MAGYAR NEMZETI BANK

VIOLA MONOSTORINÉ GROLMUSZ

OPTIMAL FORECAST COMBINATION UNDER ASYMMETRIC LOSS AND REGIME-SWITCHING

MNB WORKING PAPERS | 3

2023
DECEMBER

MAGYAR NEMZETI BANK

OPTIMAL FORECAST COMBINATION UNDER ASYMMETRIC LOSS AND REGIME-SWITCHING

MNB WORKING PAPERS | 3

2023
DECEMBER

MNB Working Papers 2023/3

Optimal forecast combination under asymmetric loss and regime-switching *

(Előrejelzések optimális kombinálása aszimmetrikus veszteségfüggvény és változó rezsimek mellett)

Written by Viola Monostoriné Grolmusz**

Budapest, December 2023

Published by the Magyar Nemzeti Bank
Publisher in charge: Eszter Hergár
Krisztina körút 55., H-1013 Budapest
www.mnb.hu
ISSN 1585-5600 (online)

*I am heavily indebted to Róbert Lieli for his guiding work as my Ph.D. advisor. In addition, I am very grateful for Sergey Lychagin, Zsolt Darvas, Lajos Szabó, Péter Gábriel and Anna Naszódi for their valuable comments and suggestions.

**Magyar Nemzeti Bank (Central Bank of Hungary), e-mail: monostorinev@mnb.hu.

Contents

Abstract

Forecast combinations have been repeatedly shown to outperform individual professional forecasts and complicated time series models in accuracy. Their ease of use and accuracy makes them important tools for policy decisions. While simple combinations work remarkably well in some situations, time-varying combinations can be even more accurate in other real-life scenarios involving economic forecasts. This paper uses a regime switching framework to model the time-variation in forecast combination weights. I use an optimization problem based on asymmetric loss functions in deriving optimal forecast combination weights. The switching framework is based on the work of Elliott and Timmermann (2005), however I extend their setup by using asymmetric quadratic loss in the optimization problem. This is an important extension, since with my setup it is possible to quantify and analyze optimal forecast biases for different directions and levels of asymmetry in the loss function, contributing to the vast literature on forecast bias. I interpret the equations for the optimal weights through analytical examples and examine how the weights depend on the model parameters, the level of asymmetry of the loss function and the transition probabilities and starting state.

JEL: C53.

Keywords: Forecast combination, Loss functions, Time-varying combination weights, Markov switching.

Összefoglaló

A kombinált előrejelzések gyakran pontosabbnak bizonyulnak mind az egyedi szakértői előrejelzéseknél, mind a bonyolult idősoros modellek predikcióinál. Egyszerű használatuk és pontosságuk miatt a döntéshozók számára is fontos eszközök lehetnek. Míg bizonyos előrejelzési helyzetekben az egyszerű, statikus kombinációk is jól teljesítenek, egyéb életszerű gazdasági szituációkban az időben változó kombinációs súlyok adnak pontosabb predikciót. Ebben a tanulmányban változó rezsimek feltételezése mellett modellezem az előrejelzési kombinációs súlyok időbeli alakulását. A kombinációs súlyok kiszámítására használt optimalizációs problémában emellett aszimmetrikus veszteségfüggvényeket feltételezek. A felhasznált rezsimváltó modell Elliott és Timmermann (2005) munkáján alapul, azonban modelljüket kiterjesztem aszimmetrikus négyzetes veszteségfüggvények használatára. Tanulmányom az előrejelzési torzítás irodalmához is jelentősen hozzájárul, hiszen a használt általánosabb keretben számszerűsíthető és vizsgálható az előrejelzési torzítás optimális mértéke a veszteségfüggvény különböző mértékű és irányú aszimmetriája esetén. Analitikus példákon keresztül értelmezem az optimális súlyokat meghatározó egyenleteket, és megmutatom, hogyan függnek a súlyok a modellparaméterektől, a veszteségfüggvény aszimmetriájának szintjétől, valamint a rezsimváltó folyamat kezdő szintjétől és átmeneti valószínűségeitől.

1 Introduction

Forecast combinations have been repeatedly shown to outperform individual professional forecasts and complicated time series models in accuracy. Since the seminal paper of Bates and Granger (1969) that introduced optimal forecast combinations, many works have shown the theoretical and empirical benefits of using combined forecasts (see, among others the papers by Clemen (1989), Diebold and Lopez (1996), Chan et al. (1999), Dunis et al. (2000), Stock and Watson (1998, 1999), Timmermann (2006), Diebold-Shin (2019)). These benefits include diversification gains from combining forecasts whose forecast errors are not perfectly correlated with one another, approximating reality with many models of different nature that are not encompassed by one complicated model and the ease of combination versus using a highly complex forecasting model (Elliott and Timmermann (2005)).

While simple combinations work remarkably well in some situations, time-varying combinations can be even more accurate in other real-life scenarios involving economic forecasts. The ranking of individual models according to accuracy is likely to change over time, as shown by Stock and Watson (2003) and Aiolfi and Timmermann (2004), among others. One forecast might be the most accurate in a period of high economic growth, but be outperformed by another forecast in times of recession. Then a combination framework with time-varying weights would work better at forecasting throughout the business cycle than one with stable weights. The idea of using time-varying forecast combination weights was first introduced by Granger and Newbold (1973), and extended to a regression framework by Diebold and Pauly (1987).

This paper uses a regime switching framework to derive optimal combination weights. I use an optimization problem based on asymmetric loss functions in deriving optimal forecast combination weights. The switching framework is based on the work of Elliott and Timmermann (2005) however, I extend their setup by using asymmetric quadratic loss in the optimization problem. This is an important extension, since with my setup it is possible to quantify and analyze optimal forecast biases for different directions and levels of asymmetry in the loss function. At the same time, this chapter also extends the findings of Elliott and Timmermann (2004). In this paper, the authors characterize the optimal combination weights for the most commonly used alternatives to mean squared error loss, but do not include state-dependence. Thus, my main contribution is the combination of state dependence with an asymmetric loss function, which, to my knowledge, has not been addressed in the literature.

In this paper I study a forecaster's problem who has access to a set of individual forecasts and wants to combine them optimally in a regime switching environment under asymmetric loss. I derive the first order conditions for an optimal linear combination and provide a numerical procedure (akin to GMM) for computing them. I interpret the optimal weights through analytical examples and examine how the weights depend on the model parameters, the level of asymmetry of the loss function and the transition probabilities and starting state. I quantify the optimal forecast bias as a function of the asymmetry parameter of the forecaster's loss function, adding to the literature on forecast bias (see Mincer and Zarnowitz (1969), Holden and Peel (1990), Batchelor (2007), Elliott et al. (2008), Dovern and Janssen (2017)). In the following paragraphs, I motivate my choices for using a Markov-switching framework, asymmetric losses, and I give context on optimal biases in forecasts.

There are different methods of using time-varying weights in forecast combinations. Using rolling window regressions to determine the combination weights for every forecast period is a popular and methodologically straightforward choice. Time-varying parameter models could also be estimated using the Kalman filter. A third choice, proposed by Deutch et al. (1994) is to determine weights based on a regime-switching model with an observable state variable. Elliott and Timmermann (2005) compare these three methods in creating combination forecasts from surveys and time series models. The authors find that the last method is the most accurate in terms of mean squared forecast error. Using the regime-switching model also enables the researcher to analyze the optimal weights and forecast errors assuming different starting regimes and different transition probabilities between regimes. This makes it possible to draw conclusions on optimal forecast biases for different economic states.

Elliott and Timmermann (2005) derives optimal combination weights in a latent state regime switching environment. The authors illustrate the result with an empirical application combining survey and time series forecasts and comparing the accuracy of combination forecasts based on different time-varying weighting methods. In the derivation of the optimal switching weights, the authors assume mean squared (MSE) loss.

Mean squared loss is widely used in the literature due to the ease of computation, analytical convenience and its favorable statistical properties[1]. However, its use is difficult to justify on economic grounds and likely does not capture the true behavior of forecasters. The arguments against the use of symmetric loss functions go back to Granger and Newbold (1986) and are developed in more recent works such as Christoffersen and Diebold (1996, 1997), Granger and Pesaran (2000), Elliott et al (2005) and (2008), Patton and Timmermann (2007), Wang and Lee (2014). The use of asymmetric loss functions is based on the idea that forecasters could be averse to 'bad' outcomes: low real GDP growth, high inflation, etc., and they could incorporate this asymmetry into their forecasts. In another forecasting situation there might be different costs in overprediction versus underprediction of sales: overprediction can lead to higher inventory holding costs, while underprediction can lead to stockout costs, loss of reputation and revenues when the demand is too high (Elliott et al. 2008). The relative costs of overprediction versus underprediction depend on the preferences of the firm, and it is reasonable to believe that the preferences are asymmetric. The forecaster is likely to be aware of the asymmetric preferences (their salary could even depend on using the right - asymmetric - loss function and producing accurate forecasts as a result), and would therefore use an asymmetric loss function in their forecasts.

Biases in economic forecasts could also be related to asymmetric loss functions. It is well documented that survey forecasts are frequently biased (Mincer and Zarnowitz (1969), Holden and Peel (1990), Batchelor (2007), Elliott et al. (2008), Dovern and Janssen (2017)). The size and direction of the bias can depend on the affiliation of the professional forecaster, as well as on the current state of the business cycle. Elliott et al. (2008) examine US Survey of Professional Forecasters (SPF) and Livingston survey data on output growth and find that close to 30 percent of individual forecasts are biased at a 5 percent significance level. The authors also find that on average, forecasters are more likely to underpredict growth (suggesting that the cost of overprediction is higher than the cost of underprediction). The biases vary by the affiliation of the forecaster: academics have almost symmetric loss functions, while banking and industry economists rely on more asymmetric loss functions (Elliott et al. (2008)).

Recent research suggest that state dependence and asymmetric loss are potentially both at play in some economic forecasts. Dovern and Janssen (2017) examine systematic forecast biases over the business cycle. On a panel of forecasts for the annual real GDP growth rate in 19 advanced economies[2] (1990-2013), they find that on average, forecasters overestimate GDP growth. However, there is a substantial difference between forecasts for different business cycle states. Forecasts made for recession periods exhibit large negative forecast errors (in advance, forecasters overestimate the growth for these periods). By contrast, forecasts for recoveries show small positive errors, while forecasts for expansions are unbiased.

As an illustration, I have reproduced Figure 1. from the paper of Dovern and Janssen (2017) using a different data set. I have used Consensus Economics surveys for annual real GDP growth for 11 Easters European countries[3], for the period between 2007 and 2019. The forecast horizons used range from 3 months to 24 months.

The figure for Eastern European economies confirms the same results as Dovern and Janssen's example of 19 advanced countries: forecasts made for recessions exhibit large negative biases, forecasts for recoveries often underpredict growth, while forecasts for expansions are on average unbiased (Figure 1). The differences between forecast biases made for different periods are large and significant (see Batchelor (2007) and Dovern and Janssen (2017)). Time series forecasts also frequently exhibit biases, especially around business cycle turning points. When constructing forecast combinations, it would be beneficent to let the weights depend on the state of the economy, as well as allow the loss function to be asymmetric. This chapter introduces an optimal combination weighting scheme that meets these criteria.

The rest of this paper is organized as follows. Section 2 shows the theoretical setup and outlines the expected loss minimization problem in the general case. Section 3 describes the procedure used for deriving the optimal weights numerically. Section 4

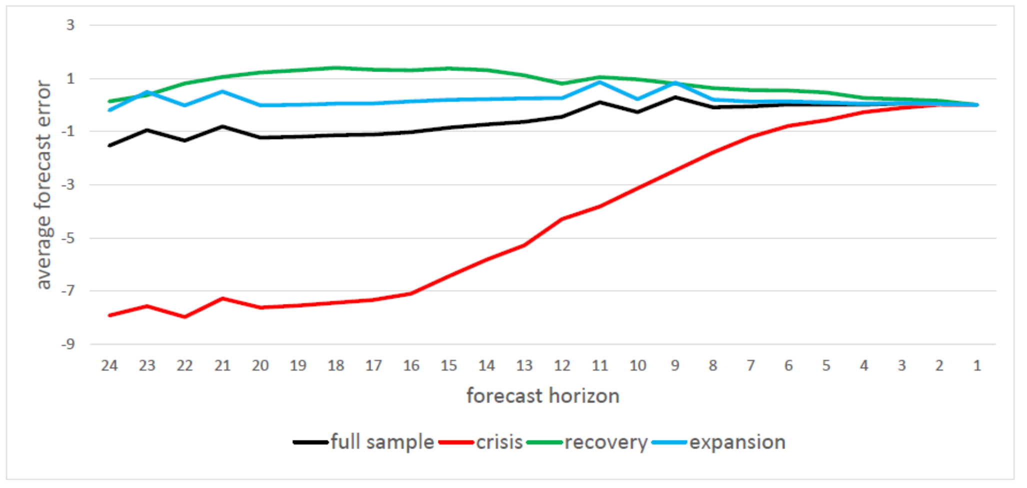
---

[1] When assuming MSE loss, the rational forecasts are unbiased and the forecast errors are uncorrelated with all variables in the current information set. Therefore, rationality testing is straightforward if quadratic loss is assumed. However, as Elliott et al. (2005, 2008) point out, testing rationality this way assumes a joint hypothesis of rationality and quadratic loss. The latter might not hold in many cases; the results of such rationality tests are not valid for forecasts constructed using asymmetric losses.

[2] Dovern and Janssen (2017) use Consensus Economics surveys for the following countries: Austria, Belgium, Canada, Switzerland, Germany, Denmark, Spain, Finland, France, Greece, Ireland, Italy, Japan, the Netherlands, Norway, Portugal, Sweden, the United Kingdom, and the United States.

[3] Bulgaria, Czech Republic, Estonia, Croatia, Hungary, Latvia, Lithuania, Poland, Romania, Slovenia, Slovakia

Figure 1
Systematic survey forecast errors by horizon



Notes: The figure shows estimates of the systematic forecast errors (in percentage points) as a function of the forecast horizon. The lines represent point estimates from regressions of the forecast errors on a set of 24 dummy variables (one for each forecast horizon).

analyzes how the optimal bias and the combination weights depend on the parameters through four analytical examples with different parametrizations. Section 5 assembles general observations from the results that could be formalized as theorems and also outlines some possible extensions. The last section concludes.

2 Theory

In the introduction, I have already argued for the high importance of allowing for asymmetric loss functions when combining forecasts. In this section, I introduce the theoretical setup and solve the expected loss minimization problem in the general case.

2.1 SETUP

We would like to forecast y_{t+1} on the basis of $I_t = \{\hat{\boldsymbol{y}}_{\tau+1}, y_\tau\}_{\tau=1}^{t}$, where

$$\hat{\boldsymbol{y}}_{\tau+1} = (\hat{y}_{1\tau+1}, \dots \hat{y}_{m\tau+1})' \tag{1}$$

is the vector of m individual forecasts. The information set includes the realized values of the target variable y_t up until the current period when the forecast is made, together with the past and current values of the m individual one-step-ahead forecasts. The last available individual forecasts in the forecaster's information set in t are the forecasts made in t for the $t+1$ horizon.

The equation for the linear combination of forecasts is the following:

$$y_{t+1} = \omega_0 + \boldsymbol{\omega}'\hat{\boldsymbol{y}}_{t+1} + e_{t+1}, \tag{2}$$

where ω_0 : is a constant, and $\boldsymbol{\omega}'$: is an m-vector of weights. The forecaster's goal is to optimally combine the individual forecasts in order to minimize her expected loss from the combined forecast. She can do this by optimizing the combination weights ω_0 and $\boldsymbol{\omega}'$ based on her specific loss function.

I assume that the joint distribution of the target y_{t+1} and the vector of individual forecasts $\hat{\boldsymbol{y}}_{t+1}$ is driven by an unobserved state variable, $S_t \in (1, \dots, k)$ that is not part of the information set; $S_t \notin I_t$. Conditional on the information set I_t and the underlying state $S_{t+1} = s_{t+1}$, assume that the joint distribution of the target and the vector of individual forecasts is Gaussian:

$$\begin{pmatrix} y_{t+1} \\ \hat{\mathbf{y}}_{\mathbf{t+1}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{ys_{t+1}} \\ \boldsymbol{\mu}_{\hat{\boldsymbol{y}}s_{t+1}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{ys_{t+1}} & \boldsymbol{\sigma}'_{y\hat{\boldsymbol{y}}s_{t+1}} \\ \boldsymbol{\sigma}_{y\hat{\boldsymbol{y}}s_{t+1}} & \boldsymbol{\Sigma}_{\hat{\boldsymbol{y}}s_{t+1}} \end{pmatrix}\right) \tag{3}$$

Given equation (2), assumption (3) implies that the corresponding conditional distribution of the error e_{t+1} is also Gaussian with some mean $\mu_{e_{s_{t+1}}}$ and standard deviation $\sigma_{e_{s_{t+1}}}$.

Finally, I also assume (following Hamilton (1989) and Elliott and Timmermann (2005)) that the states are generated by a first order Markov chain with the following transition probability matrix, where π_{ij} denotes the transition probability of arriving at state j when starting from state i:

$$\begin{pmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1k} \\ \pi_{21} & \pi_{22} & \dots & \pi_{2k} \\ \vdots & \vdots & \ddots & \pi_{k-1k} \\ \pi_{k1} & \dots & \pi_{kk-1} & \pi_{kk} \end{pmatrix} \tag{4}$$

Furthermore, if at time t the state of the process is s_t, then the probability that the process will transition to state s_{t+1} in period $t+1$ will be denoted as

$$P(s_{t+1}|s_t) = \pi_{s_{t+1},t}$$

Hence, $\pi_{s_{t+1},t}$ is the element of the matrix (4) that corresponds to row s_t and column s_{t+1}.

2.2 THE EXPECTED LOSS FUNCTION AND THE FORECASTER'S PROBLEM

Deviating from the setup of Elliott and Timmermann (2005), I choose the more flexible asymmetric quadratic (or quad-quad[4]) loss function in the forecaster's optimization problem, instead of the symmetry-assuming MSE loss[5]. The loss function takes the following form:

$$L(e) = \begin{cases} (1-\alpha)e^2, & \text{if } e > 0 \\ \alpha e^2, & \text{if } e \leq 0 \end{cases} \tag{5}$$

where $0 < \alpha < 1$. The parameter alpha in the loss function captures the asymmetry preferences of the forecaster. For alpha values lower than $\frac{1}{2}$, negative forecast errors entail a smaller cost for the forecaster as opposed to positive forecast errors, overprediction is preferred. For $\alpha > \frac{1}{2}$, positive errors entail smaller costs than negative errors, thus underprediction is preferred. $\alpha = \frac{1}{2}$ is the symmetric case, the loss function reduces to the same form as the mean squared error loss.

Assuming the loss function takes the form expressed in equation 5, the posited objective is to minimize the following expected loss formula:

$$E\{L(e_{t+1})|I_t, s_t\} = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} E\Big\{[\alpha - (2\alpha - 1)\mathbb{1}_{e_{s_{t+1}}>0}]e_{s_{t+1}}^2 \Big| I_t, s_{t+1}\Big\}, \tag{6}$$

where $\mathbb{1}_{e_{s_{t+1}}>0}$ denotes the indicator function, i.e.,

$$\mathbb{1}_{e_{s_{t+1}}>0} = \begin{cases} 1, & \text{if } e_{s_{t+1}} > 0 \\ 0, & \text{if } e_{s_{t+1}} \leq 0 \end{cases}$$

and $e_{s_{t+1}}$ is the (still random) value of the error e_{t+1} in state s_{t+1}.

Let us interpret the objective function in equation 6. The expectation on the left hand side is taken with respect to the conditional distribution of e_{t+1} given the forecaster's information set I_t and the current state s_t. This is then expanded as an iterated expectation on the right hand side. For any possible value s_{t+1} of the future state, the inner expectation is with respect to the conditional distribution of $e_{s_{t+1}}$ given I_t and s_{t+1}. This expectation is, by assumption, no longer dependent on s_t, i.e., it does not matter how the process arrives at the state s_{t+1}. The outer expectation then averages over all possible future states, using the transition probabilities corresponding to the current state s_t as weights (these are contained in the corresponding row of the transition matrix). This expectation, by contrast, is no longer dependent on I_t, as the Markov property implies that transition probabilities depend solely on the current state.[6]

To evaluate the expected loss (6) in practice, one needs to assume specific values for the transition probabilities $\pi_{s_{t+1},t}$ or estimate them based on an auxiliary model. There is a set of transition probabilities $\pi_{s_{t+1},t}$ corresponding to each possible current

---

[4] The double quadratic term refers to the type of the loss function for both negative and positive forecast errors.

[5] The asymmetric quadratic loss function I use in this chapter has been studied by other authors as well. It is a special case of the family of loss functions studied by Elliott, Komunjer and Timmermann (2005, 2008). In another paper, Elliott and Timmermann (2004) derive the optimal forecast combination in a permanent-state environment assuming the same loss function.

[6] We can summarize this discussion more formally as follows. Using the law of iterated expectations, we can write the left hand side of equation 6 as $E\{L(e_{t+1})|I_t, s_t\} = E\big\{E[L|I_t, s_t, s_{t+1}]\big|I_t, s_t\big\} = E\big\{E[L|I_t, s_{t+1}]\big|s_t\big\}$, where the last equality follows from the conditional independence conditions discussed above.

state s_t. However, s_t is not directly observed by the econometrician, which means that the evaluation of (6) also requires an assumption about the current state s_t or an estimate of it.

I now turn to the forecaster's problem. The forecaster's goal is to choose the combination weights ω_0 and $\boldsymbol{\omega}$ in equation (2) in a way that minimizes her expected loss (6). To this end, I write the value of the forecast error in state s_{t+1} as $e_{s_{t+1}} = \mu_{e_{s_{t+1}}} + \sigma_{e_{s_{t+1}}} z_{s_{t+1}}$, where $\mu_{e_{s_{t+1}}}$ and $\sigma_{e_{s_{t+1}}}$ are the state-specific mean and standard deviation, respectively, and $z_{s_{t+1}}$ is a standard normal random variable. Using equation (2) and assumption (3), these moments are given by

$$\mu_{e_{s_{t+1}}} = \mu_{y_{s_{t+1}}} - \omega_0 - \boldsymbol{\omega}' \boldsymbol{\mu}_{\hat{\mathbf{y}}_{s_{t+1}}}$$

$$\sigma^2_{e_{s_{t+1}}} = \sigma^2_{y_{s_{t+1}}} + \boldsymbol{\omega}' \boldsymbol{\Sigma}_{\hat{\mathbf{y}}_{s_{t+1}}} \boldsymbol{\omega} - 2\boldsymbol{\omega}' \boldsymbol{\sigma}_{y\hat{\mathbf{y}}_{s_{t+1}}}.$$

Substituting $e_{s_{t+1}} = \mu_{e_{s_{t+1}}} + \sigma_{e_{s_{t+1}}} z_{s_{t+1}}$ into (6) and making the corresponding change of variables in the integral yields the following expression:[7]

$$E\{L(e_{t+1})|I_t, s_t\} = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} E\{(\alpha - (2\alpha - 1)\mathbb{1}_{e_{s_{t+1}}>0}) e^2_{s_{t+1}} | I_t, s_t\} =$$

$$= \alpha \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} [\mu_e^2 + \sigma_e^2] - (2\alpha - 1) \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} \int_{-\frac{\mu_e}{\sigma_e}}^{\infty} (\mu_e + \sigma_e z_{s_{t+1}})^2 \mathrm{d}F(z_{s_{t+1}}), \tag{7}$$

where μ_e and σ_e are shorthand for $\mu_{e_{s_{t+1}}}$ and $\sigma_{e_{s_{t+1}}}$, respectively, and $F(\cdot)$ is the standard normal cumulative distribution function.

The goal is to minimize (7) with respect to the constant ω_0 and the slope coefficients (or weights) $\boldsymbol{\omega}$, where these parameters are implicit in the definition of μ_e and σ_e. However, as discussed above, the expected loss objective (6) has several 'versions' depending on the initial state s_t; there is, therefore, a corresponding set of minimizers for each possible current state. To emphasize this dependence, I will denote the optimal weights as ω^*_{0t} and $\boldsymbol{\omega}^*_t$. Thus, if the econometrician's assessment of the current state evolves from period to period, so do the optimal weights.

I will now characterize ω^*_{0t} and $\boldsymbol{\omega}^*_t$ as the solutions to the first order condition of the expected loss minimization problem outlined above.

2.3 EXPECTED LOSS MINIMIZATION IN THE GENERAL CASE

Let us minimize the expected loss function in the general case (7) by deriving the corresponding first order conditions (FOCs).

Taking the partial derivative with respect to the constant ω_{0t} yields:

$$\frac{\partial E\{L(e_{t+1})|s_t, I_t\}}{\partial \omega_{0t}} = 0 :$$

$$\alpha \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} (\mu_e) - (2\alpha - 1) \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} \left[\int_{-\frac{\mu_e}{\sigma_e}}^{\infty} (\mu_e + \sigma_e z_{s_{t+1}}) \mathrm{d}F(z_{s_{t+1}}) \right] = 0 \tag{8}$$

---

[7] See the detailed derivations in appendix A, equation 23.

Substituting μ_e and σ_e with their definitions (and omitting the state and time subscripts for clarity), we can write the FOC in the following form:

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{0t}} = 0: \\
&\alpha \sum_s \pi_s \left(\mu_{y_s} - \omega_{0t} - \boldsymbol{\omega}_t' \boldsymbol{\mu}_{\hat{\mathbf{y}}_s} \right) - \\
&\qquad (2\alpha - 1) \sum_s \pi_s \left[\int\limits_{\frac{-\mu_y + \omega_0 + \boldsymbol{\omega}_t' \boldsymbol{\mu}_{\hat{\mathbf{y}}_s}}{\sqrt{\sigma_y^2 + \boldsymbol{\omega}_t' \boldsymbol{\Sigma}_{\hat{\mathbf{y}}} \boldsymbol{\omega}_t - 2\boldsymbol{\omega}_t' \boldsymbol{\sigma}_{y\hat{\mathbf{y}}}}}}^{\infty} \left(\mu_{y_s} - \omega_{0t} - \boldsymbol{\omega}_t' \boldsymbol{\mu}_{\hat{\mathbf{y}}_s} + (\sigma_y^2 + \boldsymbol{\omega}_t' \boldsymbol{\Sigma}_{\hat{\mathbf{y}}} \boldsymbol{\omega}_t - 2\boldsymbol{\omega}_t' \boldsymbol{\sigma}_{y\hat{\mathbf{y}}})\, z \right) \mathrm{d}F(z) \right] = 0
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{0t}} = 0: \\
&\alpha \sum_s \pi_s \left(\mu_{y_s} - \omega_{0t} - \boldsymbol{\omega}_t' \boldsymbol{\mu}_{\hat{\mathbf{y}}_s} \right) - \\
&\qquad (2\alpha - 1) \sum_s \pi_s \left[\int\limits_{\frac{-\mu_y + \omega_0 + \boldsymbol{\omega}_t' \boldsymbol{\mu}_{\widehat{\mathbf{y}}_s}}{\sqrt{\sigma_y^2 + \boldsymbol{\omega}_t' \boldsymbol{\Sigma}_{\hat{\mathbf{y}}} \boldsymbol{\omega}_t - 2\boldsymbol{\omega}_t' \boldsymbol{\sigma}_{y\hat{\mathbf{y}}}}}}^{\infty} (\mu_{y_s} - \omega_{0t} - \boldsymbol{\omega}_t' \boldsymbol{\mu}_{\hat{\mathbf{y}}_s} + (\sigma_y^2 + \boldsymbol{\omega}_t' \boldsymbol{\Sigma}_{\hat{\mathbf{y}}} \boldsymbol{\omega}_t - 2\boldsymbol{\omega}_t' \boldsymbol{\sigma}_{y\hat{\mathbf{y}}})\, z) \mathrm{d}F(z) \right] = 0
\end{aligned}
\tag{10}
$$

The optimal weights ω_{0t}^* and $\boldsymbol{\omega}_t^*$ must then satisfy equation (10).

There are m more first order conditions corresponding to the partial derivatives with respect to the individual weights $\boldsymbol{\omega}_t$. These are given by:

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \boldsymbol{\omega}} = 0: \\
&\qquad \alpha \sum_s \pi_s \left(-\mu_{\hat{\mathbf{y}}} \mu_e + \boldsymbol{\Sigma}_{\hat{\mathbf{y}}} \boldsymbol{\omega} - \boldsymbol{\sigma}_{y\hat{\mathbf{y}}} \right) - \\
&- (2\alpha - 1) \sum_s \pi_s \left[\int\limits_{-\frac{\mu_e}{\sigma_e}}^{\infty} (\mu_e + \sigma_e\, z) \left(-\mu_{\hat{\mathbf{y}}} + \frac{1}{\sigma_e} \left(\boldsymbol{\Sigma}_{\hat{\mathbf{y}}} \boldsymbol{\omega} - \boldsymbol{\sigma}_{y\hat{\mathbf{y}}} \right)\, z \right) \mathrm{d}F(z) \right] = 0
\end{aligned}
\tag{11}
$$

The optimal weights ω_{0t}^* and $\boldsymbol{\omega}_t^*$ must also satisfy equation (11).

Due to the complexity of these equations, the solutions for the optimal weights cannot be given in closed form. However, it is possible to solve these equations numerically, adopting the idea behind the well-known generalized method of moments (GMM) estimator. I will describe this the general procedure in the next subsection. In Section 4 I will compute the optimal weights and consequent average losses in three specific scenarios and analyze the results in detail.

3 Numerical procedure for computing the weights

Suppose that all the parameters in equations (10) and (11) are given except for the weights ω_0 and $\boldsymbol{\omega}$. The main difficulty in solving the first order conditions lies in the evaluation of the integrals with respect to $dF(z)$, especially given that the integration limits are also dependent on the unknown weights. Let me generically represent these integrals as

$$\int_a^b g(z;\boldsymbol{\theta})dF(z), \tag{12}$$

where $\boldsymbol{\theta} = (\omega_0, \boldsymbol{\omega}')'$ stands for the vector of unknown weights and a and b may also depend on ω.

I then evaluate the first order conditions in the following way. First, I formally eliminate the integration limits by using indicator functions; that is, I represent the integrals $\int_a^b g(z;\boldsymbol{\theta})dF(z)$ as $\int g(z;\boldsymbol{\theta})\mathbb{1}_{[a,b]}(z)dF(z)$, where the latter integral is taken over the entire real line (i.e., from minus infinity to infinity). The two integrals are equal because the function $\mathbb{1}_{[a,b]}(z)$ is one if z falls into the interval $[a,b]$ and is zero otherwise.

Second, as F stands for the standard normal cdf, I can again regard these integrals as expectations over a standard normal random variable; that is,

$$\int g(z;\boldsymbol{\theta})\mathbb{1}_{[a,b]}(z)dF(z) = E\{g(Z;\boldsymbol{\theta})\mathbb{1}_{[a,b]}(Z)\}, \qquad Z \sim N(0,1). \tag{13}$$

Using this representation of the integrals with respect to $dF(z)$, the first order conditions (10) and (11) can be thought of as a set of moment conditions

$$E[m_j(Z;\boldsymbol{\theta})] = 0, \qquad j = 0,\dots,m, \tag{14}$$

where, for example,

$$\begin{aligned} m_0(Z,\boldsymbol{\theta}) &= \alpha \sum_s \pi_s \left(\mu_{y_s} - \omega_0 - \boldsymbol{\omega}'\boldsymbol{\mu}_{\hat{\mathbf{y}}_s}\right) \\ &- (2\alpha-1)\sum_s \pi_s \Big\{ \Big[\mu_{y_s} - \omega_0 - \boldsymbol{\omega}'\boldsymbol{\mu}_{\hat{\mathbf{y}}_s} + (\sigma^2_{y_s} + \boldsymbol{\omega}'\boldsymbol{\Sigma}_{\hat{\mathbf{y}}_s}\boldsymbol{\omega} - 2\boldsymbol{\omega}'\boldsymbol{\sigma}_{y\hat{\mathbf{y}}_s})\, Z\Big] \cdot \mathbb{1}_{[a_s,\infty)}(Z)\Big\} \end{aligned} \tag{15}$$

with

$$a_s = \frac{-\mu_{y_s} + \omega_0 + \boldsymbol{\omega}'\boldsymbol{\mu}_{\hat{\mathbf{y}}_s}}{\sqrt{\sigma^2_{y_s} + \boldsymbol{\omega}'\boldsymbol{\Sigma}_{\hat{\mathbf{y}}_s}\boldsymbol{\omega} - 2\boldsymbol{\omega}'\boldsymbol{\sigma}_{y\hat{\mathbf{y}}_s}}}. \tag{16}$$

Equations (10) and (11) can be written this way because the linearity of the expectations allows it to be pulled 'outside' of all the other operations.

Third, I replace the moment conditions with their 'empirical' counterparts using a large sample of artificial observations $Z_1,\dots,Z_n$ drawn from the standard normal distribution. That is, instead of expectations, I work with averages of the form

$$\frac{1}{n}\sum_{i=1}^{n} m_j(Z_i,\boldsymbol{\theta}) = 0, \qquad j = 0,\dots,m. \tag{17}$$

For large n, the law of large numbers guarantees $\frac{1}{n}\sum_{i=1}^{n} m_j(Z_i,\boldsymbol{\theta}) \approx E[m_j(Z,\boldsymbol{\theta})]$, and I can make this approximation precise by choosing n as large as computationally feasible.

Thus, in the three steps outlined above, I have reduced the computation of the optimal weights to a standard generalized method of moments (GMM) estimation problem, where the parameter vector $\boldsymbol{\theta} = (\omega_0, \boldsymbol{\omega}')'$ is just-identified. This means that one can use well-developed numerical procedures and readily available routines to compute the optimal forecast combination weights for any given parametrization of the forecaster's problem.

4 Analytical examples

In this section, I estimate and interpret the optimal weights and average losses in four different parametrizations. The simulations were carried out in order to better understand the differences between the asymmetry-allowing optimal combination weights and the ET combination weights that are based on MSE loss. For ease of interpretation, I consider only 2 states and 2 forecasts in all cases, and a one-period forecast horizon. State 1 parameters are different in the four cases, while they are always compared to the baseline parametrization in state 2 (state 2: unbiased forecasts, both variances are 1, forecasts are uncorrelated).

4.1 SCENARIO 1: ONE BIASED FORECAST

Let us assume a simple data generating process of the following form:

$$y_{t+1} = \beta_1^{s_t} x_t + \beta_2^{s_t} w_t + \epsilon_{t+1} = f1_t + f2_t + \epsilon_{t+1} \tag{18}$$

Where ϵ_{t+1} is a standard normal error term, $\epsilon_{s1} \sim N(0,1)$; $\epsilon_{s2} \sim N(0,1)$. The two individual forecasts that we would like to combine are the following:

$$\begin{aligned} f1_t &= \beta_1^{s_t} x_t \\ f2_t &= \beta_2^{s_t} w_t \end{aligned} \tag{19}$$

The linear combination of the two forecasts gives the combined forecast:

$$\hat{y}_{t+1|t} = \omega_{0t} + \omega_{1t} f1_t + \omega_{2t} f2_t \tag{20}$$

In the first parametrization, asymmetry is introduced by a small positive bias of forecast 1 in state 1 (see table 1 for the full parametrization). The other forecast stays unbiased throughout ($\mu_{f2,s1} = \mu_{f2,s2} = 0$). The variances of the forecasts are equal in both states and the two individual forecasts are uncorrelated $E(x_t w_t) = 0, \forall t$. The optimal weights are derived using the numerical procedure introduced in section 3.

I specialize the general expected loss function from equation 6 by substituting in the adoquate forms of μ_e and σ_e.

$$\begin{aligned} \mu_{e,s1} &= -\omega_{0,s1} - \omega_{1,s1}\, \mu_{f1,s1} = -\omega_{0,s1} - \omega_{1,s1}\, 0.1 \\ \mu_{e,s2} &= -\omega_{0,s1} \end{aligned}$$

$$\sigma^2_{e,s1} = \sigma^2_{e,s2} = 2 + (\omega^2_{1,s1} + \omega^2_{2,s1}) - 2(\omega_{1,s1} + \omega_{2,s1});$$

As the variances of the individual forecasts are unity and the forecasts are uncorrelated, the expected loss function and first order conditions are not overly complicated (see appendix B). The variance-covariance matrix of the two forecasts and the covariances between the target and the individual forecasts take the following forms:

Table 1
Scenario 1: one biased forecast

s_1		s_2	
β_1^{s1}	1	β_1^{s2}	1
β_2^{s1}	1	β_2^{s2}	1
μ_x^{s1}	0.1	μ_x^{s2}	0
σ_x^{s1}	1	σ_x^{s2}	1
μ_w^{s1}	0	μ_w^{s2}	0
σ_w^{s1}	1	σ_w^{s2}	1
μ_y^{s1}	0	μ_y^{s2}	0
σ_y^{s1}	$\sqrt{2}$	σ_y^{s2}	$\sqrt{2}$
$Cov(x,w)^{s1}$	0	$Cov(x,w)^{s2}$	0

$$\Sigma_{\hat{y}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \sigma_{y\hat{y}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

When there are two possible Markov states, then we get two sets of optimal weights, each referring to the starting state that is assumed to be known when making the forecast. The equation for the expected loss function and the first order conditions are stated in appendix B. Applying the GMM-based numerical procedure to these analytical results, we get the optimal combination weights outlined in table 2 and 3. For deriving the results in table 2, the symmetric transition probabilities from the matrix P_1 were used, while for the results in table 3, the asymmetric transition probabilities from P_2 were used.

$$P_1 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \qquad P_2 = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

First, let us interpret the results of table 2. The transition probability does not depend on the starting state in this scenario, therefore, the optimal weights are the same for each starting state. The optimal weights of the two individual forecasts, ω_{1t} and ω_{2t} are essentially 1 (minor estimation errors occur from the GMM procedure). This is the same as their true values, β_1 and β_2 from the DGP. The optimal bias is captured in ω_0, whose value changes as the asymmetry parameter α increases.

At $\alpha = 0.5$, the loss function is symmetric and coincides with the MSE loss. Therefore, we see that the estimated optimal weights are exactly the same in the asymmetry-allowing case and the MSE loss-based combination (ET). ω_0 takes the value that offsets the bias competely, resulting in an unbiased forecast:

$$\omega_0 = -(\text{forecast bias} \times Pr(\text{arriving in biased state})) \tag{21}$$

For lower α-s, overprediction is preferred. This is achieved in the combination forecast, by only sligthly offsetting the bias from f1; ω_0 is close to zero. As α increases toward 0.5, the preference for overprediction is weaker, therefore, ω_0 increases in absolute value, resulting in a less biased optimal combination forecast. For α-s above 0.5, underprediction is preferred, ω_0 offsets the bias coming from μ_{f1}, and produces an overall positive forecast error.

Figure 2 and 3 shows the average asymmetric quadratic losses for the transition probability matrices P_1 and P_2. The figures depict the result of a thought experiment where a one-period forecast is made and we would like to know the expected loss for the next period. In figure 2, the transition probability matrix P_1 results in a symmetric loss function that is always lower than the constant loss resulting from the ET optimal combination. The loss is lower for more extreme asymmetry preferences (α-s close to 0 and 1). Again, the starting state does not influence the results.

When the transition probability matrix takes the from of P_2, two different sets of optimal weights are calculated based on the starting state. Now, the system is likely to stay in the starting state (with probability 0.9). When this is the biased state 1, ω_0 needs to be higher in absolute value to offset the bias. The relation in equation 21 stays true; for instance when $\alpha = 0.5$, the constant from the optimal combination needs to be -0.09 to yield an unbiased forecast (this is also the optimal ω_0 for the ET loss). As the asymmetry parameter changes, we can see a similar dynamic in the change of ω_0 as in table 2: for lower α-s, the preferred overprediction of the target variable is achieved by only partly offsetting the bias from $f1$, while for α-s higher than 0.5, an ω_0 higher in absolute value is needed to produce an optimally biased combination forecast. The coefficients of $f1$ and $f2$ are 1 throughout, hitting the true β coefficients from the DGP.

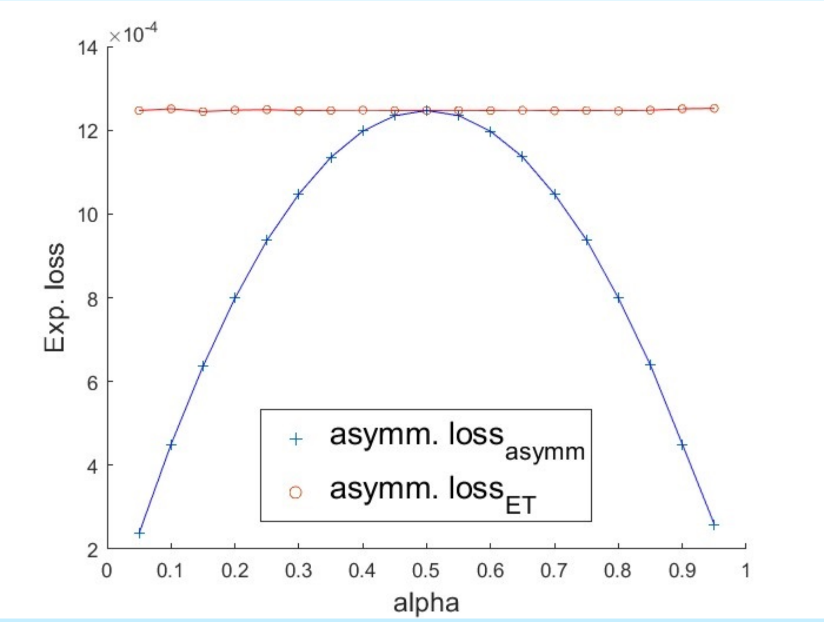
Table 2
Optimal weights from case 1, symmetric transition probabilities

	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.1	-0.010	1.000	1.000	-0.010	0.999	1.000	-0.050	0.998	1.000	-0.050	0.998	1.000
0.3	-0.030	0.998	1.000	-0.030	0.998	1.000	-0.050	0.998	1.000	-0.050	0.998	1.000
0.5	-0.050	0.998	1.000	-0.050	0.998	1.000	-0.050	0.998	1.000	-0.050	0.998	1.000
0.7	-0.070	0.997	1.000	-0.070	0.998	1.000	-0.050	0.998	1.000	-0.050	0.998	1.000
0.9	-0.090	1.000	1.000	-0.090	1.000	1.000	-0.050	0.998	1.000	-0.050	0.998	1.000

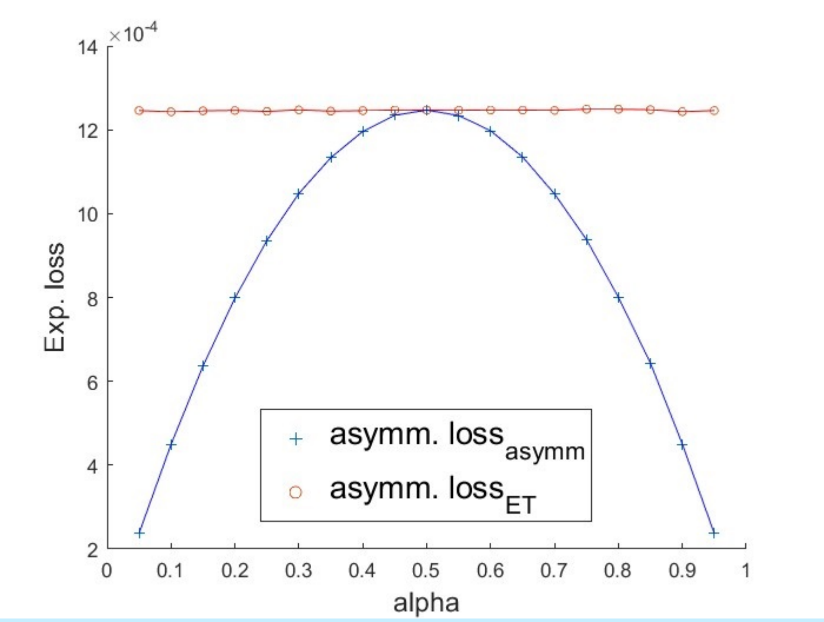
Table 3
Optimal weights from case 1, asymmetric transition probabilities

	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.1	-0.050	0.998	1.000	-0.001	1.000	1.000	-0.090	0.999	1.000	-0.010	0.999	1.000
0.3	-0.079	0.998	1.000	-0.005	1.000	1.000	-0.090	0.999	1.000	-0.010	0.999	1.000
0.5	-0.090	0.999	1.000	-0.010	0.999	1.000	-0.090	0.999	1.000	-0.010	0.999	1.000
0.7	-0.095	0.999	1.000	-0.021	0.998	1.000	-0.090	0.999	1.000	-0.010	0.999	1.000
0.9	-0.104	0.997	0.998	-0.050	0.993	1.000	-0.090	0.999	1.000	-0.010	0.999	1.000

Figure 2
Asymmetric qaudratic losses as a function of alpha based on parametrization with one biased forecast (s1); P=[0.5 0.5; 0.5 0.5]

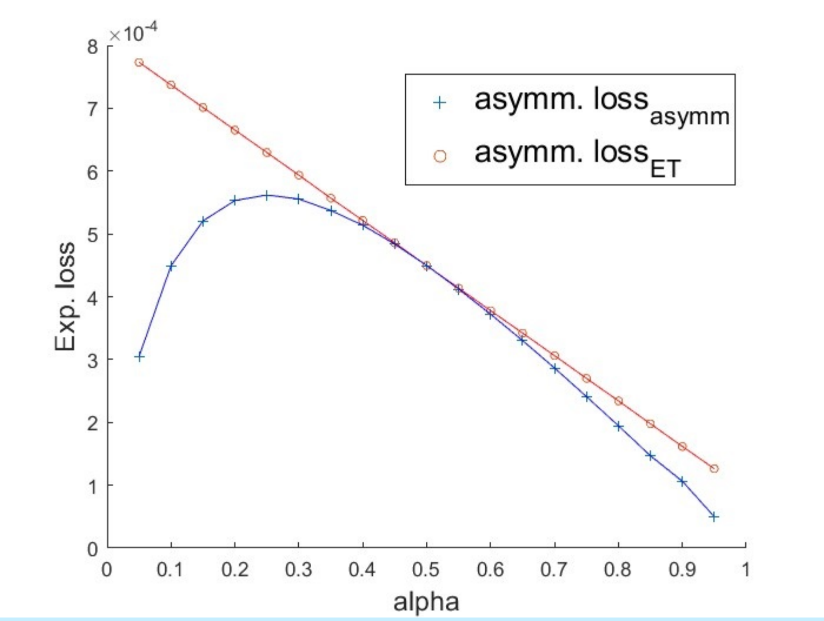


quad-quad losses, starting state: s1

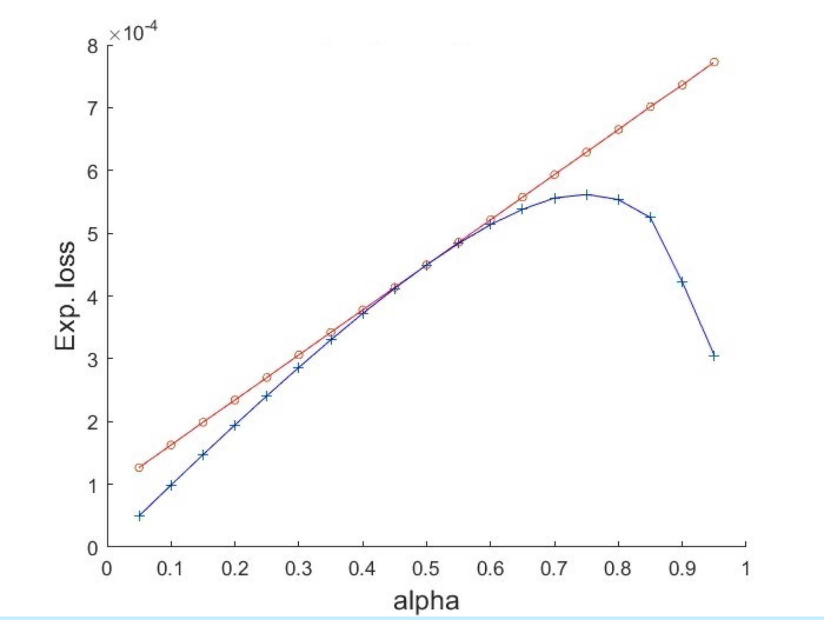


quad-quad losses, starting state: s2

Figure 3
Asymmetric qaudratic losses as a function of alpha based on parametrization with one biased forecast (s1); P=[0.9 0.1; 0.1 0.9]



quad-quad losses, starting state: s1



quad-quad losses, starting state: s2

4.2 SCENARIO 2: DIFFERENT VARIANCES OF INDIVIDUAL FORECASTS

In this scenario, both forecasts are unbiased throughout. The difference in the forecasts stems from the second forecast having a higher variance in state 1 (see table 4 for full parametrization). The two individual forecasts are uncorrelated. Again, state 2 is characterized by the baseline parametrization of equal variances and no bias. The state-dependent means and variances of the forecast error are the following:

$$\mu_{e,s1} = \mu_{e,s2} = -\omega_{0,s1}$$

$$\sigma^2_{e,s1} = 3 + (2\omega^2_{1,s1} + \omega^2_{2,s1}) - 2(2\omega_{1,s1} + \omega_{2,s1});$$
$$\sigma^2_{e,s2} = 2 + (\omega^2_{1,s1} + \omega^2_{2,s1}) - 2(\omega_{1,s1} + \omega_{2,s1});$$

In this parametrization, the variance-covariance matrix of the two forecasts and the covariances between the target and the individual forecasts are changed from the baseline to the following forms. The resulting expected loss function and first order conditions are detailed in appendix C.

$$\Sigma_{\hat{y}} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \qquad \sigma_{y\hat{y}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

The optimal combination weights are shown in tables 5 (symmetric transition probabilities characterized by P_1) and 6 (asymmetric transition probabilities characterized by P_2). It is appearent that the higher variance of $f1$ in s1 does not change the optimal weights, thus the true parameters stemming from the data generating proccess, $[\omega_0, \omega_1, \omega_2] = [0, 1, 1]$ are found. At extreme asymmetry parameters, the minor differences are due to calculation errors from the GMM procedure. When a bias is introduced to forecast 1 in state 1 in addition to the higher variance, the optimal combination weights are the same as in scenario 1.

4.3 SCENARIO 3: CORRELATED FORECASTS

Let us examine a parametrization with correlated individual forecasts in state 1. In state 1, f1 has an indirect effect on y, through its correlation with f2. Similarly to the other specifications, state 2 is characterized by the baseline DGP and forecasts.

$$y^{s1}_{t+1} = f2_t + \epsilon_{t+1}$$
$$y^{s2}_{t+1} = f1_t + f2_t + \epsilon_{t+1}$$

Where ϵ_{t+1} is a standard normal error term, $\epsilon_{s1} \sim N(0, 1)$, $\epsilon_{s2} \sim N(0, 1)$.

In state 1, each individual forecast consists of a common part, f, and an additional error term:

$$f1^{s1}_t = f_t + \zeta_t$$
$$f2^{s1}_t = f_t + \eta_t$$

$$\text{where } \zeta_t \sim N(0,1) \text{ and } \eta_t \sim N(0,0.2)$$

The forecast is the linear combination of the individual forecasts.

$$\hat{y}_{t+1|t} = \omega_{0t} + \omega_{1t} f1_t + \omega_{2t} f2_t$$

$$\Sigma_{\hat{y}} = \begin{bmatrix} 2 & 1 \\ 1 & 1.2 \end{bmatrix} \qquad \sigma_{y\hat{y}} = \begin{bmatrix} 1 \\ 1.2 \end{bmatrix}$$

$$\mu_{e,s1} = \mu_{e,s2} = -\omega_{0,s1}$$

$$\sigma^2_{e,s1} = 1.2 + 2\omega^2_{1,s1} + 1.2\omega^2_{2,s1} + 2(\omega_{1,s1}\omega_{2,s1}) - 2(\omega_{1,s1} + 1.2\omega_{2,s1});$$
$$\sigma^2_{e,s2} = 2 + (\omega^2_{1,s1} + \omega^2_{2,s1}) - 2(\omega_{1,s1} + \omega_{2,s1});$$

To better understand the results, assume first that there was no switching and the system stayed in s1. As f1 does not appear in the DGP, and has a higher variance than f2, we would expect the optimal combination weights to be $[\omega_0, \omega_1, \omega_2] = [0, 0, 1]$. In the simulation of such a case, whose results are presented in table 4.3, these weights are indeed found (at very high α-s we can see some estimation errors).

Returning to the original switching framework, let us first assume equal transition probabilities (transition probability matrix is P_1). Then the estimated optimal combination weights are those shown in table 4.3. As both forecasts are unbiased in both states, the weights do not change with the asymmetry parameter, similarly to scenario 2. ω_0 is zero throughout as there is no bias to offset coming from the individual forecasts. However, the optimal weights of the two forecasts differ in this case: the coefficient of f1 is lower (0.39) than that of f2 (0.82). The weights take values between their optimal values if the system always stayed in state 1; $[\omega_0, \omega_1, \omega_2] = [0, 0, 1]$, and their optimal values if the system always stayed in state 2; $[\omega_0, \omega_1, \omega_2] = [0, 1, 1]$. The estimated ω^*_{1t} and ω^*_{2t} are lower than the simple average of the above two sets of weights $[0.5, 1]$. This is due to the variance-minimizing objective of the forecast: the forecast with higher variance, f1, is assigned a lower combination weight. Since f2 is positively correlated to f1, it is also intuitive in light of the variance-minimizing objective that ω^*_{2t} is lower than 1.

Assuming a more persistent transition probability matrix, P_2, we can see from table 4.3 that the starting state matters for the optimal weights. When the starting state is s1, where the forecasts are correlated, f1 is assigned a low weight of 0.82 that is even lower than the probability of leaving the starting state ($P_{12} = 0.1$). ω^*_{2t} is slightly higher (0.92) than the probability of staying in state 1 ($P_{11} = 0.9$).

When the starting state is s2, the optimal weights are close to $(0, 1, 1)$, (optimal weights for a system that always stays in s2) as the probability of arriving at state 1 is low.

4.4 SCENARIO 4: COMMON FACTOR

In this scenario, the combination forecast in state 1 is again characterized by two correlated individual forecasts. In addition to these two forecasts, the data generating process includes a third forecast, f3, that is the common factor responsible for the correlation between f1 and f2. As in the other examples, state 2 is the baseline parametrization (two uncorrelated, unbiased forecasts with equal coefficients in the DGP):

$$y^{s1}_{t+1} = f1_t + f2_t + f3_t + \upsilon_{t+1}$$
$$y^{s2}_{t+1} = f1_t + f2_t + \epsilon_{t+1}$$

Where ϵ_{t+1} is an i.i.d. error, $\epsilon_{s2} \sim N(0,1)$. υ is the idiosynchratic error from the state 1 DGP with correlated variables, $\upsilon_{s1} \sim N(0,1)$.

In state 1, f1 and f2 consists of a common factor, f3, and an additional error term with different variances:

$$f1^{s1}_t = f3_t + \zeta_t$$
$$f2^{s1}_t = f3_t + \eta_t$$

$$\text{where } \zeta_t \sim N(0, 0.1) \text{ and } \eta_t \sim N(0, 9)$$

The forecast is the linear combination of forecasts f1 and f2.

$$\hat{y}_{t+1|t} = \omega_{0t} + \omega_{1t} f1_t + \omega_{2t} f2_t$$

Let us first examine the optimal weights in a constant-state system to better understand the results from the switching simulation. Assume that there is no switching and the prevailing state is always s1. Then, we would expect the optimization procedure to assign f2 lower weights than f1, due to the variance-minimizing objective.

$$\Sigma_{\hat{y}} = \begin{bmatrix} 1.1 & 1 \\ 1 & 10 \end{bmatrix} \qquad \sigma_{y\hat{y}} = \begin{bmatrix} 3.1 \\ 12 \end{bmatrix}$$

$$\mu_{e,s1} = \mu_{e,s2} = -\omega_{0,s1}$$

$$\sigma^2_{e,s1} = 16.1 + 1.1\omega^2_{1,s1} + 10\omega^2_{2,s1} + 2(\omega_{1,s1}\omega_{2,s1}) - 2(3.1\omega_{1,s1} + 12\omega_{2,s1});$$
$$\sigma^2_{e,s2} = 2 + (\omega^2_{1,s1} + \omega^2_{2,s1}) - 2(\omega_{1,s1} + \omega_{2,s1});$$

Table 4.4 shows the optimal weights from the no-switching exercise[8]. The results are intuitive: ω_0 is zero similarly to the other cases where the forecasts are unbiased, and the optimal weights do not change with α. This is also likely due to the unbiasedness of the forecasts (this conjecture and some other general observations from the results are summarized in section 5). As expected, f2 is assigned lower weights than f1, due to its higher variance. Still assuming a no-switching environment, if the system stayed in state 2 throughout, the optimal weights would be $[\omega_0, \omega_1, \omega_2] = [0, 1, 1]$, as we have seen in the previous examples.

---

[8] Results for α-s lower than 0.3 and higher than 0.7 are truncated from table 4.4, since the numerical procedure produced large estimation errors. The full table can be found in appendix F.

Returning to the regime-switching environment, let us first examine the optimal weights under symmmetric transition probabilities between states (P_2 transition probability matrix), shown in table F[9]. ω_0 is zero since the forecasts are unbiased. Also likely due to unbiasedness, the optimal weights are constant for different asymmetry parameter values. The optimal weight of f2, the forecast with the higher variance is lower than that of the other forecast. The optimal combination weights in table F are very close to the arithmetic means of the optimal weights from the previous no-switching exercises (s1: [0, 1.9, 0.01]; s2: [0, 1, 1]).

Table F shows the optimal weights assuming asymmetric transition probabilities (P_2)[10]. When the starting state is s1, the system is expected to stay in this state with a probability of 0.9, therefore, the optimal weights are close to the results from table 4.4. Conversely, when the starting state is s2, the optimal weights are close to [0,1,1], as in the scenario where the system stayed in s2 throughout.

---

[9] Results for α-s lower than 0.3 are truncated from table F due to estimation errors at these extreme values. The full table can be found in appendix F. At $\alpha = 0.7$ the outlier values are also likely due to estimation error.

[10] Again, the results for $\alpha = 0.1$ are truncated due to estimation errors, see the full table in appendix F.

Table 4
Scenario 2: one forecast has higher variance in state 1

s_1		s_2	
β_1^{s1}	1	β_1^{s2}	1
β_2^{s1}	1	β_2^{s2}	1
μ_x^{s1}	0	μ_x^{s2}	0
σ_x^{s1}	$\sqrt{2}$	σ_x^{s2}	1
μ_w^{s1}	0	μ_w^{s2}	0
σ_w^{s1}	1	σ_w^{s2}	1
μ_y^{s1}	0	μ_y^{s2}	0
σ_y^{s1}	$\sqrt{3}$	σ_y^{s2}	$\sqrt{2}$
$Cov(x,w)^{s1}$	0	$Cov(x,w)^{s2}$	0

Table 5
Optimal weights from case 2, symmetric transition probabilities

	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.1	0.000	1.000	1.000	0.003	1.000	0.999	0.000	1.000	1.000	0.000	1.000	1.000
0.3	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000
0.5	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000
0.7	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000
0.9	0.000	1.000	1.000	-0.004	1.000	0.999	0.000	1.000	1.000	0.000	1.000	1.000

Table 6
Optimal weights from case 2, asymmetric transition probabilities

	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.1	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000
0.3	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000
0.5	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000
0.7	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000
0.9	-0.008	1.002	1.002	0.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000

Table 7
Scenario 3: correlated forecasts

s_1		s_2	
β_1^{s1}	1	β_1^{s2}	1
β_2^{s1}	1	β_2^{s2}	1
μ_x^{s1}	0	μ_x^{s2}	0
σ_x^{s1}	$\sqrt{2}$	σ_x^{s2}	1
μ_w^{s1}	0	μ_w^{s2}	0
σ_w^{s1}	1	σ_w^{s2}	1
μ_y^{s1}	0	μ_y^{s2}	0
σ_y^{s1}	$\sqrt{3}$	σ_y^{s2}	$\sqrt{2}$
$Cov(x,w)^{s1}$	0	$Cov(x,w)^{s2}$	0

Table 8
Optimal weights from case 3, only one state (s1)

	optimal weights		
α	starting state: s1		
	ω_{0t}	ω_{1t}	ω_{2t}
0.1	0.000	0.000	1.000
0.3	0.000	0.000	1.000
0.5	0.000	0.000	1.000
0.7	0.000	0.000	1.000
0.9	-0.052	-0.226	1.379

Table 9
Optimal weights from case 3, symmetric transition probabilities

	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.1	-0.010	0.392	0.821	0.000	0.393	0.821	0.000	0.393	0.821	0.000	0.393	0.821
0.3	-0.002	0.393	0.821	0.001	0.393	0.821	0.000	0.393	0.821	0.000	0.393	0.821
0.5	0.000	0.393	0.821	0.000	0.393	0.821	0.000	0.393	0.821	0.000	0.393	0.821
0.7	-0.001	0.393	0.821	-0.006	0.393	0.821	0.000	0.393	0.821	0.000	0.393	0.821
0.9	-0.007	0.393	0.821	-0.017	0.395	0.821	0.000	0.393	0.821	0.000	0.393	0.821

Table 10
Optimal weights from case 3, asymmetric transition probabilities

	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.1	0.001	0.083	0.937	0.005	0.824	0.918	0.000	0.082	0.937	0.000	0.826	0.919
0.3	0.001	0.083	0.937	-0.001	0.826	0.919	0.000	0.082	0.937	0.000	0.826	0.919
0.5	0.000	0.082	0.937	0.000	0.826	0.919	0.000	0.082	0.937	0.000	0.826	0.919
0.7	0.000	0.082	0.937	0.001	0.826	0.919	0.000	0.082	0.937	0.000	0.826	0.919
0.9	-0.079	0.115	0.929	-0.005	0.824	0.918	0.000	0.082	0.937	0.000	0.826	0.919

Table 11
Scenario 2: one forecast has higher variance in state 1

s_1		s_2	
β_1^{s1}	1	β_1^{s2}	1
β_2^{s1}	1	β_2^{s2}	1
β_3^{s1}	0	β_3^{s2}	0
μ_{f1}^{s1}	0	μ_{f1}^{s2}	0
σ_{f1}^{s1}	$\sqrt{1.1}$	σ_{f1}^{s2}	1
μ_{f2}^{s1}	0	μ_{f2}^{s2}	0
σ_{f2}^{s1}	$\sqrt{10}$	σ_{f2}^{s2}	1
μ_{f3}^{s1}	0	μ_{f3}^{s2}	0
σ_{f3}^{s1}	1	σ_{f3}^{s2}	1
μ_y^{s1}	0	μ_y^{s2}	0
σ_y^{s1}	$\sqrt{16.1}$	σ_y^{s2}	$\sqrt{2}$
$Cov(x,w)^{s1}$	0	$Cov(x,w)^{s2}$	0

Table 12
Optimal weights from case 4, only one state (s1)

	optimal weights		
α	starting state: s1		
	ω_{0t}	ω_{1t}	ω_{2t}
0.3	0.000	1.900	1.010
0.5	0.000	1.900	1.010
0.7	0.000	1.900	1.010

Table 13
Optimal weights from case 4, symmetric transition probabilities

	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.3	-0.001	1.452	1.050	-0.001	1.452	1.050	0.000	1.452	1.050	0.000	1.452	1.050
0.5	0.000	1.452	1.050	0.000	1.452	1.050	0.000	1.452	1.050	0.000	1.452	1.050
0.7	0.089	0.490	1.162	0.090	0.481	1.167	0.000	1.452	1.050	0.000	1.452	1.050
0.9	0.008	1.449	1.050	0.098	0.479	1.176	0.000	1.452	1.050	0.000	1.452	1.050

Table 14
Optimal weights from case 4, asymmetric transition probabilities

	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.3	0.001	1.810	1.019	-0.001	1.094	1.048	0.000	1.810	1.019	0.000	1.094	1.048
0.5	0.000	1.810	1.019	0.000	1.094	1.048	0.000	1.810	1.019	0.000	1.094	1.048
0.7	0.001	1.810	1.019	0.000	1.094	1.048	0.000	1.810	1.019	0.000	1.094	1.048
0.9	0.000	1.810	1.019	-0.002	1.095	1.048	0.000	1.810	1.019	0.000	1.094	1.048

5 Conjectures

In this section I assemble general observations from section 4 that could be formalized as theorems given further evidence.

1. If the individual forecasts are unbiased, the optimal combination weights do not depend on the loss function's asymmetry parameter. In case 1, we have seen that the constant term in the forecast combination, ω_0 changed as α increased. However, in the other three scenarios, the optimal weights were constant despite changing asymmetry preferences. In scenarios 2 through 4, both forecasts were unbiased, only their variances and covariance changed. When cases 1 and 2 were combined (f1 was biased and had higher variance in s1), the resulting optimal weights were identical to the results from case 1; again, the optimal bias captured by ω_0 was different for different α-s.
2. If one of the individual forecasts are biased, the bias is adjusted for through ω_0, the constant in the combination. The optimal combination weight of the biased individual forecast is its true weight from the data generating process (conjecture from scenario 1).
3. If the individual forecasts are uncorrelated and unbiased, the difference in their variances does not lead to differences in their optimal combination weights. In case 2, we have seen that for such parametrization, the forecasts were assigned their true coefficients from the DGP as combination weights.
4. If f1 and f2 are correlated and have different variances, then the variance-minimization objective is taken into account in estimating their optimal weights. The individual forecast with higher variance is assigned a lower weight.

6 Conclusion

This paper uses a regime switching framework and assumes asymmetric quadratic loss function to derive the optimal combination weights of individual forecasts. The switching framework is based on the paper of Elliott and Timmermann (2005), however I extend their setup by using asymmetric quadratic loss in the optimization problem. This is an important extension, since with my setup it is possible to quantify and analyze optimal forecast biases for different directions and levels of asymmetry in the loss function, contributing to the literature on rational forecast bias.

After introducing the expected loss function and first order conditions in the general case, I present the numerical procedure used to calculate the optimal weights in specific parametrizations. The optimal forecast combination weights are calculated in four scenarios exhibiting different bias, variance and covariance properties between the individual forecasts. The general observations from these examples are summerized in section 5. The most important conjecture is that assuming an asymmetric quadratic loss function and regime switching, the optimal combination weights depend on the asymmetry parameter only in the case when one of the forecasts are biased. In this case, for asymmetric preferences, the average loss based on the asymmetric quadratic loss function strongly dominates the MSE-based average loss.

If the individual forecasts are unbiased and only their variances differ (in both uncorrelated and correlated scenarios), then the optimal weights resulting from the asymmetric loss function are the same as those resulting from the mean squared loss. The optimal weights are independent from the asymmetry parameter.

In future work, conducting simulations calibrated to the real economy and analyzing the performance of the optimal forecast combinations introduced here might prove important.

References

Aiolfi, M. and A. Timmermann (2004). Persistence of Forecasting Performance and Combination Strategies, *Mimeo*, UCSD.

Batchelor, R. (2007). Bias in Macroeconomic Forecasts, *International Journal of Forecasting, 23*(2), 189-203.

Bates, J. M. and C. W. J. Granger, (1969). The Combination of Forecasts, *Journal of the Operational Research Society, 20*, (4), 451-468.

Chan, Y. L., J. Stock and M. W. Watson (1999). A Dynamic Factor Model Framework for Forecast Combination. *Spanish Economic Review, 1*, 91–121.

Christoffersen, P. F. and F. X. Diebold (1996). Further Results on Forecasting and Model Se-lection under Asymmetric Loss. *Journal of Applied Economics, 11*, 561-571.

Christoffersen, P. F. and F. X. Diebold (1997). Optimal Prediction Under Asymmetric Loss. *Econometric Theory, 13*(6), 808-817.

Clemen, R. T. (1989). Combining Forecasts: a Review and Annotated Bibliography. *International Journal of Forecasting, 5*, 559–581.

Deutsch, M., C. W. J. Granger and T. Terasvirta (1994). The Combination of Forecasts Using Changing Weights. *International Journal of Forecasting 10*, 47–57.

Diebold, F. X. and A. Lopez (1996). Forecast Evaluation and Combination.
In G. S. Maddala, and C. R. Rao (eds.), *Handbook of Statistics.* Elsevier.

Diebold, F. X. and P. Pauly (1987). Structural Change and the Combination of Forecasts. *Journal of Forecasting, 6*, 21–40.

Diebold, F. X. and M. Shin (2019). Machine Learning for Regularized Survey Forecast Combina-tion: Partially-Egalitarian LASSO and its Derivatives. *International Journal of Forecasting, 35*(4), 1679-1691.

Dovern, J. and N. Jannssen (2017). Systematic Errors in Growth Expectations over the Business Cycle. *International Journal of Forecasting, 33* (4), 760-769.

Dunis, C., J. Laws and S. Chauvin (2000). The Use of Market Data and Model Combinations to Improve Forecast Accuracy. presented at *CF2000 Conference* in London, 2000.

Elliott, G. and A. Timmermann (2004). Optimal Forecast Combinations under General Loss Func-tions and Forecast Error Distributions. *Journal of Econometrics, 122*, 47–79.

Elliott, G. and A. Timmermann (2005). Optimal Forecast Combination Under Regime Switching. *International Economic Review, 46* (4), 1081-1102.

Elliott, G., I. Komunjer and A. Timmermann (2005). Estimation and Testing of Forecast Rationality under Flexible Loss. *Review of Economic Studies, 72*, 1107–1125.

Elliott, G., I. Komunjer and A. Timmermann (2008). Biases in Macroeconomic Forecasts: Irrationality or Asymmetric

Loss. *Journal of European Economic Association, 6*, 122–157.

Granger, C. W. J. and P. Newbold (1973). Some Comments on the Evaluation of Economic Forecasts, *Applied Economics, 5*(1), 35-47.

Granger, C. W. J. and P. Newbold (1986). Forecasting Economic Time Series, 2nd Edition. Academic Press, New York.

Granger, C. W. J. and M. H. Pesaran (2000). Economic and Statistical Measures of Forecast Ac-curacy. *Journal of Forecasting, 19*, 537–560.

Holden, K. and D. Peel (1990). On Testing for Unbiasedness and Efficiency of Forecasts. *The Manchester School of Economic and Social Studies, 58*(2), 120-27.

Mincer, J. and V. Zarnowitz (1969). The Evaluation of Economic Forecasts.
In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performan-ce*, National Bureau of Economic Research, Inc, 3-46.

Patton, A. J. and A. Timmermann (2007). Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity. *Journal of Econometrics, 140*(2), 884-918.

Stock, J.H. and M. W. Watson (1998). A Comparison of Linear and Nonlinear Models for Forecasting Macroeconomic Time Series.
In Engle, R. and H. White (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press.

Stock, J. H. and M. W. Watson (1999). Forecasting inf1ation. *Journal of Monetary Econo-mics 44*, 293–375.

Stock, J. H. and M.W. Watson (2003). Combination Forecasts of Output Growth in a Seven-Country Data Set. *Mimeo*, Harvard University.

Timmermann, A. (2006). Forecast Combinations, Chapter. 04, p. 135-196
In Elliott, G., C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Elsevier.

Wang, Y, and T. H. Lee (2014). Asymmetric Loss in the Greenbook and the Survey of Professio-nal Forecasters. *International Journal of Forecasting, 30* (2), 235-245.

Appendix A Derivation of the general expected loss function

In this appendix, I show the derivation of the general expected loss function in equation 7 by substituting $e_{s_{t+1}} = \mu_{e_{s_{t+1}}} + \sigma_{e_{s_{t+1}}} z_{s_{t+1}}$.

The forecaster needs to minimize the following expected loss:

$$E\{L(e_{t+1})|I_t, s_{t+1}\} = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} E\{((\alpha - (2\alpha - 1)\mathbb{1}_{e_{s_{t+1}}>0})(e^2_{s_{t+1}}))|I_t\} \rightarrow \min \tag{22}$$

For simplifying notation, I am going to remove the s_{t+1} subscripts from $e_{s_{t+1}}$ for the following equations: e.g. μ_e means $\mu_{e_{s_{t+1}}}$.

Note that $z_{s_{t+1}} = \frac{e_{s_{t+1}} - \mu_{e_{s_{t+1}}}}{\sigma_{e_{s_{t+1}}}}$ is the standardized forecast error. $E[z_{s_{t+1}}] = 0$; $E[z^2_{s_{t+1}}] = 1$ Taking the expected value into parts:

$$E\{L(e_{t+1})|I_t, s_{t+1}\} = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} E\{(\alpha - (2\alpha - 1)\mathbb{1}_{e_{s_{t+1}}>0})[e^2_{s_{t+1}}]\} = \tag{23}$$

$$= \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t}\ \alpha\ E[\mu_e^2 + \sigma_e^2 z^2_{s_{t+1}} + 2\mu_e\sigma_e z_{s_{t+1}}] - (2\alpha - 1) \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} E[\mathbb{1}_{e_{s_{t+1}}>0}\ e^2_{s_{t+1}}] =$$

$$\overset{\substack{E(z^2_{s_{t+1}})=1;\ E(z_{s_{t+1}})=0 \\ \downarrow}}{=} \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t}\ \alpha\ E[\mu_e^2 + \sigma_e^2] - (2\alpha - 1) \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} \int_{0}^{\infty} e^2_{s_{t+1}}\ \mathrm{d}F(e_{s_{t+1}}) =$$

$$\overset{\substack{\text{changing variables in the integral} \\ \downarrow}}{=} \alpha \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t}[\mu_e^2 + \sigma_e^2] - (2\alpha - 1) \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} \int_{-\frac{\mu_e}{\sigma_e}}^{\infty} (\mu_e + \sigma_e z_{s_{t+1}})^2 \mathrm{d}F(z_{s_{t+1}})$$

Appendix B Equations for special case 1

In this specification, f1 has an upward bias of 0.1 on state 1 (see table 1 and equations 10-12 for full specification). The expected loss takes the following form:

$$E\{L(e)|I, s1\} = \alpha[P_{11}(\mu_{e,s1}^2 + \sigma_{e,s1}^2) + P_{12}(\mu_{e,s2}^2 + \sigma_{e,s2}^2)] -$$

$$-(2\alpha - 1)\left[P_{1,1}\int_{-\frac{\mu_{e,s1}}{\sigma_{e,s1}}}^{\infty}(\mu_{e,s1} + \sigma_{e,s1}z)^2 \mathrm{d}F(z) + P_{1,2}\int_{-\frac{\mu_{e,s2}}{\sigma_{e,s2}}}^{\infty}(\mu_{e,s2} + \sigma_{e,s2}z)^2 \mathrm{d}F(z)\right] =$$

$$= \alpha\left\{P_{11}[(-\omega_{0,s1} - \omega_{1,s1}\,\mu_{f1,s1})^2 + (2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1}))] + \right.$$

$$\left. + P_{12}[(-\omega_{0,s1})^2 + (2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1}))]\right\} -$$

$$-(2\alpha - 1)\left[P_{11}\int_{\frac{\omega_{0,s1} + \omega_{1,s1}\,\mu_{f1,s1}}{\sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}}}^{\infty}\left(-\omega_{0,s1} - \omega_{1,s1}\,\mu_{f1,s1} + \right.\right.$$

$$\left. + \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})z}\right)^2 \mathrm{d}F(z) +$$

$$\left. + P_{12}\int_{\frac{\omega_{0,s1}}{\sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}}}^{\infty}(-\omega_{0,s1} + \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})z})^2 \mathrm{d}F(z)\right] \tag{24}$$

When the starting state is assumed to be s2, we get a similar expected loss function to equation 24, but the transition probailities P_{21} and P_{22} are used in place of P_{11} and P_{12}. When all elements of the transition probability matrix are 0.5, the two sets of weights are equal.

Minimizing the expected loss in equation 24 yields the following first order conditions:

$$\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_0} = 0 : \\
&\alpha\left[P_{11}(\omega_{0,s1} + \omega_{1,s1}\mu_{f1,s1}) + P_{12}(\omega_{0,s1})\right] - \\
&-(2\alpha-1)\left\{P_{11}\left[\int_{\frac{\omega_{0,s1}+\omega_{1,s1}\,\mu_{f1,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} \omega_{0,s1} + \omega_{1,s1}\,\mu_{f1,s1} - \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}\, z\,\mathrm{d}F(z)\right] + \right. \\
&\left. P_{12}\left[\int_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} \omega_{0,s1} - \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}\, z\,\mathrm{d}F(z)\right]\right\}
\end{aligned} \tag{25}$$

The optimal weights for the two individual forecasts, ω_1 and ω_2 are determined by solving first order conditions 26 and 27:

$$\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{1,s1}} = 0 : \\
&\alpha\left[P_{11}[(\omega_{0,s1} + \omega_{1,s1}\mu_{f1,s1})\mu_{f1,s1} + \omega_{1,s1} - 1] + P_{12}(\omega_{1,s1} - 1)\right] - \\
&-(2\alpha-1)\left\{P_{11}\left[\int_{\frac{\omega_{0,s1}+\omega_{1,s1}\,\mu_{f1,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s1}\left(-\mu_{f1,s1} + z\left(2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})\right)^{-1}(\omega_{1,s1} - 1)\right)\mathrm{d}F(z)\right] + \right. \\
&\left. P_{12}\left[\int_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s2}\left(z\left(2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})\right)^{-1}(\omega_{1,s1} - 1)\right)\mathrm{d}F(z)\right]\right\}
\end{aligned} \tag{26}$$

where

$$c_{s1} = \omega_{0,s1} + \omega_{1,s1}\,\mu_{f1,s1} - \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}$$

$$c_{s2} = \omega_{0,s1} - \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}$$

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{2,s1}} = 0 : \\
&\alpha\left[P_{11}(\omega_{2,s1}-1) + P_{12}(\omega_{2,s1}-1)\right] - \\
&-(2\alpha-1)\left\{P_{11}\left[\int\limits_{\frac{\omega_{0,s1}+\omega_{1,s1}\,\mu_{f1,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s1}\left(z\left(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\right)^{-1}(\omega_{2,s1}-1)\right)\mathrm{d}F(z)\right] + \right. \\
&\qquad\qquad \left. P_{12}\left[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s2}\left(z\left(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\right)^{-1}(\omega_{2,s1}-1)\right)\mathrm{d}F(z)\right]\right\}
\end{aligned}
\qquad (27)
$$

where

$$
\begin{aligned}
c_{s1} &= \omega_{0,s1} + \omega_{1,s1}\,\mu_{f1,s1} - \sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})} \\
c_{s2} &= \omega_{0,s1} - \sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}
\end{aligned}
$$

When starting from s2, the transition probabilities in the above equations change from P(1,1) and P(1,2) to P(2,1) and P(2,2), respectively.

Appendix C Equations for special case 2

In this specification, f1 has an higher variance in state 1 (see table 4 and the setup in section 4.2 for a full specification). The expected loss takes the following form:

$$
\begin{aligned}
&E\{L(e)|I,s1\} = \alpha[P_{11}(\mu_{e,s1}^2 + \sigma_{e,s1}^2) + P_{12}(\mu_{e,s2}^2 + \sigma_{e,s2}^2)] - \\
&-(2\alpha-1)\left[P_{1,1}\int_{-\frac{\mu_{e,s1}}{\sigma_{e,s1}}}^{\infty}(\mu_{e,s1}+\sigma_{e,s1}z)^2\mathrm{d}F(z) + P_{1,2}\int_{-\frac{\mu_{e,s2}}{\sigma_{e,s2}}}^{\infty}(\mu_{e,s2}+\sigma_{e,s2}z)^2\mathrm{d}F(z)\right] = \\
&= \alpha\Big\{P_{11}[(-\omega_{0,s1})^2 + (3+(2\omega_{1,s1}^2+\omega_{2,s1}^2) - 2(2\omega_{1,s1}+\omega_{2,s1}))] + \\
&\qquad +P_{12}[(-\omega_{0,s1})^2 + (2+(\omega_{1,s1}^2+\omega_{2,s1}^2) - 2(\omega_{1,s1}+\omega_{2,s1}))]\Big\} - \\
&-(2\alpha-1)\Bigg[P_{11}\int_{\frac{\omega_{0,s1}}{\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}\Big(-\omega_{0,s1}+ \\
&+\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})z}\Big)^2\,\mathrm{d}F(z) + \\
&+P_{12}\int_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}(-\omega_{0,s1}+\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})z})^2\mathrm{d}F(z)\Bigg]
\end{aligned}
\tag{28}
$$

When the starting state is assumed to be s2, we get a similar expected loss function to equation 28, but the transition probailities P_{21} and P_{22} are used in place of P_{11} and P_{12}. When all elements of the transition probability matrix are 0.5, the two sets of weights are equal.

Minimizing the expected loss in equation 28 yields the following first order conditions:

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_0} = 0 : \\
&\alpha\left[P_{11}(\omega_{0,s1}) + P_{12}(\omega_{0,s1})\right] - \\
&-(2\alpha-1)\left\{ P_{11}\left[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} \omega_{0,s1} - \sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}\, z\, \mathrm{d}F(z)\right] + \right. \\
&\qquad\qquad \left. P_{12}\left[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} \omega_{0,s1} - \sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}\, z\, \mathrm{d}F(z)\right]\right\}
\end{aligned}
\tag{29}
$$

The optimal weights for the two individual forecasts, ω_1 and ω_2 are determined by solving first order conditions 26 and 30:

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{1,s1}} = 0 : \\
&\alpha\left[P_{11}[2\omega_{1,s1}-2] + P_{12}(\omega_{1,s1}-1)\right] - \\
&-(2\alpha-1)\left\{ P_{11}\left[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s1}\left(z\left(3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})\right)^{-1}(2\omega_{1,s1}-2)\right)\mathrm{d}F(z)\right] + \right. \\
&\qquad\qquad \left. P_{12}\left[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s2}\left(z\left(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\right)^{-1}(\omega_{1,s1}-1)\right)\mathrm{d}F(z)\right]\right\}
\end{aligned}
\tag{30}
$$

where

$$
\begin{aligned}
c_{s1} &= \omega_{0,s1} - \sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})} \\
c_{s2} &= \omega_{0,s1} - \sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{2,s1}} = 0: \\
&\alpha\left[P_{11}(\omega_{2,s1}-1) + P_{12}(\omega_{2,s1}-1)\right] - \\
&\quad -(2\alpha-1)\left\{P_{11}\left[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s1}\left(z\left(3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})\right)^{-1}(\omega_{2,s1}-1)\right)\mathrm{d}F(z)\right] + \right. \\
&\qquad\qquad \left. P_{12}\left[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s2}\left(z\left(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\right)^{-1}(\omega_{2,s1}-1)\right)\mathrm{d}F(z)\right]\right\}
\end{aligned}
\tag{31}
$$

where

$$
\begin{aligned}
c_{s1} &= \omega_{0,s1} - \sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})} \\
c_{s2} &= \omega_{0,s1} - \sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}
\end{aligned}
$$

When starting from s2, the transition probabilities in the above equations change from P(1,1) and P(1,2) to P(2,1) and P(2,2), respectively.

Appendix D Equations for special case 3

DGP:

$$y = f2 + e \tag{32}$$

forecast:

$$\hat{y} = \omega_0 + \omega_1 f1 + \omega_2 f2 \tag{33}$$

$$f1 = f + \epsilon f2 = f + \nu \tag{34}$$

$$Cov(f1, f2) = Var(f) + Cov(\epsilon, \nu) = 1 \tag{35}$$

The expected loss takes the following form:

$$\begin{aligned}
E\{L(e)|I, s1\} &= \alpha[P_{11}(\mu^2_{e,s1} + \sigma^2_{e,s1}) + P_{12}(\mu^2_{e,s2} + \sigma^2_{e,s2})] - \\
&- (2\alpha - 1)\left[P_{1,1} \int\limits_{-\frac{\mu_{e,s1}}{\sigma_{e,s1}}}^{\infty} (\mu_{e,s1} + \sigma_{e,s1} z)^2 \mathrm{d}F(z) + P_{1,2} \int\limits_{-\frac{\mu_{e,s2}}{\sigma_{e,s2}}}^{\infty} (\mu_{e,s2} + \sigma_{e,s2} z)^2 \mathrm{d}F(z)\right] = \\
&= \alpha \Big\{P_{11}[(-\omega_{0,s1})^2 + 1.2 + 2\omega^2_{1,s1} + 1.2\omega^2_{2,s1} + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1}))] + \\
&\quad + P_{12}[(-\omega_{0,s1})^2 + (2 + (\omega^2_{1,s1} + \omega^2_{2,s1}) - 2(\omega_{1,s1} + \omega_{2,s1}))]\Big\} - \\
&- (2\alpha - 1)\Bigg[P_{11} \int\limits_{\frac{\omega_{0,s1}}{\sqrt{1.2 + 2\omega^2_{1,s1} + 1.2\omega^2_{2,s1} + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1}}}}^{\infty} \Big(-\omega_{0,s1} + \\
&+ \sqrt{1.2 + 2\omega^2_{1,s1} + 1.2\omega^2_{2,s1} + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1} z}\Big)^2 \mathrm{d}F(z) + \\
&+ P_{12} \int\limits_{\frac{\omega_{0,s1}}{\sqrt{2 + (\omega^2_{1,s1} + \omega^2_{2,s1}) - 2(\omega_{1,s1} + \omega_{2,s1})}}}^{\infty} \Big(-\omega_{0,s1} + \sqrt{2 + (\omega^2_{1,s1} + \omega^2_{2,s1}) - 2(\omega_{1,s1} + \omega_{2,s1})z}\Big)^2 \mathrm{d}F(z)]\Bigg]
\end{aligned} \tag{36}$$

When the starting state is assumed to be s2, we get a similar expected loss function to equation 36, but the transition probailities P_{21} and P_{22} are used in place of P_{11} and P_{12}. When all elements of the transition probability matrix are 0.5, the two sets of weights are equal.

Minimizing the expected loss in equation 36 yields the following first order conditions:

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_0} = 0 : \\
&\alpha \left[P_{11}(\omega_{0,s1}) + P_{12}(\omega_{0,s1})\right] - \\
&\qquad - (2\alpha - 1) \left\{ P_{11} \left[\int_{\frac{\omega_{0,s1}}{\sqrt{(1.2+2\omega_{1,s1}^2+1.2\omega_{2,s1}^2+2\omega_{1,s1}\omega_{2,s1}-2(\omega_{1,s1}+1.2\omega_{2,s1})}}}^{\infty} \omega_{0,s1} - \right.\right. \\
&\qquad \left. - \sqrt{1.2 + 2\omega_{1,s1}^2 + 1.2\omega_{2,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1})}\; z \,\mathrm{d}F(z) \right] + \\
&\qquad \left. P_{12} \left[\int_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} \omega_{0,s1} - \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}\; z \,\mathrm{d}F(z) \right] \right\}
\end{aligned}
\tag{37}
$$

The optimal weights for the two individual forecasts, ω_1 and ω_2 are determined by solving first order conditions 38 and 39:

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{1,s1}} = 0 : \\
&\alpha \left[P_{11}(2\omega_{1,s1} + \omega_{2,s1} - 1) + P_{12}(\omega_{1,s1} - 1)\right] - \\
&- (2\alpha - 1) \left\{ P_{11} \left[\int_{\frac{\omega_{0,s1}}{\sqrt{1.2+2\omega_{1,s1}^2+1.2\omega_{2,s1}^2+2\omega_{1,s1}\omega_{2,s1}-2(\omega_{1,s1}+1.2\omega_{2,s1})}}}^{\infty} c_{s1} \Big(z \Big(1.2 + 2\omega_{1,s1}^2 + 1.2\omega_{2,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1})\Big)^{-1} \times \right.\right. \\
&\qquad\qquad \left. \times (2\omega_{1,s1} + \omega_{2,s1} - 1)\Big) \,\mathrm{d}F(z) \right] + \\
&\qquad\qquad \left. P_{12} \left[\int_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s2} \Big(z \Big(2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})\Big)^{-1} (\omega_{1,s1} - 1) \Big) \,\mathrm{d}F(z) \right] \right\}
\end{aligned}
\tag{38}
$$

where

$$
\begin{aligned}
c_{s1} &= \omega_{0,s1} - \sqrt{1.2 + 2\omega_{1,s1}^2 + 1.2\omega_{2,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1})} \\
c_{s2} &= \omega_{0,s1} - \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{2,s1}} = 0 : \\
&\alpha \left[P_{11}(\omega_{1,s1} + 1.2\omega_{2,s1} - 1.2) + P_{12}(\omega_{2,s1} - 1)\right] - \\
&- (2\alpha - 1)\left\{ P_{11}\left[\int_{\frac{\omega_{0,s1}}{\sqrt{1.2+2\omega_{1,s1}^2+1.2\omega_{2,s1}^2+2\omega_{1,s1}\omega_{2,s1}-2(\omega_{1,s1}+1.2\omega_{2,s1})}}}^{\infty} c_{s1}\left(z\left(1.2 + 2\omega_{1,s1}^2 + 1.2\omega_{2,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1})\right)^{-1} \times \right.\right.\right. \\
&\qquad\qquad \left.\left. \times(\omega_{1,s1} + 1.2\omega_{2,s1} - 1.2)\right)\mathrm{d}F(z)\right] + \\
&\qquad \left.P_{12}\left[\int_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s2}\left(z\left(2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})\right)^{-1}(\omega_{2,s1} - 1)\right)\mathrm{d}F(z)\right]\right\}
\end{aligned}
\tag{39}
$$

where

$$
\begin{aligned}
c_{s1} &= \omega_{0,s1} - \sqrt{1.2 + 2\omega_{1,s1}^2 + 1.2\omega_{2,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1})} \\
c_{s2} &= \omega_{0,s1} - \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}
\end{aligned}
$$

When starting from s2, the transition probabilities in the above equations change from P(1,1) and P(1,2) to P(2,1) and P(2,2), respectively.

Appendix E Equations for special case 4

DGP:

$$y = f1 + f2 + f3 + e \tag{40}$$

forecast:

$$\hat{y} = \omega_0 + \omega_1 f1 + \omega_2 f2 \tag{41}$$

$$\begin{aligned} f1 &= f3 + \epsilon \\ f2 &= f3 + \nu \end{aligned} \tag{42}$$

$$Cov(f1, f2) = Var(f3) + Cov(\epsilon, \nu) = 1 \tag{43}$$

$$\begin{aligned} Cov(y, f1) &= Var(f1) + Cov(f1, f2) + Cov(f1, f3) = 1.1 + 1 + 1 = 3.1 \\ Cov(y, f2) &= Var(f1) + Cov(f1, f2) + Cov(f2, f3) = 10 + 1 + 1 = 12 \end{aligned} \tag{44}$$

The expected loss takes the following form:

$$\begin{aligned} E\{L(e)|I, s1\} = {} & \alpha[P_{11}(\mu_{e,s1}^2 + \sigma_{e,s1}^2) + P_{12}(\mu_{e,s2}^2 + \sigma_{e,s2}^2)] - \\ & - (2\alpha - 1)\left[P_{1,1} \int_{-\frac{\mu_{e,s1}}{\sigma_{e,s1}}}^{\infty} (\mu_{e,s1} + \sigma_{e,s1} z)^2 \mathrm{d}F(z) + P_{1,2} \int_{-\frac{\mu_{e,s2}}{\sigma_{e,s2}}}^{\infty} (\mu_{e,s2} + \sigma_{e,s2} z)^2 \mathrm{d}F(z)\right] = \end{aligned}$$

$$\begin{aligned} = {} & \alpha\left\{P_{11}[(-\omega_{0,s1})^2 + (16.1 + (1.1\omega_{1,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} + 10\omega_{2,s1}^2) - 2(3.1\omega_{1,s1} + 12\omega_{2,s1}))] + \right. \\ & \left. + P_{12}[(-\omega_{0,s1})^2 + (2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1}))]\right\} - \\ & - (2\alpha - 1)\left[P_{11} \int_{\frac{\omega_{0,s1}}{\sqrt{16.1 + (1.1\omega_{1,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} + 10\omega_{2,s1}^2) - 2(3.1\omega_{1,s1} + 12\omega_{2,s1})}}}^{\infty} \left(-\omega_{0,s1} + \right.\right. \\ & \left. + \sqrt{16.1 + (1.1\omega_{1,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} + 10\omega_{2,s1}^2) - 2(3.1\omega_{1,s1} + 12\omega_{2,s1})}\, z\right)^2 \mathrm{d}F(z) + \\ & \left. + P_{12} \int_{\frac{\omega_{0,s1}}{\sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}}}^{\infty} \left(-\omega_{0,s1} + \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}\, z\right)^2 \mathrm{d}F(z)]\right] \end{aligned}$$

(45)

When the starting state is assumed to be s2, we get a similar expected loss function to equation 45, but the transition probailities P_{21} and P_{22} are used in place of P_{11} and P_{12}. When all elements of the transition probability matrix are 0.5, the two sets of weights are equal.

Minimizing the expected loss in equation 45 yields the following first order conditions:

$$\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_0} = 0 : \\
&\alpha\left[P_{11}(\omega_{0,s1}) + P_{12}(\omega_{0,s1})\right] - \\
&-(2\alpha-1)\left\{P_{11}\left[\int_{\frac{\omega_{0,s1}}{\sqrt{(16.1+(1.1\omega_{1,s1}^2+2\omega_{1,s1}\omega_{2,s1}+10\omega_{2,s1}^2)-2(3.1\omega_{1,s1}+12\omega_{2,s1}))}}}^{\infty} \omega_{0,s1} - \right.\right. \\
&\left. -\sqrt{16.1 + (1.1\omega_{1,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} + 10\omega_{2,s1}^2) - 2(3.1\omega_{1,s1} + 12\omega_{2,s1})}\, z\, \mathrm{d}F(z)\right] + \\
&\left.P_{12}\left[\int_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} \omega_{0,s1} - \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}\, z\, \mathrm{d}F(z)\right]\right\}
\end{aligned} \tag{46}$$

The optimal weights for the two individual forecasts, ω_1 and ω_2 are determined by solving first order conditions 47 and 48:

$$\begin{aligned}
&\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{1,s1}} = 0 : \\
&\alpha\left[P_{11}(1.1\omega_{1,s1} + \omega_{2,s1} - 3.1) + P_{12}(\omega_{1,s1} - 1)\right] - \\
&-(2\alpha-1)\left\{P_{11}\left[\int_{\frac{\omega_{0,s1}}{\sqrt{16.1+(1.1\omega_{1,s1}^2+2\omega_{1,s1}\omega_{2,s1}+10\omega_{2,s1}^2)-2(3.1\omega_{1,s1}+12\omega_{2,s1})}}}^{\infty} c_{s1}\left(z\left(16.1 + (1.1\omega_{1,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} + 10\omega_{2,s1}^2) - \right.\right.\right.\right. \\
&\qquad \left.\left.\left. -2(3.1\omega_{1,s1} + 12\omega_{2,s1})\right)^{-1}{}_{(1.1\omega_{1,s1}+\omega_{2,s1}-3.1)}\right) \mathrm{d}F(z)\right] + \\
&\left.P_{12}\left[\int_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s2}\left(z\left(2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})\right)^{-1}(\omega_{1,s1} - 1)\right) \mathrm{d}F(z)\right]\right\}
\end{aligned} \tag{47}$$

where

$$c_{s1} = \omega_{0,s1} - \sqrt{16.1 + (1.1\omega_{1,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} + 10\omega_{2,s1}^2) - 2(3.1\omega_{1,s1} + 12\omega_{2,s1})}$$

$$c_{s2} = \omega_{0,s1} - \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}$$

$$
\begin{aligned}
&\frac{\partial E\{L(e)|s,l\}}{\partial \omega_{2,s1}} = 0: \\
&\alpha \left[P_{11}(\omega_{1,s1} + 10\omega_{2,s1} - 12) + P_{12}(\omega_{2,s1} - 1)\right] - \\
&-(2\alpha - 1)\left\{P_{11}\left[\frac{\int_{\omega_{0,s1}}^{\infty}}{\sqrt{16.1+(1.1\omega_{1,s1}^2+2\omega_{1,s1}\omega_{2,s1}+10\omega_{2,s1}^2)-2(3.1\omega_{1,s1}+12\omega_{2,s1})}} c_{s1}\left(z\left(16.1 + (1.1\omega_{1,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} + 10\omega_{2,s1}^2) - \right.\right.\right.\right. \\
&\qquad\qquad \left.\left.\left. -2(3.1\omega_{1,s1} + 12\omega_{2,s1})\right)^{-1}{}_{(\omega_{1,s1}+10\omega_{2,s1}-12)}\right) \mathrm{d}F(z)\right] + \\
&\qquad \left.P_{12}\left[\frac{\int_{\omega_{0,s1}}^{\infty}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}} c_{s2}\left(z\left(2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})\right)^{-1} (\omega_{2,s1} - 1)\right) \mathrm{d}F(z)\right]\right\}
\end{aligned}
\tag{48}
$$

where

$$
\begin{aligned}
c_{s1} &= \omega_{0,s1} - \sqrt{16.1 + (1.1\omega_{1,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} + 10\omega_{2,s1}^2) - 2(3.1\omega_{1,s1} + 12\omega_{2,s1})} \\
c_{s2} &= \omega_{0,s1} - \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}
\end{aligned}
$$

When starting from s2, the transition probabilities in the above equations change from P(1,1) and P(1,2) to P(2,1) and P(2,2), respectively.

Appendix F Optimal weights for case 4: full tables

Table 15
Optimal weights from case 4, only one state (s1)

	optimal weights		
α	starting state: s1		
	ω_{0t}	ω_{1t}	ω_{2t}
0.1	0.083	0.517	1.171
0.3	0.000	1.900	1.010
0.5	0.000	1.900	1.010
0.7	0.000	1.900	1.010
0.9	0.113	0.444	1.184

Table 16
Optimal weights from case 4, symmetric transition probabilities

	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.1	0.066	0.517	1.146	0.105	0.518	1.157	0.000	1.452	1.050	0.000	1.452	1.050
0.3	-0.001	1.452	1.050	-0.001	1.452	1.050	0.000	1.452	1.050	0.000	1.452	1.050
0.5	0.000	1.452	1.050	0.000	1.452	1.050	0.000	1.452	1.050	0.000	1.452	1.050
0.7	0.089	0.490	1.162	0.090	0.481	1.167	0.000	1.452	1.050	0.000	1.452	1.050
0.9	0.008	1.449	1.050	0.098	0.479	1.176	0.000	1.452	1.050	0.000	1.452	1.050

Table 17
Optimal weights from case 4, asymmetric transition probabilities

	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.1	0.058	0.517	1.167	-0.001	1.094	1.048	0.000	1.810	1.019	0.000	1.094	1.048
0.3	0.001	1.810	1.019	-0.001	1.094	1.048	0.000	1.810	1.019	0.000	1.094	1.048
0.5	0.000	1.810	1.019	0.000	1.094	1.048	0.000	1.810	1.019	0.000	1.094	1.048
0.7	0.001	1.810	1.019	0.000	1.094	1.048	0.000	1.810	1.019	0.000	1.094	1.048
0.9	0.000	1.810	1.019	-0.002	1.095	1.048	0.000	1.810	1.019	0.000	1.094	1.048

MNB Working Papers 2023/3

Optimal forecast combination under asymmetric loss and regime-switching

Budapest, December 2023

mnb.hu

1013 BUDAPEST, KRISZTINA KÖRÚT 55.