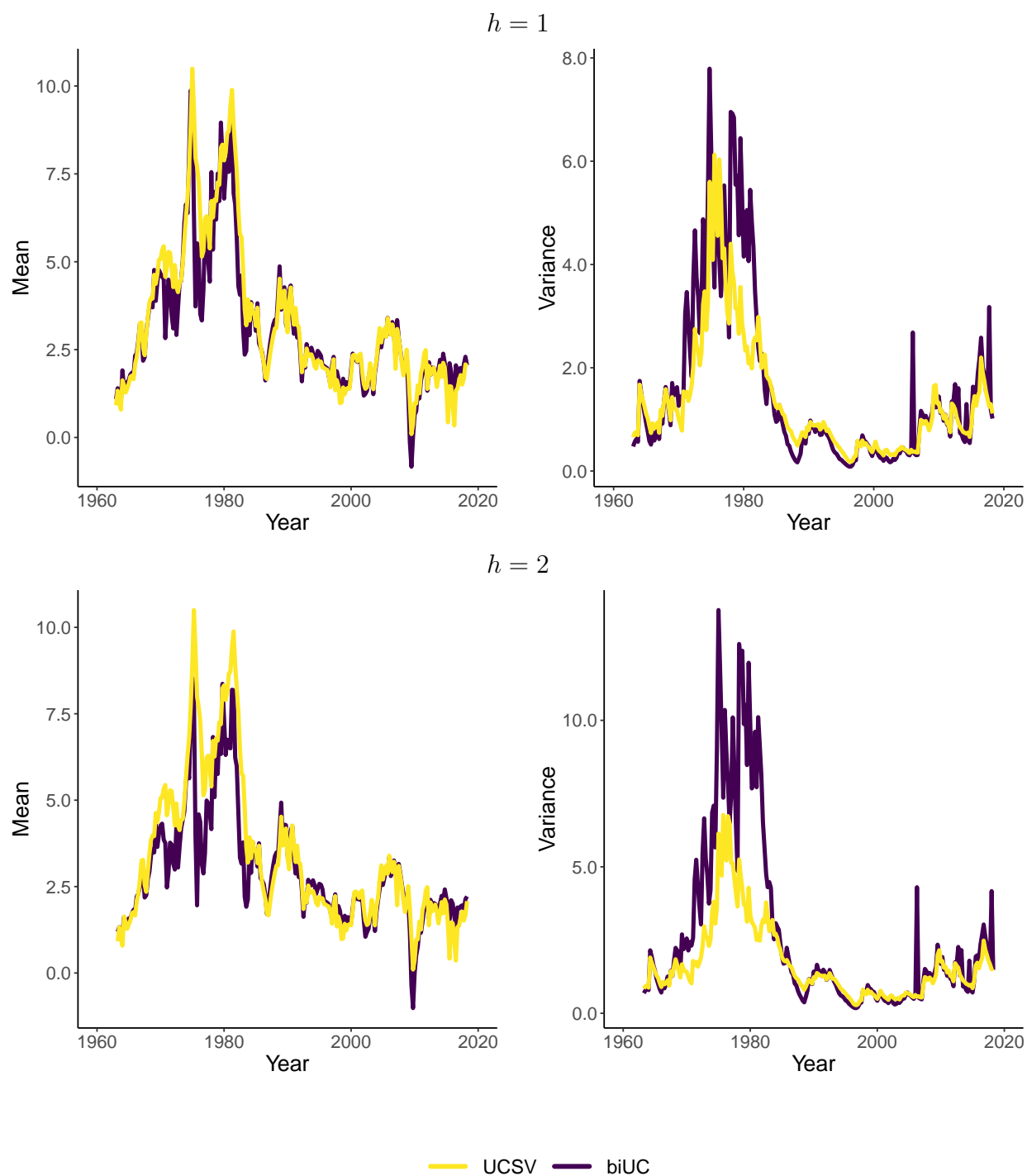# E   Additional Figures

$h = 1$

$h = 2$

Figure 5: *Inflation case study.* Predicted mean (left) and variance (right) of the two models, plotted over time.
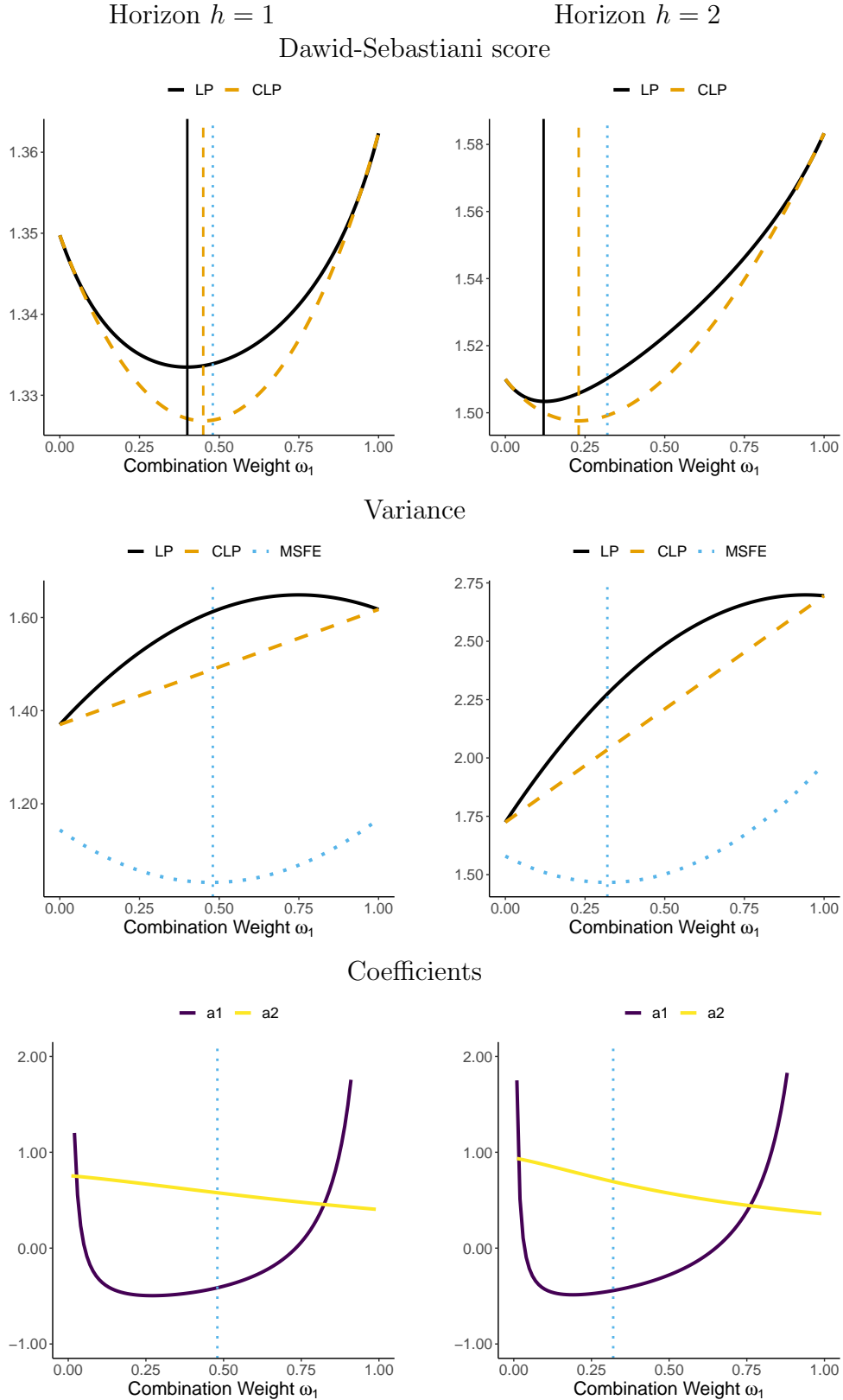
Figure 6: *Inflation case study.* Top row: Dawid-Sebastiani score plotted against the combination weight ω_1 placed on the UCSV model. A smaller score is better. The vertical lines mark the ex-post optimal weight for the pools (black and orange) and in terms of MSFE (blue). Middle row: Forecast variance plotted against combination weight; the blue curve indicates the MSFE. Bottom row: Regression coefficients in Equation (10), plotted against combination weight.
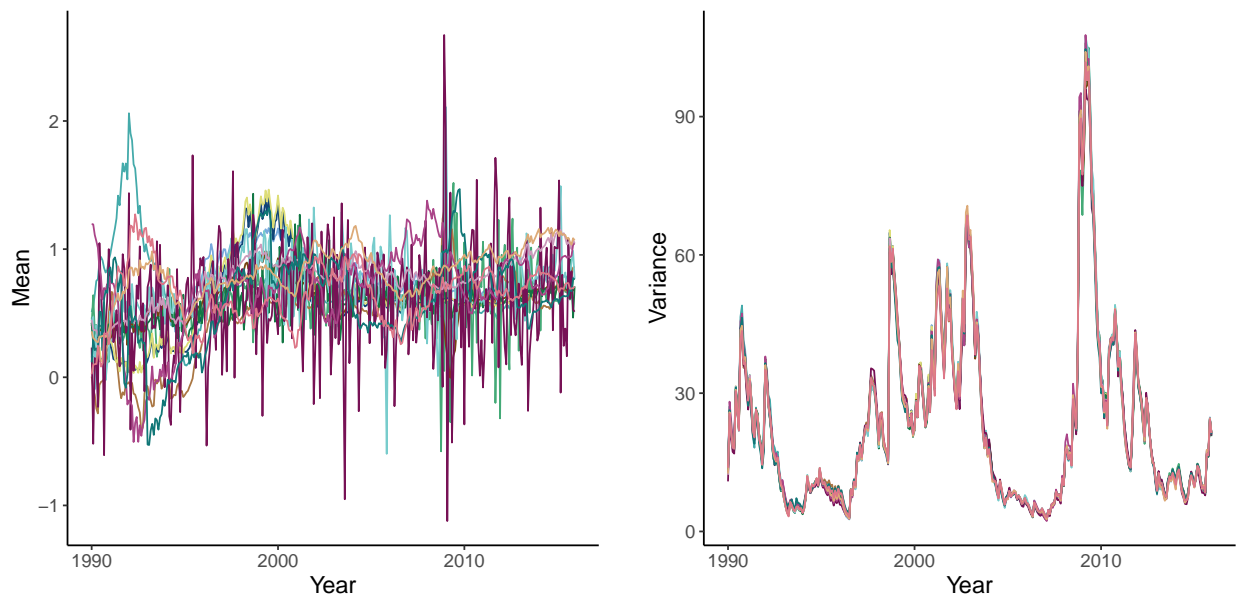
Figure 7: *Return case study.* Predicted mean (left) and variance (right) of the 15 models, plotted over time.