

Knüppel, Malte; Krüger, Fabian

**Conference Paper**

## Forecast Uncertainty, Disagreement, and Linear Pools of Density Forecasts

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Forecasting, No. E17-V2

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

Suggested Citation: Knüppel, Malte; Krüger, Fabian (2017) : Forecast Uncertainty, Disagreement, and Linear Pools of Density Forecasts, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Forecasting, No. E17-V2

This Version is available at:

<http://hdl.handle.net/10419/168294>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Forecast Uncertainty, Disagreement, and Linear Pools of Density Forecasts\*

Malte Knüppel  
Deutsche Bundesbank

Fabian Krüger  
Heidelberg University

March 1, 2017

*DRAFT VERSION (PRELIMINARY AND INCOMPLETE)*

## Abstract

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

## 1 Introduction

There is a growing recognition that measuring forecast uncertainty is of first-order importance to economic policy making. For example, many central banks have followed

---

\*This paper reflects the authors' personal opinions, but not necessarily the views of the Deutsche Bundesbank or its staff. Authors' email addresses: malte.knueppel@bundesbank.de and fabian.krueger@awi.uni-heidelberg.de. We thank seminar and workshop participants at HITS, Heidelberg University and the National Bank of Poland for helpful comments. Of course, any errors are our own responsibility. The work of Fabian Krüger (FK) has been funded by the European Union Seventh Framework Programme under grant agreement 290976. FK also thanks thank the Klaus Tschira Foundation for infrastructural support at the Heidelberg Institute for Theoretical Studies (HITS), where he was employed during the initial stage of this project.

the Bank of England’s lead in publishing probabilistic forecasts of inflation and related variables.<sup>1</sup> 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 a large number of available forecasting methods and data sources, it is now common to consider a combination of many methods, rather than a single method, for that purpose.

There are many possible ways to pool the individual densities  $f_1, \dots, f_n$  in order to construct a combined density  $f_c$ . In practice, though, much of the literature focuses on the linear pool (LP) given by  $f_c = \sum_{i=1}^n \omega_i f_i$ , where  $\{\omega_i\}_{i=1}^n$  is a set of combination weights. See, for instance, Hall and Mitchell (2007), Geweke and Amisano (2011), Waggoner and Zha (2012), or Del Negro, Hasegawa, and Schorfheide (2016). In this work, we analyze the linear pool’s implications concerning forecast uncertainty in a mean-variance prediction space framework, i.e. in a framework where we consider the joint distribution of mean forecasts, variance forecasts, and the target variable in terms of their first two moments. This general setup allows us to derive several new results.

If the individual density forecasts are dispersed correctly. i.e. if their variance forecasts are unbiased, the dispersion of the linear pool is too large with respect to the mean-squared error (MSE) of its mean forecast, as shown by Gneiting and Ranjan (2013). Using the fact that the variance of the linear pool can be decomposed into a weighted-individual-variance component and a disagreement component, as stated, *inter alia*, in Wallis (2005), we show that eliminating the disagreement component unambiguously reduces the variance bias of the linear pool, if the individual forecasts are correctly dispersed and the weights are positive. The disagreement component measures the average squared distance of the individual mean forecasts from the combined mean forecast. Its elimination is achieved simply by recentering all individual forecasts at the combined mean forecast, resulting in a centered linear pool of density forecasts. Under the conditions mentioned, the dispersion of the centered linear pool is smaller than the dispersion of the linear pool, but still larger than the MSE of the combined mean forecast.

In addition, we show that the weighted-individual-variance component encompasses the disagreement component concerning the prediction of the MSE of the combined mean forecast, if the combination weights are chosen optimally according to the MSE criterion formulated in Bates and Granger (1969), if the forecasts and the target variable are normally distributed, and if the individual variance forecasts are perfectly correlated with the MSE of the combined mean forecast. Put differently, under these conditions, the disagreement component is a noise term in the variance forecast of the linear pool.

These two results suggest that the disagreement component can be detrimental to

---

<sup>1</sup>A list of these central banks can be found in Table 1 of Franta, Baruník, Horváth, and Šmídová (2014).

the performance of the linear pool. Indeed, using the scoring rule proposed by Dawid and Sebastiani (1999), which is designed for the joint evaluation of mean and variance forecasts and which is equal to the logarithmic score if the forecast density is normal, we find that the centered linear pool outperforms the linear pool if all conditions mentioned above hold, i.e. if the individual density forecasts are dispersed correctly, the Bates-Granger-weights are used, normality is given, and the individual variance forecasts are perfectly correlated with the MSE of the combined mean forecast. Moreover, it turns out that the optimum of the Dawid-Sebastiani score can be obtained employing the Bates-Granger-weights and reducing the variance forecast of the centered linear pool by subtracting the expected disagreement.

In applications of the centered linear pool to density forecasts of macroeconomic variables and stock returns, we find that it often outperforms the linear pool, and that, in several cases, the weighted-individual-variance component encompasses the disagreement component concerning the prediction of the MSE of the combined mean forecast.

The remainder of this paper is structured as follows: Section 2 presents a baseline example which motivates our analysis. Section 3 introduces the formal setup, and Section 4 presents our main theoretical results on the linear pool. Sections 5 and 6 contains simulation and data examples, and Section 7 concludes.

## 2 Baseline example

Consider the target variable  $Y$  determined as

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

where  $X_1, X_2$  and  $U$  are distributed as

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

Forecaster 1 only observes  $X_1$ , forecaster 2 only observes  $X_2$ ; Each forecaster aims at predicting the distribution of  $Y$ , and each forecaster  $i$  has an infinite number of paired observations of  $(Y, X_i)$ . Accordingly, both forecasters state the correct forecast distribution given their information sets.

Both forecasters thus issue a Gaussian distribution, with mean  $M_i$  and variance  $V_i$  shown in Table 1. A linear pool of the two forecasts is given by  $f_c = \omega f_1 + (1 - \omega)f_2$ , where  $f_c$  is the density of the combined forecast,  $f_1$  and  $f_2$  are the individual densities, and  $0 < \omega < 1$ . Table 1 lists the formulas for the linear pool's mean  $M_c$  and variance  $V_c$  in the present example. As we argue below, the combined variance forecast  $V_c$  should have two properties:

Object	Formula
Mean of forecaster $i$	$M_i = (1 + \rho)X_i$
Variance of forecaster $i$	$V_i = V = (1 - \rho^2)\sigma_X^2 + \sigma_U^2$
Combined mean	$M_c = \omega M_1 + (1 - \omega)M_2$
Disagreement	$D = \omega (M_c - M_1)^2 + (1 - \omega) (M_c - M_2)^2$
Combined variance	$V_c = \omega V_1 + (1 - \omega)V_2 + D$
Squared error (SE) of $M_c$	$S_c = ((1 - \omega) M_1 + \omega M_2 + U)^2$
<i>Expected</i> combined variance	$\mathbb{E}(V_c) = (1 - \rho^2)\sigma_X^2 + \sigma_U^2 + 2 \omega (1 - \omega) (1 - \rho^2) (1 + \rho) \sigma_X^2$
<i>Expected</i> disagreement	$\mathbb{E}(D) = 2 \omega (1 - \omega) (1 - \rho^2) (1 + \rho) \sigma_X^2$
<i>Expected</i> SE of $M_c$	$\mathbb{E}(S_c) = (1 - \rho^2)\sigma_X^2 + \sigma_U^2 - 2 \omega (1 - \omega) (1 - \rho^2) (1 + \rho) \sigma_X^2$

Table 1: Formulas for the mean, variance, and squared forecast error of the linear pool in the motivating example.

- It should have predictive power for  $S_c$ , the squared forecast error associated with  $M_c$ . In practice,  $V_c$  should thus be correlated with  $S_c$ .
- It should have the same unconditional mean as  $S_c$ . Intuitively, the variance forecast should correctly assess the magnitude of forecast errors.

As shown in Table 1, the linear pool's variance is of the form  $V_c = a + D$ , where  $D = \omega (M_1 - M_c)^2 + (1 - \omega) (M_2 - M_c)^2$  reflects disagreement between the two point forecasts, and  $a$  is a constant. Strikingly, the linear pool's variance  $V_c$  fails both of the requirements mentioned above:

- For the leading case  $\omega = 0.5$ , it can be shown that  $D$  is *independent* of  $S_c$ .<sup>2</sup> For other choices of  $\omega$ , the relation between  $D$  and  $S_c$  depends on  $\rho, \sigma_x^2$  and  $\sigma_U^2$ , but often implies weak correlation between  $D$  and  $S_c$ .
- The linear pool's expected variance,  $\mathbb{E}(V_c)$  exceeds the expected squared error  $\mathbb{E}(S_c)$  for all admissible values of  $\omega$ . The linear pool can therefore be labeled underconfident or overdispersed. Further inspection shows that the upward bias of  $\mathbb{E}(V_c)$  is equal to  $2 \times \mathbb{E}(D)$ . Hence, the linear pool would overpredict the true forecast uncertainty even if the disagreement component was eliminated from the pool's variance forecast.

In the above example, disagreement can thus be viewed as a positive-valued noise term which harms the linear pool's forecast variance.

<sup>2</sup> $\omega = 0.5$  is a popular default choice in practice, and is the MSE-optimal choice in the present example.

### 3 Formal setup

Here we describe methods for evaluating and constructing forecasts, as well as a probability model for the joint distribution of forecasts and realizations.

#### 3.1 Dawid-Sebastiani score for forecast evaluation

We consider combining  $n$  alternative mean and variance forecasts that could stem from different data sources, different statistical methods, or could represent the judgement of different forecasters. We view the forecasts as random variables, collected in the random vector  $Z \equiv (M_1, \dots, M_n, V_1, \dots, V_n)'$  with support  $\mathcal{Z}$ . After observing  $Z = z$ , a forecast combination method returns a combined mean forecast  $m = f_m(z)$  and a combined variance forecast  $v = f_v(z)$ . The combination method is characterized by the functions  $f_m$  and  $f_v$ . While we will study specific combination methods below,  $f_m$  and  $f_v$  can initially be thought of as general functions. We evaluate the combined forecast with the Dawid and Sebastiani (1999) score (DSS):

$$DSS(m, v, y) = \log v + \frac{(y - m)^2}{v}; \quad (1)$$

here  $y$  denotes a realization of the target variable  $Y$ , and a smaller score corresponds to a better forecast.<sup>3</sup> Our goal is to choose functions  $f_m$  and  $f_v$  which minimize the expected DSS:

$$\min_{f_m, f_v} \mathbb{E} [DSS(f_m(Z), f_v(Z), Y)],$$

where the expectation  $\mathbb{E}$  is over both  $Z$  and  $Y$ . Using the law of iterated expectations, the above problem is solved by setting

$$\begin{aligned} m &= f_m(z) = \mathbb{E}_z [Y], \\ v &= f_v(z) = \mathbb{E}_z [S], \end{aligned} \quad (2)$$

where  $\mathbb{E}_z$  denotes expectation conditional on  $Z = z$  and  $S$  is defined as

$$S = (Y - f_m(z))^2.$$

In words, optimality requires to set  $m$  equal to the conditional expectation of  $Y$ , and  $v$  equal to the conditional expectation of  $(Y - f_m(z))^2$ , the squared forecast error ( $S$ ).

Equation (2) yields another implication of forecast optimality:

$$\mathbb{E}[(Y - f_m(Z))^2] = \mathbb{E}[f_v(Z)], \quad (3)$$

i.e., the expected squared error of the mean forecast should be equal to the expected forecast variance.

---

<sup>3</sup>The DSS is a legitimate loss function for density forecasts, in that it is a proper scoring rule as studied by Gneiting and Raftery (2007). It is equivalent to the logarithmic score of a Gaussian distribution.

### 3.2 Linear pools

The linear prediction pool, which is typically attributed to Stone (1961), is a specific, very popular combination method which sets

$$f_m(z) = \bar{m} = \sum_{i=1}^n \omega_i m_i \quad (4)$$

$$f_v^{LP}(z) = \sum_{i=1}^n \omega_i v_i + \sum_{i=1}^n \omega_i (m_i - \bar{m})^2 \quad (5)$$

where  $\sum_{i=1}^n \omega_i v_i$  is the weighted-individual-variance component and  $\sum_{i=1}^n \omega_i (m_i - \bar{m})^2$  is the disagreement component of the linear pool's variance forecast. In what follows, we will 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 the combination

$$f_v^{CLP}(z) = \sum_{i=1}^n \omega_i v_i \quad (6)$$

for the variance forecast, which we refer to as the *centered linear pool* (CLP). In terms of density forecasts, this combination method relocates each individual density such that its resulting mean  $m_i^{CLP}$  is equal to  $\bar{m}$ .

The performance of the linear pools depends on the joint properties of the forecasts (i.e., the individual forecast means and variances) and the predictand. As a next step, we therefore propose a model for this object.

### 3.3 Model for forecasts and realizations

We assume that the joint distribution of mean and variance forecasts is given by

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

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

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

where  $Y$  is the demeaned predictand. The mean forecasts are collected in  $M \in \mathbb{R}^n$ , which is an  $n$ -variate continuous random vector. Since  $\mathbb{E}[Y] = 0$ ,  $\mu_M$  denotes the biases of the mean forecasts.  $V \in \mathbb{R}_+^n$  is an  $n$ -variate random vector with positive continuous entries denoting the variance forecasts.

The independent scalar random variable  $\eta \in \mathbb{R}_+$  has  $\mathbb{E}[\eta] = 1$  and scales the variances of  $M$ ,  $U$ , and  $Y$ , implying that these variances are perfectly correlated. If we consider the case where the mean forecasts are based on different information sets, this assumption implies that the variances of the variables in these information sets are perfectly correlated with the variance of the predictand  $Y$ . While this assumption might appear restrictive, the following results hold conditional on a specific value of  $\eta$ . This means that, for example, in a time series context where  $\eta$  varies over time, these results actually hold within each period. Moreover, in macroeconomics, the variances of many variables actually appear to be strongly correlated, as documented in Carriero, Clark, and Marcellino (2016).

Note that the example described in 2 is a special case of the general setup presented here with  $\eta$  being constant and, hence, equal to 1, with

$$\begin{bmatrix} M_1 \\ M_2 \\ V_1 \\ V_2 \\ U \end{bmatrix} \sim \left( \begin{bmatrix} 0 \\ 0 \\ (1 - \rho^2) \sigma_X^2 + \sigma_U^2 \\ (1 - \rho^2) \sigma_X^2 + \sigma_U^2 \\ 0 \end{bmatrix}, \begin{bmatrix} (1 + \rho)^2 \sigma_X^2 & \rho(1 + \rho)^2 \sigma_X^2 & 0 & 0 & 0 \\ \rho(1 + \rho)^2 \sigma_X^2 & (1 + \rho)^2 \sigma_X^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_U^2 \end{bmatrix} \right)$$

and with

$$Y = [M_1 \quad M_2] \begin{bmatrix} (1 + \rho)^{-1} \\ (1 + \rho)^{-1} \end{bmatrix} + U.$$

## 4 Theoretical results

In this section, we derive our main theoretical results.

### 4.1 List of assumptions considered

Unless explicitly noted otherwise, we assume that the prediction space framework of Section 3.3 applies. We also consider the following restrictions on the latter:

- A1 The individual variance forecasts  $V$  are unconditionally unbiased, i.e.  $\mathbb{E}[V_i] = \mathbb{E}[Y - M_i]^2, i = 1, \dots, n$ .
- A2 The forecast means are all identical, i.e.  $\mu_M = [\kappa, \dots, \kappa]'$  for some  $\kappa \in \mathbb{R}$ .
- A2+ The forecast means are all equal to zero, i.e.  $\mu_M = [0, \dots, 0]'$ .
- A3 The combination weights  $\omega$  are deterministic and sum to one.



A3+ The combination weights  $\omega$  are deterministic, *positive* and sum to one.

A4 The combination weights have been chosen such that  $\mathbb{E}[S]$  is minimized, subject to the constraint of adding to one (Bates and Granger (1969)).

A5 The joint distribution of  $M$  and  $U$  conditional on  $\eta$  is normal.

A6 It holds that  $\Sigma_V = 0$ .

**Remark 4.1.** *The optimal weights  $\omega^*$  fulfilling A4 do not depend on  $\eta$ , and, thus, they also minimize  $\mathbb{E}[S|\eta]$ .*

## 4.2 Biases in the variance forecasts of linear pools

The following result uses the prediction space framework of Section 3.3 to derive a specific formula for  $\mathbb{E}[V_{lp}]$ , the variance of the LP.

**Theorem 4.1.** *Under A1 and A3, it holds that  $\mathbb{E}[S] = \mathbb{E}[\omega'V] - \mathbb{E}[D]$ . Hence the LP and CLP systematically deviate from  $S$  according to*

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

*Proof.* See Appendix. □

**Corollary 4.1.** *The DSS is minimized by using the weights from A4 for the mean forecast and setting  $v = \mathbb{E}[\omega'V|\eta] - \mathbb{E}[D|\eta]$ . This follows from the facts that, given any variance forecast  $v \in \mathbb{R}_+$ , the DSS-optimal mean forecast minimizes  $\mathbb{E}[S|\eta]$ , and that the optimal variance forecast  $v$  equals  $\mathbb{E}[S|\eta]$ .*

Interestingly, a qualitative statement similar to that of Theorem 4.1 can be derived without the prediction space framework of Section 3.3.

**Theorem 4.2.** *Consider an arbitrary joint distribution of  $M$ ,  $V$  and  $Y$ . Under A1 and A3+, both the linear pool and the centered linear pool are overdispersed, and the centered linear pool is less overdispersed than the linear pool.*

*Proof.* See Appendix. □

Note that Theorems 4.1 and 4.2 have the same qualitative interpretation: If the individual variances are unbiased, then both the LP and the CLP are underconfident. However, Theorem 4.1 exploits the prediction space framework to make a precise quantitative statement on the pools' underconfidence.

### 4.3 Disagreement encompassed by weighted individual variance forecasts

Consider the following linear regressions:

$$\begin{aligned} S &= a_0 + a_1 D + a_2 \omega'V + \text{error}; \\ S &= b_1 D + b_2 \omega'V + \text{error}; \end{aligned} \tag{7}$$

We have that

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

where  $\tilde{\alpha}$  and  $\tilde{\beta}$  are the coefficients from a linear regression of  $D$  on a constant and  $\omega'V$ . A very similar expression holds for  $b_1$  (same formula, except that  $\text{Cov}[A, B]$  is replaced by  $\mathbb{E}[AB]$  and  $\mathbb{V}[A]$  is replaced by  $\mathbb{E}[A^2]$ ).

**Theorem 4.3.** *Assume that A2, A4, A5 and A6 hold. Then,  $a_1 = b_1 = 0$ , i.e.  $\omega'V$  encompasses disagreement in the prediction of  $S$ .*

*Proof.* A5 allows to derive expressions for the relevant variance and covariance terms (see Appendix B). Under A2 and A4, some of the terms simplify considerably (see Lemma B.1 in the appendix). The result then follows from using A6.  $\square$

**Remark 4.2.** *Assumption 4 can imply negative weights.*

### 4.4 Relative precision of $\omega'V$ and $D$ in the prediction of $S$

Theorem 4.3 relies on the assumption  $\Sigma_V = 0$ . If this assumption does not hold, the relative importance of  $D$  and  $\omega'V$  for the prediction of  $S$  turns out to depend on their coefficients of variation.

**Theorem 4.4.** *Under assumptions A2+, A4, and A5, predictions for  $S$  based on  $\omega'V$  do not produce a higher mean-squared error than predictions based on  $D$  if the condition*

$$\frac{\sqrt{\mathbb{V}[\omega'V]}}{\mathbb{E}[\omega'V]} \leq \frac{\sqrt{\mathbb{V}[D]}}{\mathbb{E}[D]}. \quad (8)$$

*holds. If the inequality is strict, using  $\omega'V$  yields more precise predictions than using  $D$ .*

*Proof.* See Appendix. □

It might be interesting to assess the probability that (8) holds in practice. Since  $\mathbb{V}[\omega'V] = 0$  if  $\Sigma_V = 0$ , the question is whether  $\Sigma_V$  could easily become large enough for the condition (8) to be violated. While it is difficult to investigate this issue in a general manner, the case with  $n = 2$  forecasts and  $\eta$  being constant yields valuable insights.

**Theorem 4.5.** *Assume that each variance forecast  $V_i$  equals  $V_i = (y - m_i)^2$ , where  $m_i$  is a single mean forecast and  $y$  is the corresponding realization of  $Y$ . Under assumptions A2+ and A5, and with  $n = 2$  and  $\eta = 1$ , the coefficients of variation of  $\omega'V$  and  $D$  are related by*

$$\frac{\sqrt{\mathbb{V}[\omega'V]}}{\mathbb{E}[\omega'V]} \leq \frac{\sqrt{\mathbb{V}[D]}}{\mathbb{E}[D]} = \sqrt{2}.$$

*If the forecasts are not perfectly correlated, and if no forecast receives a weight of one, the inequality is strict.*

*Proof.* See Appendix. □

If the assumptions of Theorem 4.5 hold, but  $V_i$  is determined using  $R$  forecast errors, i.e. by  $V_i = V_i^{(R)} = \frac{1}{R} \sum_{r=1}^R (y_r - m_{i,r})^2$ , and if the forecast errors  $y_{r_1} - m_{i,r_1}$  and  $y_{r_2} - m_{i,r_2}$  are uncorrelated for  $r_1 \neq r_2$ , we have that  $\sqrt{\mathbb{V}[\omega'V^{(R)}]} / \mathbb{E}[\omega'V^{(R)}] = R^{-1/2} \sqrt{\mathbb{V}[\omega'V^{(1)}]} / \mathbb{E}[\omega'V^{(1)}]$ , where  $V^{(1)}$  corresponds to the variance forecasts described in Theorem 4.5. The fact that the coefficient of variation of  $\omega'V$  will be considerably smaller than its counterpart for  $D$ , even if the former is estimated from a very small number of previous forecast errors only, makes it seem unlikely that in empirical applications,  $\Sigma_V$  will be large enough for the condition (8) to be false.

For the baseline example in Section 2 with  $w = 1/2$ , under the assumption of Theorem 4.5 one obtains

$$\frac{\sqrt{\mathbb{V}[\omega'V]}}{\mathbb{E}[\omega'V]} = \sqrt{\frac{2\sigma_U^4 + c_{UX} 2\sigma_U^2\sigma_X^2 + c_X \sigma_X^4}{\sigma_U^4 + c'_{UX} 2\sigma_U^2\sigma_X^2 + c'_X \sigma_X^4}}$$

with  $c_{UX} = (1 + \rho)(1 - \rho)^2$ ,  $c_X = (1 + \rho^2)(1 - \rho)^2(1 + \rho)^2$ ,  $c'_{UX} = (1 - \rho^2)$ , and  $c'_X = (1 - \rho^2)^2$ . Since  $c_{UX} / c'_{UX} = 1 - \rho$  and  $c_X / c'_X = \rho^2 + 1$ , the inequalities  $0 \leq c_{UX} / c'_{UX} \leq 2$  and  $0 \leq c_X / c'_X \leq 2$  hold, implying that  $\sqrt{\mathbb{V}[\omega'V]} / \mathbb{E}[\omega'V] \leq 2$ .

## 5 Monte Carlo results

[TO BE COMPLETED]

## 6 Empirical case studies

We next investigate the properties of the LP and CLP in two case studies from macroeconomics and finance.

### 6.1 Nowcasting quarterly macro data

- Forecast 1: Model based nowcasts, obtained as in Krüger, Clark, and Ravazzolo (forthcoming) Forecast 2: SPF nowcast, plus variance from stochastic volatility specification. The latter is obtained by fitting a Bayesian linear model with SV on the SPF nowcast, and shrinking the coefficient on the SPF to be essentially equal to one. Appendix D provides a formal description of the estimation methodology and prior choices.
- Data and variable definitions closely follow Krüger, Clark, and Ravazzolo (forthcoming)
- Results:
  - Disagreement is quite large (accounting for 19% of LP’s variance for GDP, 25% for INF, 38% for TBI, and 64% for UNE), see Table 2.
  - Nowcast models are underconfident, in the sense that average variance exceeds  $SE$ . Disagreement leads to even more underconfidence.
  - Regressions of  $SE$  on  $D$  and  $V$  (Table 3):  $D$  is significant in case of GDP and TBI.  $V$  is significant in case of INF and TBI.
  - CLP has lower (i.e., better) scores than LP for all variables. DM tests are significant for TBI and UNE.
  - Results suggest that, on the whole,  $D$  harms LP. Hence underconfidence effect (Table 2) seems more important than correlation with  $SE$  (Table 3).

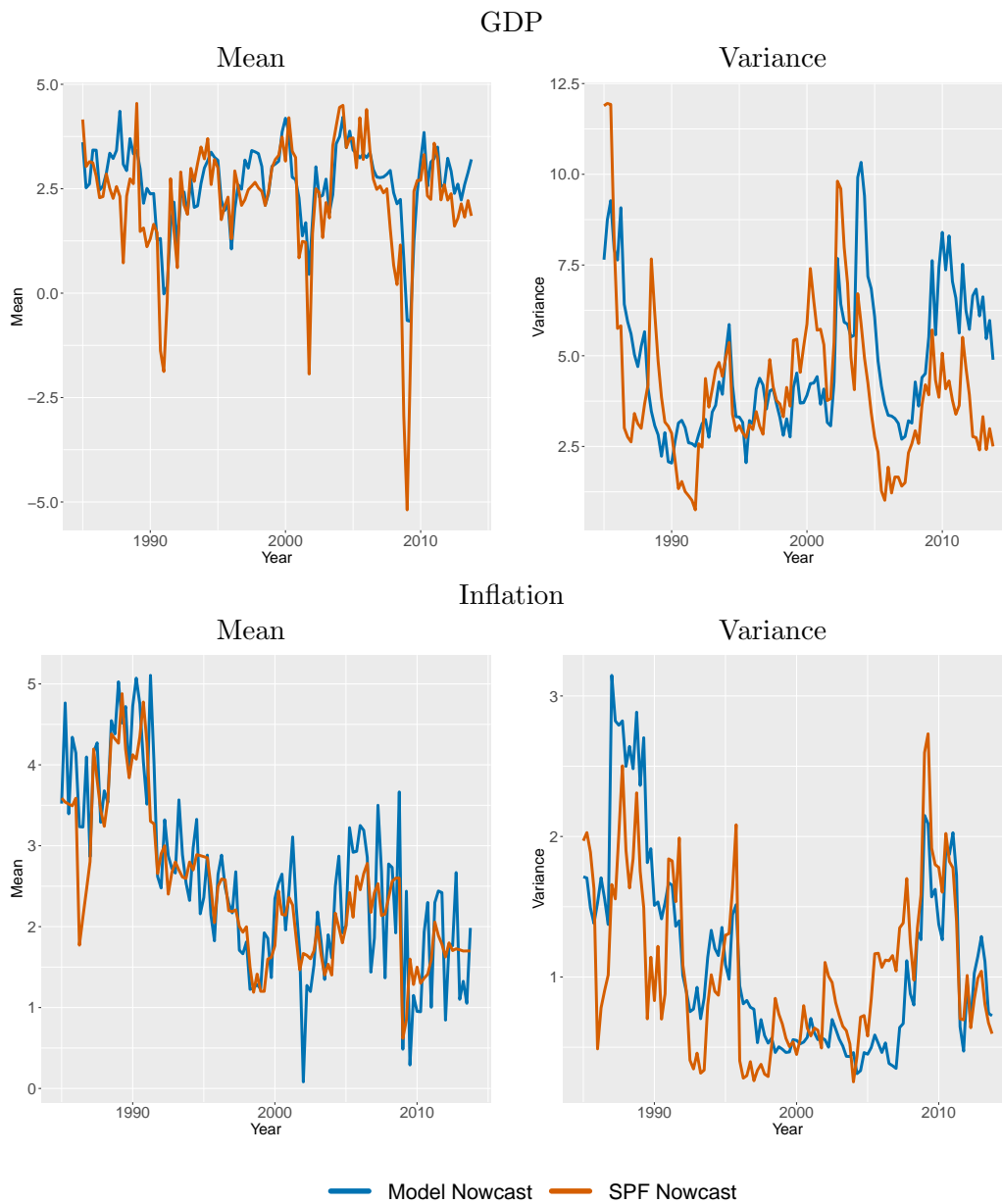


Figure 1: Mean and variance predictions of the two nowcasting methods, plotted over time.

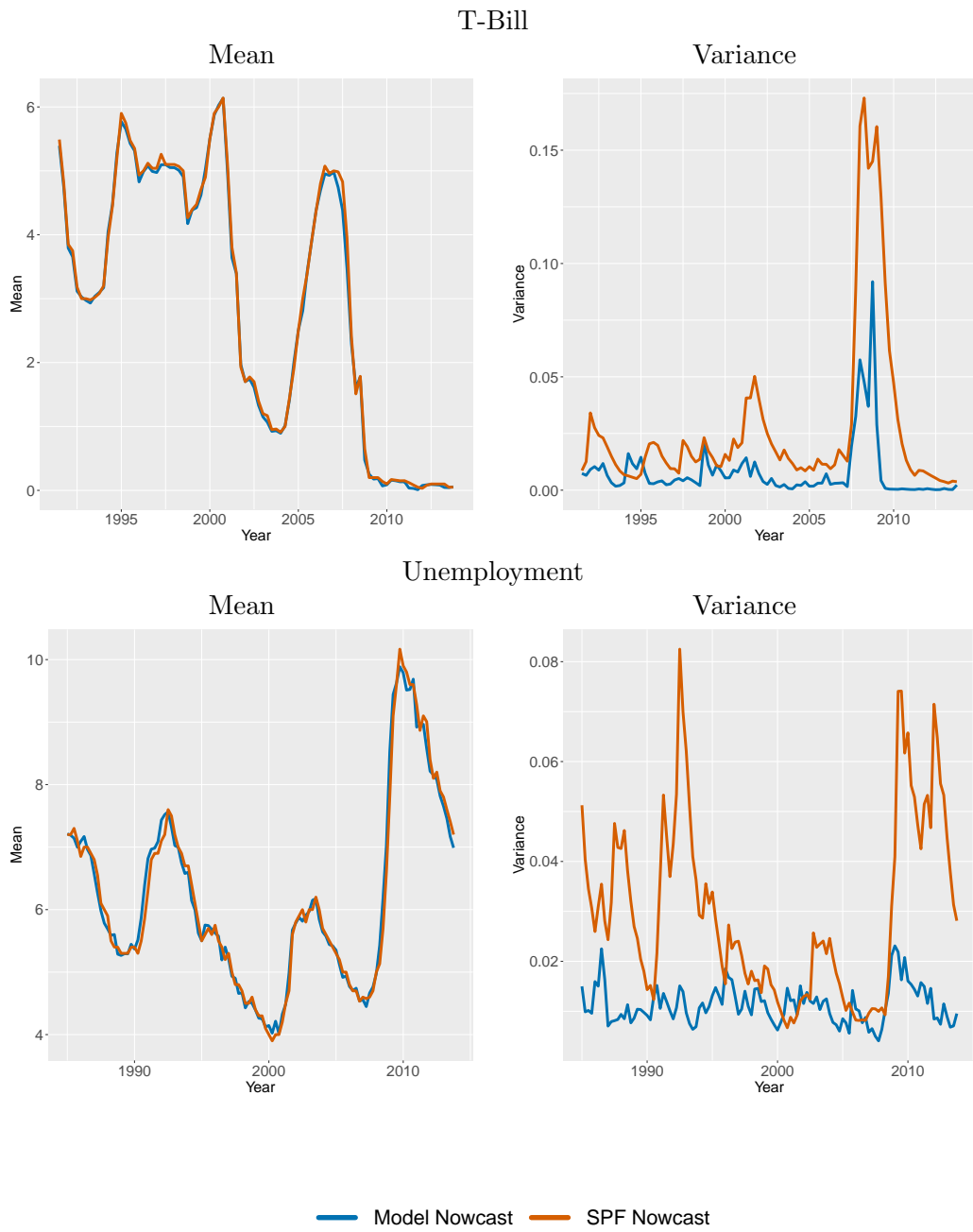


Figure 1: Mean and variance predictions of the two nowcasting methods, plotted over time.

	$V_{LP}$	$\omega'V$	D	SE	$DSS_{LP}$	$DSS_{CLP}$
GDP	5.431	4.441	0.990	2.745	2.129	2.098
INF	1.461	1.095	0.366	0.857	0.863	0.825
TBI	0.027	0.017	0.010	0.009	-3.878	-4.017
UNE	0.056	0.020	0.035	0.005	-3.147	-3.691

Table 2: Nowcast case study. Sample averages for several variables. The sample period is 1985:Q1 to 2013:Q4, except for TBI (1991:Q3 to 2013:Q4).

	(Intercept)	$D$	$\omega'V$
GDP	1.211 (0.655)	0.827 (0.234)	0.161 (0.111)
INF	0.298 (0.146)	0.522 (0.374)	0.336 (0.148)
TBI	-0.001 (0.001)	0.378 (0.025)	0.399 (0.095)
UNE	0.004 (0.001)	0.013 (0.013)	0.042 (0.071)

Table 3: Nowcast case study. Regression of squared errors on disagreement and average variance. Heteroscedasticity robust standard errors in parentheses (Newey-West with truncation at lag zero). The sample period is as defined below Table 2.

## 6.2 Forecasting monthly excess returns

- Models of the form

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

where  $R_t$  is the monthly excess return,  $X_t$  is one out of 15 predictors considered by Welch and Goyal (2008), and  $\varepsilon_t$  is an error term with stochastic volatility. Note: Forecast combination for excess returns has been considered by Rapach, Strauss, and Zhou (2010) and others.

- The model is estimated via Bayesian methods, see Appendix D. Based on expanding window, with observations ranging back until January 1970. Evaluation period is from January 1990 to December 2015.
- Results:
  - The individual models are underconfident, with average variance clearly exceeding the LP's squared error (Table 4).
  - On average, disagreement is much smaller than variance. Accordingly, the LP and the CLP perform very similarly (Table 4).
  - Disagreement does not predict squared forecast errors, but average variance does (Table 5)

$V_{LP}$	$\omega'V$	D	SE	$DSS_{LP}$	$DSS_{CLP}$
23.255	23.192	0.063	18.101	3.776	3.776

Table 4: Excess return case study: Sample averages for several variables. The sample period is January 1990 to December 2015.

(Intercept)	$D$	$\omega'V$
6.478	14.364	0.462
(3.308)	(39.067)	(0.106)

Table 5: Excess return case study: Regression of squared errors on disagreement and average variance. Heteroscedasticity robust standard errors in parentheses (Newey-West with truncation at lag zero). The sample period is as defined below Table 4.



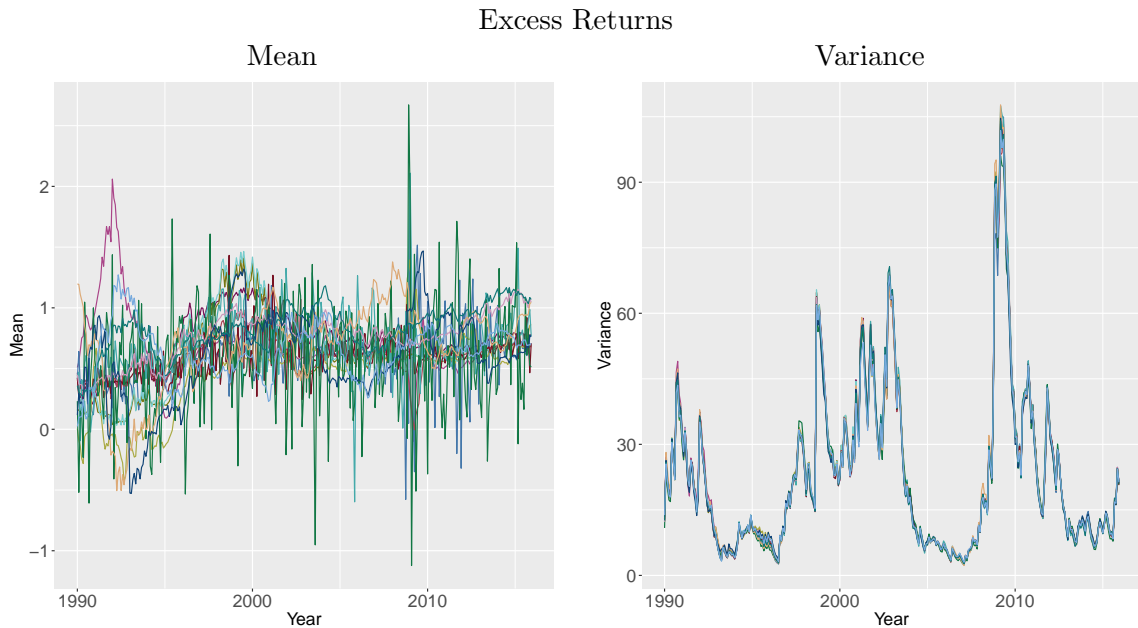


Figure 2: Mean and variance predictions of models for excess returns, plotted over time. Each of the 15 lines represents a model based on a specific predictor variable.

## 7 Conclusion

In this work, we have established a general setup for modeling the joint distribution of mean forecasts, variance forecasts, and the target variable. This setup has allowed us to derive several new results concerning uncertainty forecasts based on linear pools.

Under the assumption of correctly dispersed individual density forecasts, we can quantify the variance bias of the linear pool which simply equals twice the expected disagreement. Moreover, we find that under certain conditions, disagreement is encompassed by the weighted individual variance forecasts concerning the prediction of the MSE of the combined mean forecast. These conditions are the normality of forecasts and the target variable, the use of Bates-Granger weights, the absence of noise in the variance forecasts, and the requirement that the mean forecasts be unbiased or identically biased. If all conditions mentioned are fulfilled, the disagreement component unambiguously deteriorates the performance of the linear pool according to the scoring rule proposed by Dawid and Sebastiani (1999), which is designed for the joint evaluation of mean and variance forecasts. Therefore, a simple modification of the linear pool, the elimination of the disagreement component which produces a centered linear pool, leads to an improved forecast performance.

While it is unlikely that all conditions mentioned are met in practice, it could be that the deviations are small enough for the centered linear pool to yield better results than the linear pool. For example, concerning the assumption of absence of noise in the variance forecasts, we find that this noise is likely to be very small compared to the implicit noise in the disagreement component.

In the empirical applications considered, it turns out that the centered linear pool tends to outperform the linear pool if the disagreement component is not too small. This is the case for the nowcasts of quarterly macroeconomic variables, where the centered linear pool attains better scores than the linear pool. The better performance is mainly due to the bias reduction in the variance forecasts. The disagreement component is encompassed by the weighted individual variance forecasts only in one case. For the forecasts of monthly stock returns, the disagreement component is very small, leading to virtually identical results for both pools considered. However, the disagreement component is encompassed by the weighted individual variance forecasts in the prediction of the MSE of the combined mean forecast.

Since the construction of the centered linear pool is very simple, the theoretical and empirical results suggest to try it out as a promising alternative to the traditional linear pool.

## References

- BATES, J. M., AND C. W. J. GRANGER (1969): “The Combination of Forecasts,” *Journal of the Operational Research Society*, 20, 451–468.
- CARRIERO, A., T. E. CLARK, AND M. MARCELLINO (2016): “Common Drifting Volatility in Large Bayesian VARs,” *Journal of Business & Economic Statistics*, 34, 375–390.
- 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*, pp. 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.
- 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.
- GEWEKE, J., AND G. AMISANO (2011): “Optimal Prediction Pools,” *Journal of Econometrics*, 164, 130–141.

- 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.
- HALL, S. G., AND J. MITCHELL (2007): “Combining Density Forecasts,” *International Journal of Forecasting*, 23, 1–13.
- 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.
- KRÜGER, F., T. E. CLARK, AND F. RAVAZZOLO (forthcoming): “Using Entropic Tilt-  
ing to Combine BVAR Forecasts with External Nowcasts,” *Journal of Business & Economic Statistics*, pp. 1–48.
- MANSKI, C. F. (2015): “Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern,” *Journal of Economic Literature*, 53, 631–53.
- 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.
- STONE, M. (1961): “The Opinion Pool,” *The Annals of Mathematical Statistics*, 32, 1339–1342.
- WAGGONER, D. F., AND T. ZHA (2012): “Confronting Model Misspecification in Macroeconomics,” *Journal of Econometrics*, 171, 167–184.
- 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.

## Appendix

### A Derivations for baseline example

Here we prove that  $D$  and  $S_c$  are independent for  $\omega = 0.5$ . Consider the vector  $W = [X_1, X_2, U]'$ , which has mean zero and VCV

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

We can write  $W$  as  $W = C Z$ , where  $C$  is the lower-diagonal Cholesky matrix of  $\Omega$ , 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

$$(X_1 - X_2)^2 = \underbrace{([1 \ -1 \ 0] \times W)}_{=a'}^2 = Z' C' a a' C Z;$$

furthermore,

$$S_c = \underbrace{([\omega(1+\rho) \ (1-\omega)(1+\rho) \ 1] W)}_{=b'}^2 = Z' C' b b' C Z.$$

A standard result (Craig, 1943) states that  $D$  and  $S_c$  are independent if  $C' a a' C C' b b' C = 0$ . Simple but tedious algebra shows that this holds true 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. Note that we refer to A0 as the assumption that the prediction space framework of Section 3.3 applies.

Object	Expression	Underlying Assumption(s)
$\mathbb{E}(D)$	$Tr(A\Sigma_M) + \mu'_M A \mu_M$	
$\mathbb{E}(S)$	$Tr(B\Sigma_M) + (\mu'_M w)^2 + \Sigma_U$	A0
$\mathbb{E}(\omega'V)$	$\omega' \mu_V$	
$Cov(D, S)$	$2 \mathbb{E}(\eta^2) Tr(A\Sigma_M B \Sigma_M) + 4\mu'_M A \Sigma_M B \mu_M + V(\eta) \times Tr(A\Sigma_M) \times (Tr(B\Sigma_M) + \Sigma_U)$	A0, A5
$Cov(S, \omega'V)$	$\mathbb{V}(\eta) \omega' \mu_V \times (Tr(B\Sigma_M) + \Sigma_U)$	A0
$Cov(D, \omega'V)$	$\mathbb{V}(\eta) Tr(A\Sigma_M) \omega' \mu_V$	
$\mathbb{V}(\omega'V)$	$\mathbb{E}(\eta^2) \omega' \Sigma_V \omega + \mathbb{V}(\eta) (\omega' \mu_V)^2$	
$\mathbb{V}(D)$	$2\mathbb{E}(\eta^2) Tr(A\Sigma_M A \Sigma_M) + 4\mu'_M A \Sigma_M A \mu_M + \mathbb{V}(\eta) (Tr(A\Sigma_M))^2$	A5

Table 6: Formulas and underlying assumptions. Matrices  $A$  and  $B$  are defined as follows:  $A = diag(\omega) - \omega\omega'$ ,  $B = (\gamma - \omega)(\gamma - \omega)'$ .

## C Proofs of theorems

### Theorem 4.1

*Proof.* From A0 and A1, we have that  $\mu_{V,i}$  (expected variance of forecast  $i$ ) is given by

$$\mu_{V,i} = (\gamma - \iota_{(i)})' \Sigma_M (\gamma - \iota_{(i)}) + \Sigma_U + \iota'_{(i)} \mu_M \mu'_M \iota_{(i)},$$

where  $\iota_{(i)}$  is a vector of length  $n$ , with all elements equal to zero except the  $i$ th element equal to one. Stacking the expected variances in  $\mu_V = [\mu_{V,1}, \dots, \mu_{V,n}]'$ , we get

$$\mu_V = (\gamma' \Sigma_M \gamma + \Sigma_U) \iota + \widetilde{\Sigma}_M + \widetilde{\mu_M \mu'_M} - 2 \Sigma_M \gamma, \quad (9)$$

where  $\iota$  is an  $n \times 1$  vector of ones, and  $\widetilde{A}$  is an  $n \times 1$  vector containing the diagonal elements of matrix  $A$ . From (9) and A3, the weighted average variance is given by

$$\omega' \mu_V = \gamma' \Sigma_M \gamma + \Sigma_U + \omega' \widetilde{\Sigma}_M + \omega' \widetilde{\mu_M \mu'_M} - 2 \omega' \Sigma_M \gamma. \quad (10)$$

We then have that

$$\begin{aligned}
\mathbb{E}(D) + \mathbb{E}(S) &= Tr(diag(\omega)\Sigma_M) - \omega'\Sigma_M\omega + \mu'_M diag(\omega)\mu_M - (\omega'\mu_M)^2 + \gamma'\Sigma_M\gamma - 2\gamma'\Sigma_M\omega + \\
&\quad \omega'\Sigma_M\omega + (\omega'\mu_M)^2 + \Sigma_U, \\
&= \gamma'\Sigma_M\gamma + \Sigma_U + \omega'\widetilde{\Sigma}_M + \omega'\widetilde{\mu}_M\mu'_M - 2\omega'\Sigma_M\gamma,
\end{aligned} \tag{11}$$

where we use the formulas in Appendix B and the facts that i)  $Tr(ABC) = Tr(BCA)$  and ii) the trace of a scalar is the scalar itself. Thus,  $\omega'\mathbb{E}(V) = \mathbb{E}(D) + \mathbb{E}(S)$  and the proof is complete.  $\square$

## Theorem 4.2

*Proof.* Denote by  $M_{lp} = \omega'M$  and  $V_{lp} = \omega'V + \sum_{i=1}^n \omega_i(M_i - M_{lp})^2$  the mean and variance of the linear pool. Similarly, let  $M_{clp} = M_{lp}$  and  $V_{clp} = \omega'V$  denote the mean and variance of the centered linear pool. We have that

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

where the inequality in the first line follows from the convexity of the square function, the second line follows from the assumption that the weights  $\omega$  are deterministic, the third line follows from neutral dispersion of the components as stated in assumption A1, and the last two lines follow from the above expressions for the variances of both pooling methods.  $\square$

## Lemma for Theorem 4.3

**Lemma C.1.** *Under A0, A2 and A4, the following relations hold:*

- $Tr(A\Sigma_M B\Sigma_M) = 0$
- $\mu'_M A = 0$   $\iota'$

*Proof.* We first derive an expression for  $B = (\gamma - \omega^*)(\gamma - \omega^*)'$ , where  $\omega^*$  denotes the optimal combination weights. Since the weights  $\omega^*$  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

$$\omega^* = \gamma - \underbrace{\zeta_M^{-1} \iota (\iota' \zeta_M^{-1} \iota)^{-1}}_{=P} (\iota' \gamma - 1),$$

where  $\zeta_M = E(MM')$ . A2 implies that we can write  $\zeta_M = \kappa^2 \iota \iota' + \Sigma_M$  for some  $\kappa \in \mathbb{R}$ . From the inverse formula in Equation (160) of Petersen and Petersen (2012), we find that

$$\begin{aligned} \zeta_M^{-1} \iota &= \Sigma_M^{-1} \iota (1 + \kappa^2 \iota' \Sigma_M^{-1} \iota)^{-1}, \\ (\iota' \zeta_M^{-1} \iota)^{-1} &= (1 + \kappa^2 \iota' \Sigma_M^{-1} \iota)^{-1} (\iota' \Sigma_M^{-1} \iota)^{-1}, \end{aligned}$$

such that

$$P = \Sigma_M^{-1} \iota (\iota' \Sigma_M^{-1} \iota)^{-1},$$

i.e. the parameter  $\kappa$  does not enter the probability limit of the optimal weights. Hence,

$$\begin{aligned} B &\equiv (\gamma - \omega^*)(\gamma - \omega^*)' \\ &= \frac{(\iota' \gamma - 1)^2}{\underbrace{(\iota' \Sigma_M^{-1} \iota)^2}_{\equiv \underline{c} \in \mathbb{R}_+}} (\Sigma_M^{-1} \iota \iota' \Sigma_M^{-1}). \end{aligned} \tag{12}$$

Equation (12) and the definition  $A = [G - \omega^* \omega^{*'}]$  imply that

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

such that  $Tr(A \Sigma_M B \Sigma_M) = 0$ . We further have that  $\mu'_M A = \kappa \iota' [G - \omega^* \omega^{*'}] = \kappa [\omega^{*'} - \omega^{*'}] = 0 \iota'$ .

□

### Theorem 4.4

*Proof.* Note that under assumption A2+, the conditional expectations and variances of  $S$ ,  $D$  and  $\omega'V$  are given by

$$\begin{aligned}\mathbb{E}[S|\eta] &= \eta \times c_{\mathbb{E}[S]} \\ \mathbb{E}[D|\eta] &= \eta \times c_{\mathbb{E}[D]} \\ \mathbb{E}[\omega'V|\eta] &= \eta \times c_{\mathbb{E}[\omega'V]}\end{aligned}$$

and

$$\begin{aligned}\mathbb{V}[D|\eta] &= \eta^2 \times c_{\mathbb{V}[D]} \\ \mathbb{V}[\omega'V|\eta] &= \eta^2 \times \omega'\Sigma_V\omega\end{aligned}$$

where  $c_{\mathbb{E}[S]}$ ,  $c_{\mathbb{E}[D]}$ ,  $c_{\mathbb{E}[\omega'V]}$ , and  $c_{\mathbb{V}[D]}$  are constants referring to the respective quantities conditional on  $\eta = 1$ , i.e.  $c_{\mathbb{E}[D]} = \mathbb{E}[D|\eta = 1]$  etc. Moreover, under assumption A4 we have

$$Cov(S, D|\eta) = Cov(S, \omega'V|\eta) = Cov(D, \omega'V|\eta) = 0.$$

Rewriting the regression (7) as

$$S^* = c_1 D^* + c_2 (\omega'V)^* + \text{error}$$

with  $S^* = S/E[S|\eta]$ ,  $D^* = D/E[D|\eta]$ , and  $(\omega'V)^* = \omega'V/E[\omega'V|\eta]$ , the regressand and the regressors have an expectation of 1. The regressors  $D^*$  and  $(\omega'V)^*$  can be regarded as noisy measures of the quantities  $E[D^*] = E[(\omega'V)^*] = 1$ . Noting that  $c_1 = 1 - \lambda$  and  $c_2 = \lambda$  with  $\lambda \in \mathbb{R}$  if  $D^* = E[D^*]$  and  $(\omega'V)^* = E[(\omega'V)^*]$ , employing the standard formula for regressors with measurement errors yields

$$plim \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{bmatrix} \mathbb{V}[(\omega'V)^*] / (\mathbb{V}[(\omega'V)^*] \times \mathbb{V}[D^*] + \mathbb{V}[(\omega'V)^*] + \mathbb{V}[D^*]) \\ \mathbb{V}[D^*] / (\mathbb{V}[(\omega'V)^*] \times \mathbb{V}[D^*] + \mathbb{V}[(\omega'V)^*] + \mathbb{V}[D^*]) \end{bmatrix}$$

which shows that both coefficients  $c_1$  and  $c_2$  are subject to attenuation bias if  $\mathbb{V}[(\omega'V)^*] > 0$ , i.e. if  $\Sigma_V > 0$ . If  $\mathbb{V}[(\omega'V)^*] < \mathbb{V}[D^*]$ ,  $(\omega'V)^*$  is a more precise predictor (in a mean-squared-error sense) of  $S^*$  than  $D^*$ , resulting in  $plim(c_1) < plim(c_2)$ . The condition  $\mathbb{V}[(\omega'V)^*] < \mathbb{V}[D^*]$  can be rewritten in terms of the conditional coefficients of variation, yielding

$$\frac{\sqrt{\mathbb{V}[\omega'V|\eta]}}{\mathbb{E}[\omega'V|\eta]} < \frac{\sqrt{\mathbb{V}[D|\eta]}}{\mathbb{E}[D|\eta]}. \quad (13)$$

Noting that

$$\mathbb{V}[\omega'V|\eta] = \eta^2 \times \mathbb{V}[(\omega'V|\eta = 1)]$$

the *unconditional* coefficient of variation can be determined based on

$$\mathbb{V}[\omega'V] = \mathbb{E}[\eta^2] \times \mathbb{E}[(\omega'V|\eta = 1)^2] - (\mathbb{E}[\eta])^2 \times (\mathbb{E}[(\omega'V|\eta = 1)])^2.$$



Using  $\mathbb{E}[(\omega'V|\eta=1)] = \mathbb{E}[\omega'V]$  and  $\mathbb{E}[\eta] = 1$ , this can be rewritten as

$$\mathbb{V}[\omega'V] = \mathbb{E}[\eta^2] \times \left( \mathbb{V}[(\omega'V|\eta=1)] + (\mathbb{E}[\omega'V])^2 \right) - (E[\eta])^2 \times (\mathbb{E}[\omega'V])^2$$

yielding

$$\frac{\sqrt{\mathbb{V}[\omega'V]}}{\mathbb{E}[\omega'V]} = \sqrt{\mathbb{V}[\eta] (1 + C_{\omega'V}) + C_{\omega'V}} \quad (14)$$

where  $C_{\omega'V}$  denotes the squared *conditional* coefficient of variation of  $\omega'V$

$$C_{\omega'V} = \frac{\mathbb{V}[\omega'V|\eta]}{(\mathbb{E}[\omega'V|\eta])^2} = \frac{\mathbb{V}[\omega'V|\eta=1]}{(\mathbb{E}[\omega'V])^2}.$$

The same calculations can be performed replacing  $\omega'V$  by  $D$ , yielding

$$\frac{\sqrt{\mathbb{V}[D]}}{\mathbb{E}[D]} = \sqrt{\mathbb{V}[\eta] (1 + C_D) + C_D} \quad (15)$$

with  $C_D = \mathbb{V}[D|\eta] / (\mathbb{E}[D|\eta])^2$ . This implies that

$$\frac{\sqrt{\mathbb{V}[\omega'V]}}{\mathbb{E}[\omega'V]} < \frac{\sqrt{\mathbb{V}[D]}}{\mathbb{E}[D]}$$

holds iff the inequality (13) holds and vice versa. □

### Theorem 4.5

*Proof.* Concerning the *conditional* coefficient of variation of  $D$ , under A2+, the *conditional* expectation and variance of  $D$  are given by

$$\begin{aligned} \mathbb{E}[D|\eta] &= \eta \times \text{Tr}(A\Sigma_M) \\ \mathbb{V}[D|\eta] &= 2\eta^2 \times \text{Tr}(A\Sigma_M A\Sigma_M). \end{aligned}$$

If the number of forecasters equals  $n = 2$ ,  $A$  and  $\Sigma_M$  equal

$$\begin{aligned} A &= \begin{bmatrix} w(1-w) & -w(1-w) \\ -w(1-w) & w(1-w) \end{bmatrix} \\ \Sigma_M &= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}. \end{aligned}$$

with  $\omega = [w, 1-w]'$ . In this case, the eigenvalues of  $A\Sigma_M$  and  $A\Sigma_M A\Sigma_M$  are given by

$$\begin{aligned} \text{eig}(A\Sigma_M) &= \begin{bmatrix} w(1-w)(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) \\ 0 \end{bmatrix} \\ \text{eig}(A\Sigma_M A\Sigma_M) &= \begin{bmatrix} (w(1-w)(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}))^2 \\ 0 \end{bmatrix}. \end{aligned}$$

Since the trace equals the sum of the eigenvalues, the *conditional* coefficient of variation is given by

$$C_D = \frac{\sqrt{\mathbb{V}[D|\eta]}}{\mathbb{E}[D|\eta]} = \sqrt{2}.$$

Concerning the *conditional* coefficient of variation of  $w'V$ , under assumptions A2+ and A5, with  $\eta$  being constant and with  $n = 2$  forecasters, the joint distribution of the forecast errors  $q_1$  and  $q_2$  is given by

$$\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda_1^2 & \rho\lambda_1\lambda_2 \\ \rho\lambda_1\lambda_2 & \lambda_2^2 \end{bmatrix} \right),$$

where  $\rho$  denotes the correlation of the forecast errors. The squared coefficient of variation equals

$$\frac{\mathbb{V}[(\omega'V)^2]}{(\mathbb{E}[(\omega'V)])^2} = \frac{\mathbb{E}[(\omega'V)^2]}{(\mathbb{E}[(\omega'V)])^2} - 1$$

where numerator and denominator are given by

$$\begin{aligned} \mathbb{E}[(\omega'V)^2] &= \mathbb{E}[(w_1q_1^2 + w_2q_2^2)^2] \\ &= 3w_1^2\lambda_1^4 + 2(1 + 2\rho^2)w_1w_2\lambda_1^2\lambda_2^2 + 3w_2^2\lambda_2^4 \\ (\mathbb{E}[(\omega'V)])^2 &= (\mathbb{E}[w_1q_1^2 + w_2q_2^2])^2 \\ &= (w_1\lambda_1^2 + w_2\lambda_2^2)^2 \end{aligned}$$

with  $\omega = [w_1, w_2]'$ . Since

$$\partial \left( \mathbb{E}[(\omega'V)^2] / (\mathbb{E}[(\omega'V)])^2 - 1 \right) / \partial (\rho^2) > 0$$

setting  $\rho = 1$  or  $\rho = -1$  maximizes the squared coefficient of variation and yields for the numerator

$$\mathbb{E}[(\omega'V)^2 | \rho = 1] = 3(w_1\lambda_1^2 + w_2\lambda_2^2)^2.$$

Thus, the coefficient of variation cannot exceed

$$\sqrt{\frac{\mathbb{E}[(\omega'V)^2 | \rho = 1]}{(\mathbb{E}[(\omega'V)])^2} - 1} = \sqrt{\frac{3(w_1\lambda_1^2 + w_2\lambda_2^2)^2}{(w_1\lambda_1^2 + w_2\lambda_2^2)^2} - 1} = \sqrt{2}.$$

The same equality is obtained if  $w_1 = 0$  or  $w_2 = 0$ . □

## D Details on the SV model

This appendix provides details on the model used in Sections 6.1 and 6.2, which is of the form

$$Y_t = X'_{t-1}\beta + \varepsilon_t, \quad (16)$$

$$\varepsilon_t \sim \mathcal{N}(0, \exp(h_t)), \quad (17)$$

$$h_t = h_{t-1} + \nu_t, \quad (18)$$

$$\nu_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_\nu^2). \quad (19)$$

We denote the sample period by  $t = 1, \dots, T$ , and let  $X = [X_0 \ X_1 \ \dots \ X_{T-1}]'$  and  $Y = [Y_1, \dots, Y_T]$ .

### Sketch of Gibbs sampler

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

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

$$\bar{V}_\beta = \left( \underline{V}_\beta^{-1} + X'W^{(i-1)}X \right)^{-1}, \quad (20)$$

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

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

2. Draw  $h_1^{(i)}, \dots, h_T^{(i)}$  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 7).
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  $m_{T+1}^{(i)} = X'_{T+1}\beta^{(i)}$  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 7 summarizes the prior parameters. For the survey-based nowcast model in Section 6.1,  $X_t$  is a scalar containing the SPF nowcast at date  $t$ , and we strongly shrink its coefficient towards one. For the return prediction models in Section 6.2,  $X_t$  contains an intercept and one out of 14 financial predictors; we use loose priors for  $\beta$ . In all cases, we use a diffuse prior for (the initial log variance)  $h_0$ , and our choices of  $\underline{a}_\nu$  and  $\underline{b}_\nu$  allow for a considerable amount of stochastic volatility.

Parameter	Value	
	Sec. 6.1	Sec. 6.2
$\underline{m}_\beta$	1	$[0, 0]'$
$\underline{V}_\beta$	$1e - 10$	$1000 \times I_2$
$\underline{m}_{h_0}$		0
$\underline{V}_{h_0}$		100
$\underline{a}_\nu$		1
$\underline{b}_\nu$		5

Table 7: 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.