Essays on Economic Forecasting Under General Loss Functions

by
Viola Monostoriné Grolmusz

Submitted to
Central European University
Department of Economics

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Supervisor: Róbert Lieli

Budapest, Hungary
2023

10.14754/CEU.2023.02

CENTRAL EUROPEAN UNIVERSITY
DEPARTMENT OF ECONOMICS AND BUSINESS

Author: **Viola Monostoriné Grolmusz**
Title: **Essays on Economic Forecasting Under General Loss Functions**
Degree: **Ph.D.**
Dated: **June 7, 2023**

Hereby I testify that this thesis contains no material accepted for any other degree in any other institution and that it contains no material previously written and/or published by another person except where appropriate acknowledgement is made.

Monostoriné Grolmusz Viola

Signature of the author

10.14754/CEU.2023.02

DISCLOSURE OF CO-AUTHORS CONTRIBUTION

Chapter 2: Problems in Identifying Loss Functions

Co-authors: Róbert Lieli and Maxwell Stinchcombe
This chapter is an extended version of section 5.2 in our co-authored paper, Lieli et al. (2019). The original paper was developed in close cooperation with Róbert Lieli and Maxwell Stinchcombe. Róbert and Maxwell contributed to the main idea and the theoretical arguments of the paper. I contributed to the detailed empirical examples by data collection, programming, analysis and writing.

10.14754/CEU.2023.02

Abstract

This thesis consists of three chapters on economic forecasting under general loss functions. The first two chapters make contributions to the theory of econometric forecasting, while the third one presents new evidence that stock analyst forecasts are biased. In the first chapter, I use a regime switching framework and assume asymmetric loss functions when deriving optimal forecast combination weights. In the second chapter, co-authored with Róbert Lieli and Maxwell Stinchcombe, we explore a serious identification problem in the estimation of the asymmetry parameter in the seminal 2005 paper of Elliott, Komunjer and Timmermann. In the third chapter, I estimate an asymmetry parameter capturing stock analyst's relative cost from overpredicting versus underpredicting the stock performance and present new evidence on analysts' bias.

Chapter 1: Optimal forecast combination under asymmetric loss and regime-switching

Forecast combinations have been repeatedly shown to outperform individual professional forecasts and complicated time series models in accuracy. 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.

10.14754/CEU.2023.02

Chapter 2: Problems in Identifying Loss Functions

(joint with Róbert Lieli and Maxwell Stinchcombe)

The seminal paper by Elliott, Komunjer and Timmermann (2005; henceforth EKT) proposes a method for estimating a forecaster's loss function based on a moment condition derived from the first order condition of the forecaster's expected loss minimization problem. This chapter demonstrates that the moment condition used for identification is fundamentally non-unique; that is, a very diverse class of loss functions can give rise to the same moment condition. More specifically, if one estimates the asymmetry parameter of a lin-lin or quad-quad loss function, there exist other loss functions with completely different asymmetry properties that are observationally equivalent. In EKT this serious identification problem remains hidden by the fact that they only consider loss functions that depend purely on the forecast error and not separately on the level of the forecast or the realization. However, once such level effects are allowed, identification completely breaks down. Hence, it is critical for any practical application of EKT to provide theoretical justification for why the assumption of a purely error-dependent loss function is appropriate in the situation at hand. Nevertheless, this discussion is usually missing. A version of this chapter has already been published as Section 5.2 of the paper by Lieli, Stinchcombe and Grolmusz (2019, International Journal of Forecasting), which deals with several technical issues in identifying loss functions.

Chapter 3: Recovering Stock Analysts' Loss Functions from Buy/Sell Recommendations

I carry out an empirical analysis to recover stock analysts' loss functions from observations on forecasts, actual realizations and a proxy for the publicly observed part of the analyst's information set. The forecasts I use are analyst stock (buy/hold/sell) recommendations for two Blue Chip stocks. I estimate an asymmetry parameter that captures the analyst's relative cost from overpredicting versus underpredicting the stock performance. I find that the results are sensitive to the categorization of 'hold' recommendations. When substituting 'holds' with the recommendation from the previous period, in most cases the estimated bounds for the asymmetry parameter suggest that analysts are more likely to issue a 'false buy' than a 'false sell' recommendation. This is in line with the frequent statement from the analyst recommendations literature, that optimism relative to the consensus is rewarded in analyst recommendations.

10.14754/CEU.2023.02

Acknowledgments

I am highly indebted to my supervisor, Róbert Lieli for his constant support and invaluable guidance during my PhD work. He spared no time to discuss my work regularly, and always motivated me by showing interest in my results. I learned a lot from him on how to find interesting research problems and how to see the economic intuition in empirical results.

I am also very grateful for faculty members at CEU whom I could always turn to with my questions and who gave valuable comments and suggestions on my papers: Sergey Lychagin, László Mátyás, Attila Rátfai, Ádám Reiff, Ádám Zawadowski, Marc Kauffman, Arieda Muco, Andrea Weber. I am also indebted to fellow PhD students and colleagues: Lajos Szabó, Judit Krekó, Péter Gábriel, Anna Naszódi. I also thank the staff members for their continuous and kind help.

Last but not least, I am thankful to my family and friends for believing in me. I am especially grateful to my husband for always supporting me in pursuing my dreams and to my children for their patience, understanding and love throughout the long years of my PhD studies.

10.14754/CEU.2023.02

Table of Contents

10.14754/CEU.2023.02

10.14754/CEU.2023.02

List of Tables

10.14754/CEU.2023.02

10.14754/CEU.2023.02

List of Figures

10.14754/CEU.2023.02

10.14754/CEU.2023.02

Chapter 1

Optimal forecast combination under asymmetric loss and regime-switching

1.1. 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.

[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.

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.

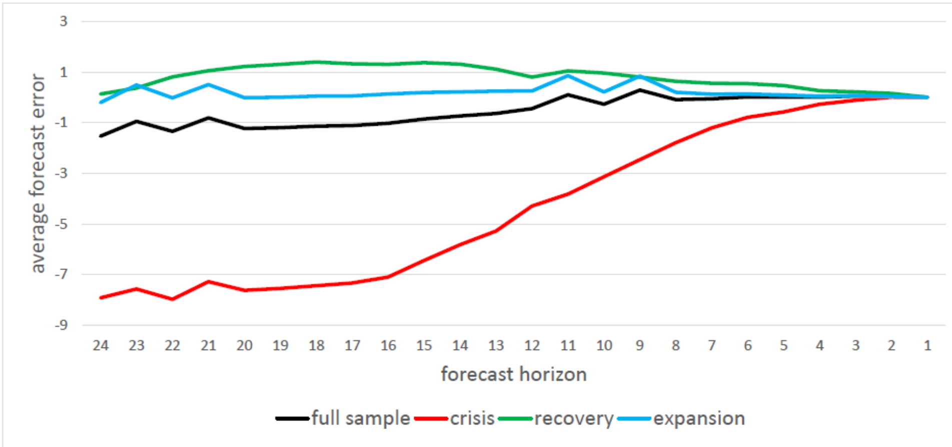
As an illustration, I have reproduced Figure 1. from the paper of Dovern and Janssen (2017) using a different data set. I have used Consensus Economics surveys for annual real GDP growth for 11 Easters European countries[3], for the period between 2007 and 2019. The forecast horizons used range from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

[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 1.1: 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).

pansions are on average unbiased (Figure 1.1). 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 chapter is organized as follows. Section 1.2 shows the theoretical setup and outlines the expected loss minimization problem in the general case. Section 1.3 describes the procedure used for deriving the optimal weights numerically. Section 2.4 analyzes how the optimal bias and the combination weights depend on the parameters through four analytical examples with different parametrizations. Section 1.5 assembles general observations from the results that could be formalized as theorems and also outlines some possible extensions. The last section concludes.

1.2. 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.

1.2.1 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

$$\hat{\boldsymbol{y}}_{\tau+1} = (\hat{y}_{1\tau+1}, \dots \hat{y}_{m\tau+1})' \tag{1.1}$$

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:

$$y_{t+1} = \omega_0 + \boldsymbol{\omega}'\hat{\boldsymbol{y}}_{t+1} + e_{t+1}, \tag{1.2}$$

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:

$$\begin{pmatrix} y_{t+1} \\ \hat{\boldsymbol{y}}_{\boldsymbol{t+1}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{ys_{t+1}} \\ \boldsymbol{\mu}_{\hat{\boldsymbol{y}}s_{t+1}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{ys_{t+1}} & \boldsymbol{\sigma}'_{y\hat{\boldsymbol{y}}s_{t+1}} \\ \boldsymbol{\sigma}_{y\hat{\boldsymbol{y}}s_{t+1}} & \boldsymbol{\Sigma}_{\hat{\boldsymbol{y}}s_{t+1}} \end{pmatrix} \right) \tag{1.3}$$

Given equation (1.2), assumption (1.3) 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:

$$\begin{pmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1k} \\ \pi_{21} & \pi_{22} & \dots & \pi_{2k} \\ \vdots & \vdots & \ddots & \pi_{k-1k} \\ \pi_{k1} & \dots & \pi_{kk-1} & \pi_{kk} \end{pmatrix} \tag{1.4}$$

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 (1.4) that corresponds to row s_t and column s_{t+1}.

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:

$$L(e) = \begin{cases} (1-\alpha)e^2, & \text{if } e > 0 \\ \alpha e^2, & \text{if } e \leq 0 \end{cases} \tag{1.5}$$

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.

[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.

Assuming the loss function takes the form expressed in equation 1.5, the posited objective is to minimize the following expected loss formula:

$$E\{L(e_{t+1})|I_t, s_t\} = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} E\Big\{[\alpha - (2\alpha - 1)\mathbb{1}_{e_{s_{t+1}}>0}]e^2_{s_{t+1}}\Big|I_t, s_{t+1}\Big\}, \tag{1.6}$$

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 1.6. 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 (1.6) 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 state s_t. However, s_t is not directly observed by the econometrician, which means that the evaluation of (1.6) 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 (1.2) in a way that minimizes her expected loss (1.6). To

[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 1.6 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.

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 (1.2) and assumption (1.3), these moments are given by

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

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

Substituting $e_{s_{t+1}} = \mu_{e_{s_{t+1}}} + \sigma_{e_{s_{t+1}}} z_{s_{t+1}}$ into (1.6) and making the corresponding change of variables in the integral yields the following expression:[7]

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

$$= \alpha \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} [\mu_e^2 + \sigma_e^2] - (2\alpha - 1) \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} \int_{-\frac{\mu_e}{\sigma_e}}^{\infty} (\mu_e + \sigma_e z_{s_{t+1}})^2 \mathrm{d}F(z_{s_{t+1}}), \tag{1.7}$$

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 (1.7) 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 (1.6) 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.

1.2.3 Expected loss minimization in the general case

Let us minimize the expected loss function in the general case (1.7) by deriving the corresponding first order conditions (FOCs).

[7] See the detailed derivations in appendix 1, equation 3.2.

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

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

$$\alpha \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} \left(\mu_e\right) - (2\alpha - 1) \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} \left[\int_{-\frac{\mu_e}{\sigma_e}}^{\infty} (\mu_e + \sigma_e \, z_{s_{t+1}}) \mathrm{d}F(z_{s_{t+1}}) \right] = 0 \tag{1.8}$$

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:

$$\frac{\partial E\{L(e)|s, I\}}{\partial\omega_{0t}} = 0 :$$

$$\alpha \sum_{s} \pi_s \left(\mu_{y_s} - \omega_{0t} - \boldsymbol{\omega}_t'\boldsymbol{\mu}_{\hat{\boldsymbol{y}}_s}\right) -$$

$$(2\alpha - 1) \sum_{s} \pi_s \left[\int_{\frac{-\mu_y + \omega_0 + \boldsymbol{\omega}_t'\boldsymbol{\mu}_{\hat{\boldsymbol{y}}_s}}{\sqrt{\sigma_y^2 + \boldsymbol{\omega}_t'\boldsymbol{\Sigma}_{\hat{\boldsymbol{y}}}\boldsymbol{\omega}_t - 2\boldsymbol{\omega}_t'\boldsymbol{\sigma}_{y\hat{\boldsymbol{y}}}}}}^{\infty} (\mu_{y_s} - \omega_{0t} - \boldsymbol{\omega}_t'\boldsymbol{\mu}_{\hat{\boldsymbol{y}}_s} + (\sigma_y^2 + \boldsymbol{\omega}_t'\boldsymbol{\Sigma}_{\hat{\boldsymbol{y}}}\boldsymbol{\omega}_t - 2\boldsymbol{\omega}_t'\boldsymbol{\sigma}_{y\hat{\boldsymbol{y}}}) \, z) \mathrm{d}F(z) \right] = 0 \tag{1.9}$$

The optimal weights ω_{0t}^* and $\boldsymbol{\omega}_t^*$ must then satisfy equation (1.9).

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:

$$\frac{\partial E\{L(e)|s, I\}}{\partial\boldsymbol{\omega}} = 0 :$$

$$\begin{aligned} &\alpha \sum_{s} \pi_s \left(-\mu_{\hat{\boldsymbol{y}}}\mu_e + \boldsymbol{\Sigma}_{\hat{\boldsymbol{y}}}\boldsymbol{\omega} - \boldsymbol{\sigma}_{y\hat{\boldsymbol{y}}}\right) - \\ - (2\alpha - 1) &\sum_{s} \pi_s \left[\int_{-\frac{\mu_e}{\sigma_e}}^{\infty} (\mu_e + \sigma_e \, z) \left(-\mu_{\hat{\boldsymbol{y}}} + \frac{1}{\sigma_e} \left(\boldsymbol{\Sigma}_{\hat{\boldsymbol{y}}}\boldsymbol{\omega} - \boldsymbol{\sigma}_{y\hat{\boldsymbol{y}}} \right) \, z \right) \mathrm{d}F(z) \right] = 0 \end{aligned} \tag{1.10}$$

The optimal weights ω_{0t}^* and $\boldsymbol{\omega}_t^*$ must also satisfy equation (1.10).

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.

10.14754/CEU.2023.02

1.3. Numerical procedure for computing the weights

Suppose that all the parameters in equations (1.9) and (1.10) 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

$$\int_a^b g(z;\boldsymbol{\theta})dF(z), \tag{1.11}$$

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,

$$\int g(z;\boldsymbol{\theta})\mathbb{1}_{[a,b]}(z)dF(z) = E\{g(Z;\boldsymbol{\theta})\mathbb{1}_{[a,b]}(Z)\}, \qquad Z \sim N(0,1). \tag{1.12}$$

Using this representation of the integrals with respect to $dF(z)$, the first order conditions (1.9) and (1.10) can be thought of as a set of moment conditions

$$E[m_j(Z;\boldsymbol{\theta})] = 0, \qquad j = 0,\ldots,m, \tag{1.13}$$

where, for example,

$$\begin{aligned} m_0(Z,\boldsymbol{\theta}) &= \alpha\sum_s \pi_s\left(\mu_{y_s} - \omega_0 - \boldsymbol{\omega'\mu_{\hat{y}_s}}\right) \\ &\quad - (2\alpha-1)\sum_s \pi_s\Big\{\left[\mu_{y_s} - \omega_0 - \boldsymbol{\omega'\mu_{\hat{y}_s}} + (\sigma^2_{y_s} + \boldsymbol{\omega'\Sigma_{\hat{y}_s}\omega} - 2\boldsymbol{\omega'\sigma}_{y\boldsymbol{\hat{y}}_s})\,Z\right]\cdot\mathbb{1}_{[a_s,\infty)}(Z)\Big\} \end{aligned} \tag{1.14}$$

with

$$a_s = \frac{-\mu_{y_s} + \omega_0 + \boldsymbol{\omega'\mu_{\hat{y}_s}}}{\sqrt{\sigma^2_{y_s} + \boldsymbol{\omega'\Sigma_{\hat{y}_s}\omega} - 2\boldsymbol{\omega'\sigma}_{y\boldsymbol{\hat{y}}_s}}}. \tag{1.15}$$

Equations (1.9) and (1.10) can be written this way because the linearity of the expectations

10.14754/CEU.2023.02

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

$$\frac{1}{n}\sum_{i=1}^{n} m_j(Z_i, \boldsymbol{\theta}) = 0, \qquad j = 0, \dots, m. \tag{1.16}$$

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.

1.4. 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).

1.4.1 Scenario 1: one biased forecast

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

$$y_{t+1} = \beta_1^{s_t} x_t + \beta_2^{s_t} w_t + \epsilon_{t+1} = f1_t + f2_t + \epsilon_{t+1} \tag{1.17}$$

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:

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

Table 1.1: Scenario 1: one biased forecast

$$\begin{aligned} f1_t &= \beta_1^{s_t} x_t \\ f2_t &= \beta_2^{s_t} w_t \end{aligned} \tag{1.18}$$

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

$$\hat{y}_{t+1|t} = \omega_{0t} + \omega_{1t} f1_t + \omega_{2t} f2_t \tag{1.19}$$

In the first parametrization, asymmetry is introduced by a small positive bias of forecast 1 in state 1 (see table 1.1 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 2). The variance-covariance matrix of the two forecasts and the covariances between the target and the individual forecasts take the following forms:

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

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 2. 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:

$$\omega_0 = -(\text{forecast bias} \times Pr(\text{arriving in biased state})) \tag{1.20}$$

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 1.2, the transition probability matrix P_1 results in a symmetric loss function

	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 1.2: Optimal weights from case 1, symmetric transition probabilities

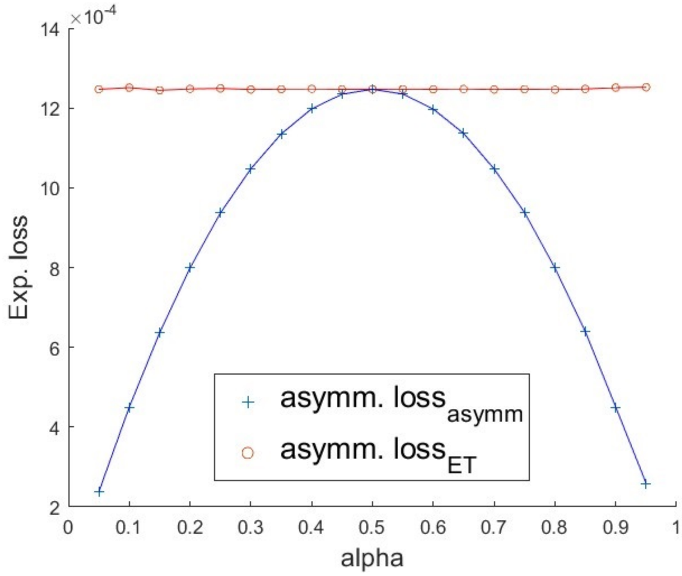
	optimal weights						ET optimal weights					
α	starting state: s1			starting state: s2			starting state: s1			starting state: s2		
	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}	ω_{0t}	ω_{1t}	ω_{2t}
0.1	-0.050	0.998	1.000	-0.001	1.000	1.000	-0.090	0.999	1.000	-0.010	0.999	1.000
0.3	-0.079	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

Table 1.3: Optimal weights from case 1, asymmetric transition probabilities

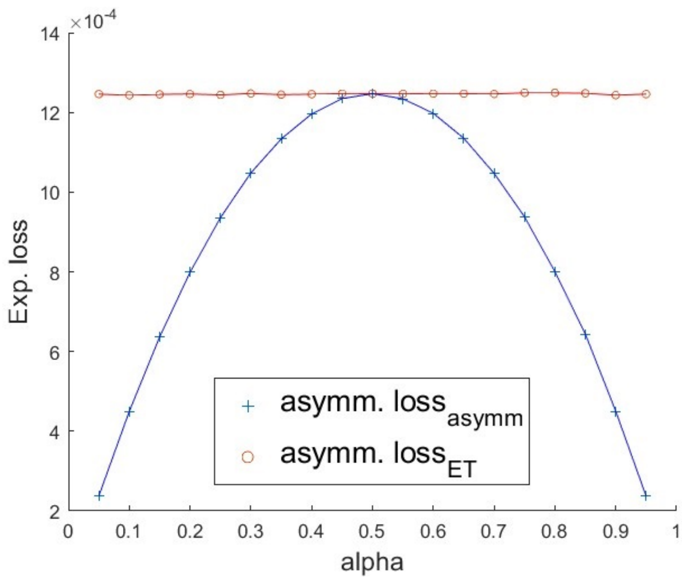
that is always lower than the constant loss resulting from the ET optimal combination. The loss is lower for more extreme asymmetry preferences (α-s close to 0 and 1). Again, the starting state does not influence the results.

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 1.20 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.

Figure 1.2: 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]



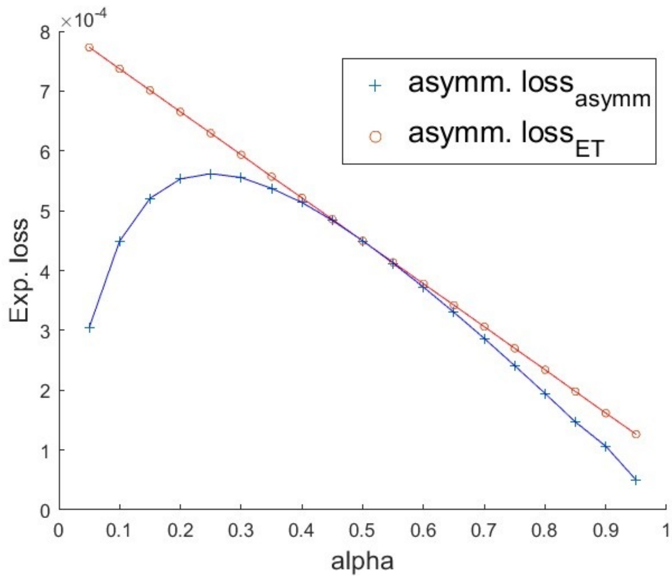
quad-quad losses, starting state: s1



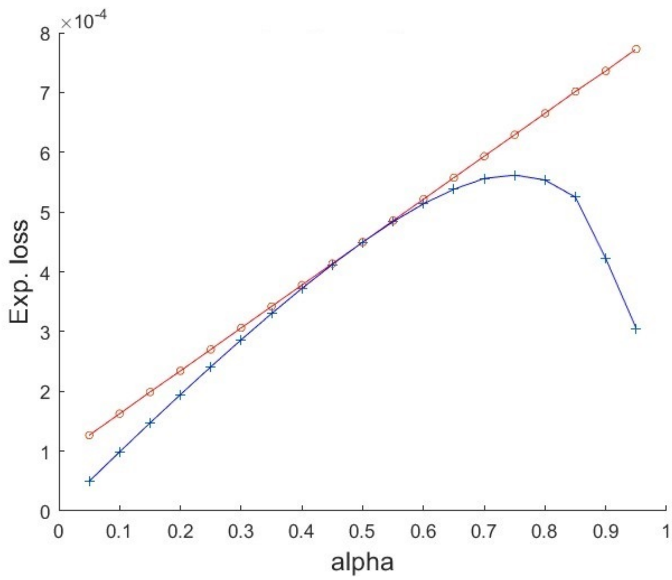
quad-quad losses, starting state: s2

10.14754/CEU.2023.02

Figure 1.3: 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

1.4.2 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 3.

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

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 1.4: Scenario 2: one forecast has higher variance in state 1

α	optimal weights						ET optimal weights					
	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 1.5: Optimal weights from case 2, symmetric transition probabilities

α	optimal weights						ET optimal weights					
	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 1.6: Optimal weights from case 2, asymmetric transition probabilities

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.

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.

10.14754/CEU.2023.02

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

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.

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

s_1		s_2	
β^{s1}_1	1	β^{s2}_1	1
β^{s1}_2	1	β^{s2}_2	1
μ^{s1}_x	0	μ^{s2}_x	0
σ^{s1}_x	$\sqrt{2}$	σ^{s2}_x	1
μ^{s1}_w	0	μ^{s2}_w	0
σ^{s1}_w	1	σ^{s2}_w	1
μ^{s1}_y	0	μ^{s2}_y	0
σ^{s1}_y	$\sqrt{3}$	σ^{s2}_y	$\sqrt{2}$
$Cov(x,w)^{s1}$	0	$Cov(x,w)^{s2}$	0

Table 1.7: Scenario 3: correlated forecasts

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

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

α	optimal weights		
	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 1.8: Optimal weights from case 3, only one state (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 1.8, 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 1.9. 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 1.10 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 ($P_{11} = 0.9$).

	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 1.9: 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	0.001	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 1.10: Optimal weights from case 3, asymmetric transition probabilities

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.

1.4.4 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):

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

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, $\upsilon_{s1} \sim N(0, 1)$.

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 1.11: Scenario 4: common factor

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

$$f1_t^{s1} = f3_t + \zeta_t$$
$$f2_t^{s1} = f3_t + \eta_t$$

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

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.

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

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

α	optimal weights		
	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 1.12: Optimal weights from case 4, only one state (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 1.12 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 1.5). 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.

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 1.13[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 1.13 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 1.14 shows the optimal weights assuming asymmetric transition probabilities

[8]Results for α-s lower than 0.3 and higher than 0.7 are truncated from table 1.12, since the numerical procedure produced large estimation errors. The full table can be found in appendix 6.

[9]Results for α-s lower than 0.3 are truncated from table 1.13 due to estimation errors at these extreme values. The full table can be found in appendix 6. At $\alpha = 0.7$ the outlier values are also likely due to estimation error.

α	optimal weights						ET optimal weights					
	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 1.13: Optimal weights from case 4, symmetric transition probabilities

α	optimal weights						ET optimal weights					
	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

Table 1.14: Optimal weights from case 4, 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 1.12. 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.

1.5. Conjectures

In this section I assemble general observations from section 2.4 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

[10] Again, the results for $\alpha = 0.1$ are truncated due to estimation errors, see the full table in appendix 6.

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.

1.6. 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 1.5. The most important conjecture is that assuming an asymmetric quadratic loss function and regime switching, the optimal combination weights depend on the asymmetry parameter

10.14754/CEU.2023.02

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.

10.14754/CEU.2023.02

Chapter 2

Problems in Identifying Loss Functions[1]

2.1. Introduction

A biased forecast (a systematic underprediction or overprediction of the target) is often interpreted as evidence of the forecaster's irrationality (Mishkin (1981), Zarnowitz (1985), Davies and Lahiri (1995), de Mendonca et al. (2021)). However, the property that the optimal forecast should be unbiased is based on a very specific assumption — that the forecaster's loss function is quadratic. Under more general, possibly asymmetric, loss functions the optimal forecast involves a bias term (Granger 1969, Patton and Timmerman 2007, Capistrán and Timmermann (2009), Franses (2021)). In general, there is no reason to assume that economic forecasts are made under conditions that are well captured by a symmetric loss function such as quadratic loss (Granger 1969, Granger and Newbold (1986), Christoffersen and Diebold (1997), Döpke et al. (2010)). For example, a central bank may face very different costs if it underpredicts inflation or overpredicts it. Thus, when seeing biased forecasts in the data, an alternative interpretation is that it was produced by a rational forecaster possessing an asymmetric loss function.

Starting from this last observation, Elliott, Komunjer and Timmermann (2005) propose a method for estimating a forecaster's loss function using observations on the forecasts, the corresponding realizations, and possibly some of the predictors in the forecaster's information set. They model the candidate loss functions as a parametric class, which includes lin-lin and quad-quad as special cases. Importantly, all members of this class depend on the forecast error only, i.e., the difference between the realization and the forecast. The proposed estimation procedure is based on the first order condition for the forecaster's expected loss minimization problem. For example, if the forecaster has quadratic loss, the first order condition states that the forecast error is uncorrelated with all variables in the forecaster's information set. For general loss functions the analogous condition is that the first derivative of the loss function with respect to the forecast, evaluated at the observed forecast-realization pair, is uncorrelated with the predictors. These moment conditions can then be used to derive a standard GMM estimator for the unknown loss function parameters.

The EKT method has been applied in a number of contexts to recover forecasters' preferences. For example, Capistrán (2008) uses the EKT estimator to estimate the loss function of the Federal Reserve before and during the Volcker era. Capistrán finds that under the chairmanship of Volcker, having inflation above the target was costlier than having inflation below the target. Before Volcker, the loss suggests an asymmetry of the opposite direction. Döpke et al. apply the EKT method to German annual inflation and output growth forecasts covering the 1970-2007 period. The authors find limited evidence of asymmetric loss for the economic forecasts analyzed; only pooled forecasts proved to be asymmetric, but not individual ones. When quad-quad functional form was assumed, the authors have found a tendency of forecasters to produce overly optimistic GDP forecasts. Pierdzioch et al. (2012) also use EKT's estimator to analyze the asymmetry of inflation and output growth forecasts issued by the Bank of Canada. The authors find evidence for asymmetry only for the next-year forecasts, but not for the forecasts made for the end of the current year. Overestimating next-year inflation was associated with a higher loss than that implied by underestimation. The asymmetry of output growth forecasts pointed into the other direction (Pierdzioch, et al. (2012)).

The contribution of this paper is to demonstrate that the EKT loss function estimation method depends critically on some implicit identifying assumptions — a fact that is not recognized in the literature. In particular, the results obtained by the EKT estimator are very sensitive to the assumption that the posited parametric loss functions depend on the forecast error only. More generally, however, the forecaster's loss may also depend on the level of the target variable and/or the forecast. (We will say that such loss functions are endowed with "level effects.") For example, in the context of central bank inflation forecasts, it is not only reasonable to entertain the possibility that the pertinent loss function is asymmetric, but also that a 1 percentage point forecast error has different implications when average inflation is at 3 percent and when it is at, say, 9 percent.

If loss functions with level effects are not ruled out, one can use a transformation called "Osband's principle" (after Osband (1985)) to show that the first order conditions described above do not uniquely determine the underlying loss function. To outline Osband's principle, I have to define elicitability first. Let us consider a loss function and a statistical functional (e.g. mean, median). If the correct forecast uniquely minimizes the functional, we call the functional elicitable (Fissler and Ziegel (2016)). Osband's principle shows that a functional that is not elicitable can be a component of an elicitable vector-valued functional (Fissler and Ziegel (2016), Gneiting (2011), Osband (1985)).

The key observation that allows the application of the transformation based on Osband's principle is that any given forecast is naturally a function of the information that was available to the forecaster at the time. Hence, any function of the forecast must be uncorrelated with the first derivative of the loss function with respect to the forecast. As I will show, the generality of this condition permits substantial ambiguity about the underlying loss function. I also use concrete examples to demonstrate that loss functions with completely different directions of asymmetry may generate the same first order conditions and are hence observationally equivalent.

For example, I embed the standard asymmetric lin-lin loss function with asymmetry parameter α into a larger class that has an additional shape parameter b. A change in b causes an economically meaningful change in the loss function and yet the first order condition that defines the optimal forecast is invariant to the value of b. Hence, this parameter is unidentified from any data on observed forecasts and realizations. This is very problematic because for $b = b1$ the forecaster may prefer negative forecast errors when, say, the realization is $y = 1$ whereas a positive forecast error would be preferred when $b = b2$ and $y = 1$. Thus, nothing can be inferred from the data about the direction of asymmetry of the underlying loss function (at least in the presence of level effects). I present a similar embedding for quad-quad loss functions.

While the identification problem exposed in this paper is fundamental, it does not mean that the EKT approach to loss function estimation should be abandoned altogether. Rather, it clarifies what additional arguments are needed for the application of the method in practice. In particular, a careful theoretical argument should be made that level effects are not relevant in the situation at hand, and hence restricting attention to the error-dependent loss functions posited by EKT is appropriate. In this situation the asymmetry parameter α is identified and meaningful both in the lin-lin and quad-quad cases. Nevertheless, we note that such a discussion is lacking in the empirical papers reviewed above.

The rest of the paper is organized as follows. Section 2.2 describes the forecaster's expected loss minimization problem. Section 2.3 introduces the identification problem of EKT's approach, and section 2.4 illustrates the identification problem by showing examples of how small changes in the assumed loss give rise to vastly different asymmetry parameter estimates. The last section concludes.

2.2. Setup

Let $\hat{Y}_{t+1}$ denote the forecast of the target variable Y_{t+1} made by the forecaster at time t. The information set available to the forecaster on which the forecast is based is denoted as Ω_t; the predictive distribution p_t is the conditional distribution of Y_{t+1} given Ω_t. The loss function possessed by the forecaster is modeled by the econometrician as $\ell(\hat{y}, y; \theta)$, where θ is a finite dimensional vector of parameters. If the model for the loss function is correctly specified, there exists some value θ_0 of the parameters such that the observed forecast $\hat{Y}_{t+1}$ minimizes the expectation of $\ell(\hat{y}, y; \theta_0)$ with respect to p_t, and hence satisfies the corresponding FOC:

$$\frac{d}{d\hat{y}} \int \ell(\hat{Y}_{t+1}, y; \theta_0) dp_t(y) = \int \ell_{\hat{y}}(\hat{Y}_{t+1}, y; \theta_0) dp_t(y) = 0. \tag{2.1}$$

While p_t itself is unobserved by the econometrician, (2.1) and the law of iterated expectations imply

$$E[W_t \ell_{\hat{y}}(\hat{Y}_{t+1}, Y_{t+1}; \theta_0)] = 0, \tag{2.2}$$

where W_t is any random vector measurable with respect to Ω_t for which the expectation is well defined. Thus, as put forward by EKT, one can estimate the loss function parameters θ_0 using moment conditions of the form (2.2) without full knowledge of Ω_t. All that is required is some instrument vector W_t that is plausibly available to the forecaster at time t.

While the estimation strategy described above cleverly handles the problem that p_t is not observed by the econometrician, Osband's principle implies that strong additional assumptions are needed to identify the forecaster's loss function. The identification problem is detailed further in the next section.

10.14754/CEU.2023.02

2.3. The Identification Problem

2.3.1 Implications of Osband's principle for moment based loss function estimation[2]

To illustrate the practical relevance of the abstract identification problems addressed in this paper, we draw out the consequences of Osband's principle for the moment based loss function estimation method proposed by Elliott, Timmermann and Komunjer (2005; henceforth, EKT)[3]. In EKT's framework identification is achieved by assuming very specific classes of loss functions: asymmetric absolute loss (lin-lin) or asymmetric quadratic loss (quad-quad) with the asymmetry parameter α to be estimated. Nevertheless, Osband's principle implies that given α, there is a multitude of other losses that explain the data equally well (e.g., other generalized α-piecewise linear (α-GPL)[4] losses in case of lin-lin). Thus, any conclusion drawn about the shape of the underlying loss from the estimate of α is extremely sensitive to the specific parametrization used. We will now make this argument formal and provide a partly empirical example.

2.3.2 The ambiguity of the EKT moment conditions[5]

The estimation strategy described in the setup cleverly handles the problem that p_t is not observed by the econometrician, Osband's principle implies that strong additional assumptions are needed to identify the forecaster's loss function. In particular, equation (17) in Lieli at al. (2019) shows that given a suitable weight function $w(\cdot)$, the loss function

---

[2]This subsection was published as section 5. in Lieli et al (2019).

[3]Appendix 1 gives a more detailed description of Osband's principle.

[4]α-GPL loss functions are defined in the following way. For each $\alpha \in (0, 1)$, the loss functions in $\mathcal{L}^{\alpha}_{GPL}$ are those of the form

$$\ell(\hat{y}, y) = [\mathbb{1}(y < \hat{y}) - \alpha][\psi(\hat{y}) - \psi(y)], \tag{2.3}$$

where $\psi(\cdot)$ is any continuous, strictly increasing function and $\mathbb{1}(\cdot)$ denotes the indicator function.

[5]This subsection was published as section 5.1. in Lieli et al (2019).

$\ell^{\dagger}(\hat{y}, y; \theta_0) = \int_{-\infty}^{\hat{y}} w(t)\ell_{\hat{y}}(t, y; \theta_0)dt$ satisfies

$$\frac{d}{d\hat{y}} \int \ell^{\dagger}(\hat{Y}_{t+1}, y; \theta_0)dp_t(y) = \int w(\hat{Y}_{t+1})\ell_{\hat{y}}(\hat{Y}_{t+1}, y; \theta_0)dp_t(y) = 0.$$

Again, for any Ω_t-measurable random vector W_t this implies

$$E[W_t \ell^{\dagger}_{\hat{y}}(\hat{Y}_{t+1}, Y_{t+1}; \theta_0)] = E[W_t w(\hat{Y}_{t+1})\ell_{\hat{y}}(\hat{Y}_{t+1}, Y_{t+1}; \theta_0)] = 0, \tag{2.4}$$

provided that the expectations exist. The identification problem arises because $W_t w(\hat{Y}_{t+1})$ is also Ω_t-measurable, since $\hat{Y}_{t+1}$ is, of necessity, a function of information available at time t. Hence, the moment condition (2.4) has two equally valid interpretations: it can either be regarded as a consequence of the forecaster's FOC under the loss $\ell(\hat{y}, y; \theta_0)$ and instrument choice $W_t w(\hat{Y}_{t+1})$ or, alternatively, the loss $\ell^{\dagger}(\hat{y}, y; \theta_0)$ and instrument choice W_t. Even if these moment conditions point-identify θ_0, the loss functions ℓ and $\ell^{\dagger}$ may look very different. Knowledge of θ_0 alone says very little, if anything, about the shape of the underlying loss.

2.4. Empirical examples

2.4.1 Empirical example: generalized piecewise-linear losses, different parameters[6]

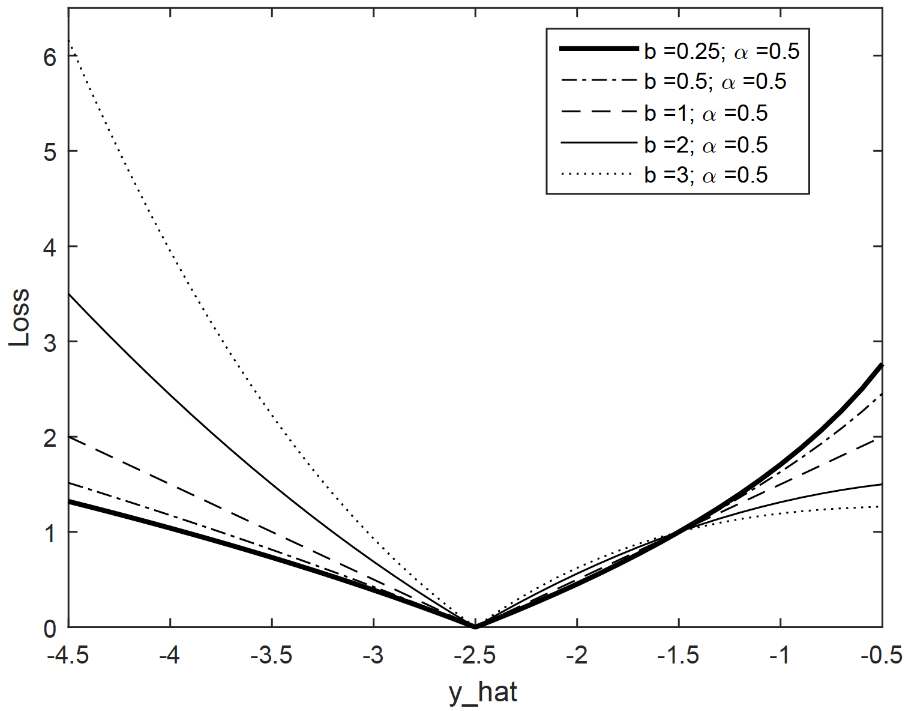
An example will help reinforce the argument. Let us embed lin-lin losses into a larger class of generalized piecewise-linear (GPL) loss functions, introduced by Patton (2016). Let y denote the target variable and $\hat{y}$ denote the forecast and let $\mathbb{1}(\cdot)$ denote the indicator function. For $b > 0$ the function $\psi(t) = sgn(t)|t|^b$ is strictly increasing, so

$$\ell(\hat{y}, y; \alpha) = [\mathbb{1}(y - \hat{y} < 0) - \alpha] \cdot [sgn(\hat{y})|\hat{y}|^b - sgn(y)|y|^b], \quad \alpha \in (0, 1) \tag{2.5}$$

is indeed a collection of GPL losses for any $b > 0$. Setting $b = 1$ corresponds to lin-lin loss, but other values of b give rise to very differently shaped loss functions asymmetric in either direction; see Figure 2.1 for an illustration with α fixed at 0.5.

[6]This subsection was published as section 5.2. in Lieli et al (2019).

Figure 2.1: The loss functions (2.5) for different values of b and $\alpha = 1/2$. The realization y is fixed at -2.5 and the forecast $\hat{y}$ varies.



While each loss function exhibited on Figure 2.1 has the property that the median of p_t is the optimal point forecast for any distribution p_t, these losses are not economically equivalent. For example, suppose that the forecaster is presented with the following question: "If $Y_{t+1} = 1$, how much would you be willing to pay to avoid a forecast error of size $+1$ versus size -1?" Clearly, a forecaster whose loss is given by the dashed line would respond differently than a forecaster whose loss is given by the dotted line.

Let us turn to the moment conditions for estimating α that result from specification (2.5). In this case equation (2.2) takes the form

$$E\Big\{W_t|\hat{Y}_{t+1}|^{b-1} \times \big[-\alpha 1(Y_{t+1} > \hat{Y}_{t+1}) + (1-\alpha)1(Y_{t+1} < \hat{Y}_{t+1})\big]\Big\} = 0. \tag{2.6}$$

For any given $b > 0$ the generalized method of moments estimator of α derived from the sample analog of (2.6) is isomorphic to EKT's estimator with the instrument choice $W_t|\hat{Y}_{t+1}|^{b-1}$ (in EKT's setting $b = 1$). To highlight the ambiguity in interpreting estimates of α, we revisit one of EKT's original applications involving annual budget deficit forecasts

	$b = 0.25$	$b = 0.5$	$b = 1$	$b = 2$	$b = 3$
$\hat{\alpha}$	0.46	0.46	0.45	0.40	0.33
s.e.	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
p-value ($\alpha = .5$)	[0.36]	[0.35]	[0.32]	[0.18]	[0.07]

Note: Based on the IMF's current year budget deficit forecasts for France. Sample period: 1980-2017; W_t = constant, lagged budget deficit (EKT's instrument 3). The GMM weighting matrix $\hat{S}$ is specified as in EKT. The case $b = 1$ corresponds to the original EKT estimator (their estimate of α is 0.54).

Table 2.1: Estimated α parameters for various values of b

published by the IMF for various countries. Using (2.6), we re-estimate the α parameter for different values of b, while setting W_t equal to one of the original instruments considered by EKT.[7]

An important argument should be taken into consideration when interpreting the results from Table 2.1 and comparing them to the results from EKT[8]. If there are differences between the point estimates that go beyond small sample variation, a plausible explanation is that the family of losses (2.6) is misspecified in the broader sense that the point forecasts reported by the forecaster do not correspond to a fixed quantile of the underlying predictive distributions. Other possible interpretations include instrument invalidity (W_t is not in the forecaster's information set) or forecaster irrationality.

A small set of results is shown in Table 2.1. While $\hat{\alpha}$ varies somewhat as a function of b, the null hypothesis that $\alpha = 0.5$ cannot formally be rejected at the 5% level in any of the cases (albeit the conclusion is borderline for $b = 3$). If, as in EKT, one takes lin-lin as the underlying loss, these estimates suggest no significant deviation from symmetry—the different values of b correspond simply to different instruments. Nevertheless, an alternative, and *a priori* equally plausible interpretation of (2.6) is that the underlying losses belong to the set (2.5) with some value of b different from 1. Plotting the loss functions corresponding to different $(b, \hat{\alpha})$ pairs in Table 2.1 would yield a picture similar to Figure 2.1 — the observed budget forecasts and realizations can also be rationalized by loss functions

[7]As we do not have access to EKT's original data set, we collected our own data for the sample period 1980-2017. Note that this is not perfectly aligned with EKT's original sample (1975-2001 for OECD data and 1976-2000 for IMF data). We did not find the IMF and OECD forecasts made before 1980 and we included data made after 2000 in order to provide a larger sample.

[8]To ease comparison, the results of Table 2 in EKT (2005) are included in appendix 2, table 4

asymmetric in either direction (see figure 2.2 in the next section)!

The conservative interpretation of the α estimates reported in EKT's Table 2 (and our Table 2.1) is that they simply approximate which quantile of the forecaster's predictive distribution is used as the point forecast. There are then a diverse class of α-GPL losses that can potentially rationalize this behavior. Hence, any conclusions drawn by EKT about the (a)symmetry of the underlying loss[9] is conditional on *complete* trust in the lin-lin specification. However, our results in Section 3 of Lieli et al. (2019) show that different α-GPL losses are observationally equivalent even in controlled environments, so it requires strong *theoretical* arguments to single out any particular subclass as the appropriate model of forecaster behavior.

A property that makes lin-lin losses special among GPL losses is that the forecaster's loss is a function of the forecast error $y - \hat{y}$ only, independently of the level of $\hat{y}$ or y. All of EKT's statements about symmetry are critically dependent on this implicit identifying assumption. In general, any loss function estimation exercise that uses the lin-lin (or quad-quad) specification should argue the point that a loss function solely dependent on the forecast error is appropriate for the situation at hand.

2.4.2 Systematic replication of EKT with various loss functions

In this subsection I conduct a systematic replication of the EKT results and present a broader set of results compared to subsection 2.4.1. My sample period (1980-2017) is not perfectly aligned with EKT's original sample (1975-2001 for OECD data and 1976-2000 for IMF data), so this might explain in part the different estimates. The reason why the sample periods differ from those of EKT is that I did not have access to their original dataset. I did not find the IMF and OECD forecasts made before 1980 and I included data made after 2000 in order to provide a larger sample.

I follow the same methodology as in subsection 2.4.1, and re-estimate the α parameter for annual budget deficit forecasts published by the IMF and the OECD for various countries for different values of b, while setting W_t equal to the original instrument sets considered by EKT. Setting $b = 1$ corresponds to lin-lin loss, but other values of b give rise to

[9] One example: "[T]he point estimates of α suggest strong asymmetries in the loss function... For some countries they indicate that underpredictions of budget deficits are viewed as up to three times costlier than overpredictions." (EKT, p. 1117)

10.14754/CEU.2023.02

		IMF							OECD				
		Canada	France	Germany	Italy	Japan	UK	US	US	France	Germany	Italy	UK
Current year													
Inst=1	alpha	0.47	0.43	0.43	0.43	0.38	0.38	0.29	0.41	0.38	0.33	0.35	0.41
	s.e	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.09	0.10	0.09	0.09	0.09
	p-value	0,81	0,51	0,51	0,51	0,26	0,26	0,06	0,33	0,23	0,07	0,10	0,33
Inst=2	alpha	0.43	0.44	0.40	0.45	0.28	0.35	0.31	0.42	0.40	0.30	0.36	0.42
	s.e	0.12	0.11	0.11	0.11	0.10	0.11	0.12	0.10	0.10	0.09	0.10	0.10
	p-value	0,55	0,61	0,35	0,62	0,03	0,15	0,10	0,42	0,30	0,03	0,14	0,43
Inst=3	alpha	0.43	0.45	0.40	0.45	0.30	0.35	0.27	0.40	0.39	0.26	0.36	0.42
	s.e	0.12	0.11	0.11	0.11	0.10	0.11	0.11	0.10	0.10	0.09	0.10	0.10
	p-value	0,57	0,64	0,36	0,65	0,05	0,15	0,04	0,32	0,25	0,00	0,14	0,42
Inst=4	alpha	0.43	0.44	0.39	0.44	0.20	0.34	0.23	0.40	0.37	0.26	0.36	0.42
	s.e	0.12	0.11	0.11	0.11	0.09	0.11	0.11	0.10	0.10	0.09	0.10	0.10
	p-value	0,54	0,61	0,33	0,62	0,00	0,13	0,01	0,32	0,19	0,00	0,14	0,42
1-year ahead													
Inst=1	alpha	0.40	0.33	0.40	0.47	0.67	0.40	0.40	0.42	0.40	0.42	0.54	0.50
	s.e	0.13	0.12	0.13	0.13	0.12	0.13	0.13	0.10	0.10	0.10	0.10	0.10
	p-value	0,43	0,17	0,43	0,80	0,17	0,43	0,43	0,43	0,31	0,43	0,68	1,00
Inst=2	alpha	0.33	0.35	0.38	0.50	0.96	0.39	0.41	0.44	0.37	0.44	0.52	0.54
	s.e	0.13	0.13	0.13	0.13	0.05	0.13	0.13	0.10	0.10	0.10	0.10	0.10
	p-value	0,19	0,25	0,33	1,00	0,00	0,40	0,50	0,53	0,20	0,55	0,83	0,70
Inst=3	alpha	0.33	0.36	0.42	0.50	0.94	0.42	0.41	0.43	0.34	0.44	0.53	0.52
	s.e	0.13	0.13	0.13	0.13	0.06	0.13	0.13	0.10	0.10	0.10	0.10	0.10
	p-value	0,18	0,26	0,54	1,00	0,00	0,52	0,50	0,50	0,11	0,52	0,81	0,83
Inst=4	alpha	0.31	0.35	0.30	0.50	1.00	0.35	0.39	0.43	0.34	0.44	0.53	0.54
	s.e	0.12	0.13	0.12	0.13	0.00	0.13	0.13	0.10	0.10	0.10	0.10	0.10
	p-value	0,12	0,25	0,11	0,99	0,00	0,24	0,40	0,48	0,10	0,52	0,80	0,68

Note: Sample period: 1980-2017; The four instrument sets are based on EKT and are the following from inst=1 to inst=4: (i) constant; (ii) constant, lagged forecast error; (iii) constant, lagged budget deficit; (iv) constant, lagged forecast error and lagged budget deficit. The GMM weighting matrix $\hat{S}$ is specified as in EKT. The case $b = 1$ corresponds to the original EKT estimator. p values refer to the null hypothesis of $\alpha = 0.5$.

Table 2.2: Estimated α parameters for $b = 1$

very differently shaped loss functions asymmetric in either direction. To ease comparison, the results of Table 2 in EKT (2005) are included in appendix 2, table 3.6.

The results for the b=1 case are shown in Table 2.2. Starting with the end-of-year IMF budget deficit forecasts using only the constant as instrument (Inst=1), we see no significant rejection of symmetry out of seven countries (at 5% significance level). The original lin-lin loss based estimates of EKT show four rejections as opposed to my results (see Table 3.6 in the appendix). The pattern of low rejection rates is prevalent under the other instrument sets as well: 1/7 when using the constant and a lagged forecast error as instruments (Inst=2), also 1/7 when using lagged budget deficit alongside the constant (Inst=3), and 2/7 in the fourth case, where both lagged forecast error and lagged budget deficit are used as instruments alongside the constant. These rates can be compared to the much higher rates of the original EKT estimates, 4/7 in all four cases. Symmetry was rejected by EKT's results for Italy, Japan, the UK and the US no matter which instrument set was used.

The rejection rates are similar for the end-of-year OECD forecasts: using only the constant as instrument, symmetry of the $\alpha-$estimates is not rejected for either country. When using the other instrument sets, only Germany's estimated α was significantly different from 1/2, while EKT estimated asymmetric α's for Germany, France and Italy in all four cases.

My estimates for the symmetry parameter concerning IMF's next-year budget deficit forecasts differed significantly from 1/2 only in the case of Japan, for the instrument sets 2 to 4. For OECD's next-year budget deficit forecasts, I estimated no significant deviation from symmetry for either country's estimated α parameter.This is in contrast to the EKT estimates; while there are only two out of 16 rejections in the OECD estimates (Germany for the third and fourth instrument set), symmetry of α's are rejected for half of EKT's IMF estimates.

	b=0.25	b=0.5	b=1	b=2	b=3	$\sum$
IMF end-of-year	0.32	0.29	0.14	0.32	0.43	0.30
IMF 1-year ahead	0.25	0.21	0.11	0.43	0.54	0.31
IMF $\sum$	0.29	0.25	0.13	0.38	0.48	0.30
OECD end-of-year	0.25	0.10	0.15	0.20	0.25	0.19
OECD 1-year ahead	0.00	0.00	0.00	0.25	0.60	0.17
OECD $\sum$	0.13	0.05	0.08	0.23	0.43	0.18
$\sum$	0.22	0.17	0.10	0.31	0.46	0.25

Table 2.3: Rejection rates across b-s and forecast subgroups

I have included the tables of detailed results for $b = [0.25; 0.5; 2; 3]$ in the appendix (tables 3.7 through 3.10). In table 2.3, rejection rates are summerized for different b's, forecasters and horizons (cases for different instruments are pooled). The results show higher rejection rates for the symmetry of estimated α's for $b = 2$ and $b = 3$ compared to the estimates for lower b's. Rejection rates do not differ much depending on the forecast horizon for either institution. However, the forecasting institution does seem to make a difference as on average, rejection rates for IMF forecasts are higher (0.3 averaged over the two horizons) than rejection rates for OECD (0.18 averaged over the two horizons) forecasts.

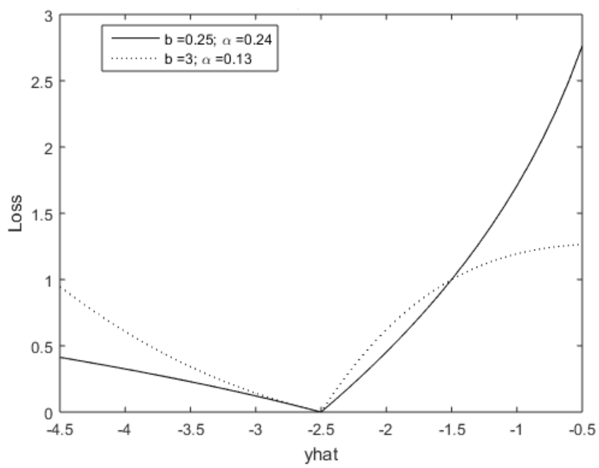
Overall, we can conclude that my estimated rejection rates of the hypothesis $\alpha = 0.5$ are much lower than EKT's estimated rejection rates (an average of 0.25 as opposed to EKT's 0.46). If, following EKT, one takes lin-lin as the underlying loss, these estimates suggest few significant deviations from symmetry, with the different values of b simply corresponding to different instruments. Nevertheless, an alternative, and *a priori* equally plausible interpretation of Eq. 2.6 is that the underlying losses belong to the set in Eq. 2.5, with some value of b being different from 1. To illustrate this, I plot four loss functions corresponding to different $(b, \hat{\alpha})$ pairs in Figure 2.2. Depending on b, the observed budget forecasts and realizations can also be rationalized by loss functions asymmetric in either direction!

Appendix 7 includes further examples using losses of different parametric families that are forecast equivalent to quad-quad.

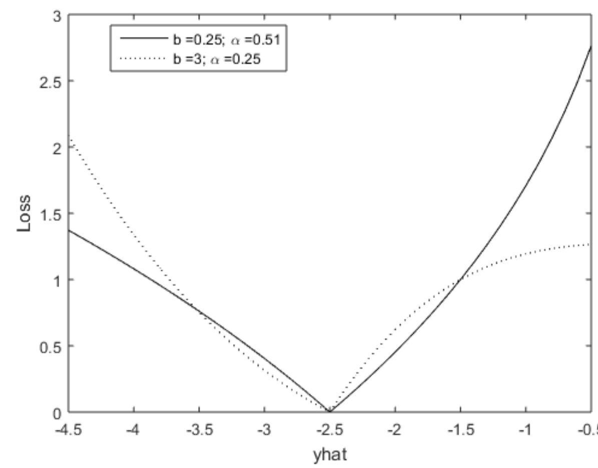
10.14754/CEU.2023.02

Figure 2.2: Loss functions asymmetric in either direction depending on b. The realization y is fixed at -2.5 and the forecast $\hat{y}$ varies.

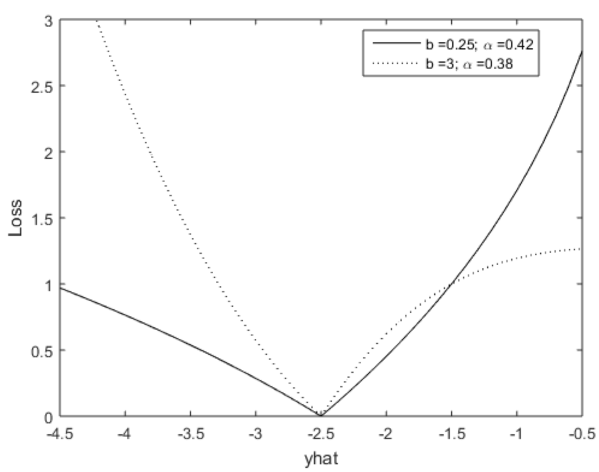
(a) US, instrument 1



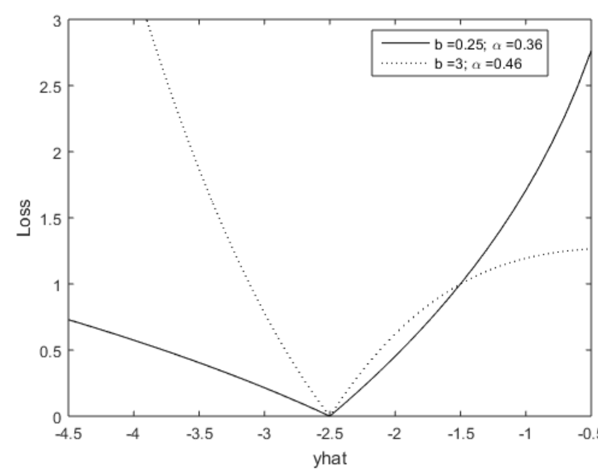
(b) Germany, instrument 2



(c) France, instrument 3



(d) Canada, instrument 4



2.5. Conclusion

The seminal paper by Elliott, Komunjer and Timmermann (2005) proposes a method for estimating a forecaster's loss function based on a moment condition derived from the first order condition of the forecaster's expected loss minimization problem. The contribution of this paper is to demonstrate that the EKT loss function estimation method depends critically on some implicit identifying assumptions — a fact that is not recognized in the literature.

In particular, the results obtained by the EKT estimator are very sensitive to the assumption that the posited parametric loss functions depend on the forecast error only. More generally, however, the forecaster's loss may also depend on the level of the target variable and/or the forecast. (Such loss functions are endowed with "level effects.") For example, in the context of central bank inflation forecasts, it is not only reasonable to entertain the possibility that the pertinent loss function is asymmetric, but also that a 1 percentage point forecast error has different implications when average inflation is at 3 percent and when it is at, say, 9 percent.

If loss functions with level effects are not ruled out, one can use a transformation called "Osband's principle" to show that the first order conditions described above do not uniquely determine the underlying loss function. The key observation that allows the application of this transformation is that any given forecast is naturally a function of the information that was available to the forecaster at the time. Hence, any function of the forecast must be uncorrelated with the first derivative of the loss function with respect to the forecast. As I show, the generality of this condition permits substantial ambiguity about the underlying loss function. I also use concrete examples to demonstrate that loss functions with completely different directions of asymmetry may generate the same first order conditions and are hence observationally equivalent.

For example, I embed the standard asymmetric lin-lin loss function with asymmetry parameter α into a larger class that has an additional shape parameter b. A change in b causes an economically meaningful change in the loss function and yet the first order condition that defines the optimal forecast is invariant to the value of b. Hence, this parameter is unidentified from any data on observed forecasts and realizations. This is very problematic because for $b = b1$ the forecaster may prefer negative forecast errors when, say, the realization is $y = 1$ whereas a positive forecast error would be preferred when $b = b2$ and $y = 1$. Thus, nothing can be inferred from the data about the direction of asymmetry of the underlying loss function (at least in the presence of level effects). I present a similar embedding for quad-quad loss functions, and additional functional forms as well.

10.14754/CEU.2023.02

Chapter 3

Recovering Stock Analysts' Loss Functions from Buy/Sell Recommendations

3.1. Introduction

In this paper, I estimate bounds for the parameter characterizing analysts' loss functions in making stock recommendations. In a binary variable forecasting environment, it is possible to set-identify the parameter that accounts for the forecaster's relative cost for overestimating versus underestimating the target even if the stock analyst's information set is not fully observed (Lieli and Stinchcombe (2013)). In this empirical application of the Lieli and Stinchcombe result, I use binarized stock recommendations as forecasts: buy recommendations account for positive, while hold or sell recommendations account for negative forecasts. The forecast is compared to the one-month-ahead price performance of the stock relative to the market. In the estimation, I also use a proxy for the publicly observed part of the forecaster's information set. The proxy I use is the smooth price per equity ratio. I have chosen this proxy by following Campbell and Thompson (2008), who show that the smooth P/E ratio could be used to predict excess stock returns once weak

10.14754/CEU.2023.02

restrictions hold for the signs of coefficients[1].

My empirical results show high sensitivity to the categorization of hold recommendations. When I assume that 'hold' means 'sell', the estimated asymmetry parameters are relatively high. This suggests that we can rule out analysts' extreme reluctance to propose a 'sell'; they are more likely to issue 'false sells' than 'false buys'. However, when categorizing 'hold' into the buy category, the reverse is found: in almost all cases the highest possible values for the asymmetry parameter are ruled out. When imputing 'hold' with the previous recommendation, again the highest values are ruled out in more than half of the cases.

While financial professionals do not all agree on the information content of analyst stock recommendations, their widespread use and several pieces of evidence from the literature confirm that they are in fact relevant and useful forecasts for the future performance of stocks. It has been shown that analysts' earnings forecasts are superior to mechanical time series models (Brown and Rozeff (1978), Bradshaw et al. (2012)). Empirical evidence also shows that recommendations have some investment value, as they are successful in predicting short-run stock returns (Womack (1996), Loh and Mian (2006)). In their 1998 paper, Barber et al. document that an investment strategy based on the consensus recommendations of security analysts earns positive returns. For the analyzed period between 1986 and 1996, purchasing stocks most highly recommended and selling short those with the worst recommendations yielded a return of 102 basis points a month (Barber et al (1998)). The statement from Barber et al. is confirmed by more recent findings as well: see Jegadeesh et al. (2004) and Green (2006).

Another straightforward argument on the relevance of analyst recommendations is that brokerage houses produce and sell them for millions of dollars every year[2]. If they were in fact useless, why would so much money be spent on their production and sale?

We can see that analyst stock recommendations are in fact relevant. This is also confirmed if we look at the massive attention analyst recommendations get in the academic literature (for a comprehensive picture, see the review on the financial analyst forecasting literature by Ramnath et al. (2008)).

[1]An earlier version of this paper appeared in the Spring Wind 2016 conference volume (Grolmusz (2016)).

[2]A first year equity analyst earned a yearly base salary of \$68,200 plus a bonus of \$48,100 on average in 2013, as reported by the Wall Street Oasis 2013 Compensation Report (Rapoza (2013).

10.14754/CEU.2023.02

We can deduct some important inference from this large body of academic literature on what characteristics of analysts' recommendations are rewarded. First, unsurprisingly, evidence suggests that forecast accuracy is important for an analyst's prestige and career prospects. In their 2003 paper, Hong and Kubik relate earnings forecasts made by security analysts to job separations. They find that forecast accuracy is indeed a substantial factor in an analyst's career outcomes, such as how prestigious is her employer brokerage house, or what kind of stocks is she assigned to cover (Hong and Kubik (2003)). Forecasts are not directly evaluated on their accuracy, but for building reputation and influence among the buy side, it is substantial for the analyst to make the right calls (Hong and Kubik (2003)).

Although accuracy is important, evidence suggests that it is not everything: for the best career perspectives, an analyst also has to publish relatively optimistic recommendations. Controlling for accuracy, analysts who issue a large fraction of forecasts that are more optimistic than the consensus are much more likely to move up the career hierarchy ladder (Hong and Kubik (2003)). This observation is confirmed by Lim (2001), among others. Lim argues that incorporating positive bias in earnings forecasts is a rational action.

Anecdotal evidence also supports the above statement. Lim argues that it is widely known throughout the financial analyst profession that a negative report on a company might result in the involved company's management limiting or eliminating the pessimistic analyst's information flow (Lim (2001)). Other pieces of anecdotal evidence emphasize that analysts need to go along with the management's optimistic projections, or if they do not, they risk being passed over for more loyal analysts (Hong and Kubik (2003), Lim(2001)). The importance of following the management's guidelines is even higher for young and inexperienced analysts, as their risk of unfavorable job separation is much higher than it is for their older colleagues (Hong et al. (2000)). This is the reason why younger analysts tend to avoid making bold forecasts and are more likely to herd (Hong et al. (2000)).

Different theories on the driving forces behind creating analyst recommendations suggest different implications for the direction of bias in the observed recommendations. The above arguments support low risk aversion in analysts for making buy-side recommendations: as analysts are rewarded for issuing relatively optimistic recommendations, they tend to incorporate a positive bias into their recommendations. However, sound arguments for the reverse can also be found. Consider that if an analyst issues a buy recommendation, then in the case of underperformance of the stock, her client will lose money for sure. However, if the analyst recommendation is 'sell', then the client can still lose in the sense of opportunity cost, but it might not be as painful for her (due to loss aversion), and the client might not even observe the performance of the stock as it is not anymore in her portfolio. This argument suggests that a risk-averse analyst should only issue a 'buy', if the probability of the stock outperforming the market is very high. Thus, analysts should be motivated to avoid making overly optimistic recommendations.

The contribution of this paper to the literature is twofold. First, I derive confidence intervals for the bounds of the loss function asymmetry parameter introduced by Lieli and Stinchcombe (2013), and second, I develop an empirical application of their theoretical result in a binary forecasting setting. More concretely, I inspect stock analysts' relative costs for overprediting versus underpredicting the stock's performance, by using a flexible and general method that has not been used up to now. By doing this, I am able to draw conclusions on the relative empirical relevance of the above two channels.

The remainder of the paper is organized as follows. In section 3.2, I outline the theoretical background for preference recovery in a binary forecasting environment, relying on the results from Lieli and Stinchcombe (2013). In section 3.3, I introduce the methodology and the data used in the empirical application. Section 3.4 presents and interprets the results, while the last section concludes.

3.2. Preference Recovery in a Binary Forecasting Environment

The theoretical background for the empirical investigation used in this paper comes from the 2013 paper of Lieli and Stinchcombe. In a binary variable forecasting environment, Lieli and Stinchcombe's paper provides a set identifiction result for the parameter characterizing the forecaster's loss function. In this section, I summarize this theoretical result.

3.2.1 Expected Loss Minimization Problem

Let Y_t be the time series of binary values, and $\hat{Y}_t$ be the time series of Y_t's forecasts made in the previous period ($\hat{Y}_t = \hat{Y}_{t|t-1}$); $t = 1, 2, \ldots, T$; $T < \infty$. In a binary variable forecasting setting, $Y_t, \hat{Y}_t \in \{0, 1\}$, and a loss function can be represented in the following way:

	$Y_t = 1$	$Y_t = 0$
$\hat{Y}_t = 1$	0	ℓ(1,0)
$\hat{Y}_t = 0$	ℓ(0,1)	0

Where $\ell(\hat{Y}_t, Y_t)$ is the loss from forecasting $\hat{Y}_t$ when the realization will be Y_t. We assume that $\ell(1{,}0) \geq 0$ and $\ell(0{,}1) \geq 0$. For expected loss minimizing forecasters, assuming that the loss is zero when the forecaster *hits* the target ($Y_t = \hat{Y}_t$) is true without loss of generality[3].

We assume that forecasters produce their forecasts by minimizing expected loss. Let I_t denote the information set of the forecaster. Then the forecaster solves the following problem:

$$min_{\hat{Y}_t \in \{0,1\}} \ell(\hat{Y}_t, 0) P(Y_t = 0 \mid I_t) + \ell(\hat{Y}_t, 1) P(Y_t = 1 \mid I_t)$$

The solution to this problem is to predict one if $P(Y_t = 1 \mid I_t) > c$, where $c = \frac{1}{1+\frac{\ell(0,1)}{\ell(1,0)}}$, $c \in [0, 1]$. Let us denote c as the asymmetry parameter. The asymmetry parameter depends on the forecaster's relative loss from overpredicting versus underpredicting the target. It is the parameter I would like to estimate.

The key identification problem is that the econometrician does not observe the whole information set on which the forecast is based, but only the public part of it. Following Lieli and Stinchcombe (2013), let us partition the information set I_t into two subsets: the part that the econometrician also observes, Z_t, and the private information of the forecaster, Z'_t. The forecast is based on the whole information set that is only partly observed by the econometrician, that is, the forecaster predicts one if $p_{Z_t, Z'_t} \equiv P(Y_t = 1 \mid Z_t, Z'_t) > c$. Therefore, the econometrician cannot identify the asymmetry parameter exactly, she can only estimate a set in which the parameter lies (Lieli and Stinchcombe (2013)).

Let us define the unconditional and sample probabilities of Y_t and $\hat{Y}_t$ in the following way:

$$p = P(Y_t = 1)$$

$$q = P(\hat{Y}_t = 1)$$

$$\hat{p}_T = \frac{1}{T} \sum_{t=1}^{T} Y_t$$

$$\hat{q}_T = \frac{1}{T} \sum_{t=1}^{T} \hat{Y}_t$$

[3]This fact is due to the following standardization: $\ell^c(\hat{Y}_t, Y_t) = \ell(\hat{Y}_t, Y_t) - \ell(Y_t, Y_t)$, where ℓ^c is the canonical form of the loss function (Lieli and Stinchcombe (2013)).

Then, we can define p_{Z_t} and q_{Z_t} as probabilities conditional on Z_t; the part of the forecaster's information set that the econometrician also observes:

$$p_{Z_t} = P(Y_t = 1 \mid Z_t)$$
$$q_{Z_t} = P(\hat{Y}_t = 1 \mid Z_t) = P(p_{Z_t, Z'_t} > c \mid Z_t),$$

where q_{Z_t} is the proportion of times $\hat{Y}_t = 1$ is observed conditional on Z_t. It is true by the law of iterated expectations[4], that $E[p_{Z_t, Z'_t} \mid Z_t] = p_{Z_t}$.

Using this relationship, Lieli and Stinchcombe (2013) derive the following bounds for the asymmetry parameter:

$$\tfrac{p_{Z_t} - q_{Z_t}}{1 - q_{Z_t}} \leq c \leq \tfrac{p_{Z_t}}{q_{Z_t}}.$$

Let us denote the lower bound as L_t, and the upper bound as U_t: $L_t = \frac{p_{Z_t} - q_{Z_t}}{1 - q_{Z_t}}, U_t = \frac{p_{Z_t}}{q_{Z_t}}$. It is easy to show that $L_t \leq U_t$. It can happen that $U_t \geq 1$ or $L_t \leq 0$, in these cases the bound is not informative. p_{Z_t} and q_{Z_t} could be estimated from the data using logit regressions, and using these estimates, we can give lower and upper bounds L_t and U_t for c.

Lieli and Stinchcombe highlight that their result is very general, as there are no assumptions about the number of omitted variables Z'_t, nor about their distributions. This makes loss function parameter identification possible in a general framework.

3.2.2 Confidence Intervals

To check the statistical significance of the estimates, we need to derive confidence intervals. I do this by setting up a central limit theorem for the averages $\hat{p}_T$ and $\hat{q}_T$, and derive the variances for the estimated upper and lower bounds that are approximated as linear combinations of $\hat{p}_T$ and $\hat{q}_T$.

Definition:

[4] $p_{Z_t} = E[Y_t \mid Z_t] \overset{\text{LIE}}{=} E[E(Y_t \mid Z_t, Z'_t) \mid Z_t] = E[p_{Z_t, Z'_t} \mid Z_t]$

Let Γ_h be the following:

$$\Gamma_h = E\left[\begin{pmatrix} Y_t - p \\ \hat{Y}_t - q \end{pmatrix} \begin{pmatrix} Y_{t-h} - p & \hat{Y}_{t-h} - q \end{pmatrix}\right], h = 0, \pm 1, \pm 2, ...$$

L and U are the lower and upper bounds for the asymmetry parameter c:

$$L = \frac{p-q}{1-q} \leq c \leq \frac{p}{q} = U;$$

$$\hat{L}_T = \frac{\hat{p}_T - \hat{q}_T}{1-\hat{q}_T},\ \hat{U}_T = \frac{\hat{p}_T}{\hat{q}_T}$$

Assumptions:

- Y_t and $\hat{Y}_t$ are weakly stationary,
- Y_t and $\hat{Y}_t$ have absolutely summable covariances: $\sum_{h=0}^{\infty} \Gamma_h < \infty$.

Theorem 1 *Distribution of $\hat{U}_T$ and $\hat{L}_T$*

1. $\sqrt{T}(\hat{U}_T - U) \xrightarrow{\mathrm{d}} N(0, \lambda_U' V \lambda_U)$,
   where $V = \sum_{h=-\infty}^{\infty} \Gamma_h$, *and* $\lambda_U = \begin{pmatrix} \frac{1}{q} \\ -\frac{p}{q^2} \end{pmatrix}$

2. $\sqrt{T}(\hat{L}_T - L) \xrightarrow{\mathrm{d}} N(0, \lambda_L' V \lambda_L)$,
   where $\lambda_L = \begin{pmatrix} 1 \\ \frac{p-1}{(1-q)^2} \end{pmatrix}$

Theorem 2 *Distribution of $\hat{U}_T$ and $\hat{L}_T$*

1. $\sqrt{T}(\hat{U}_T - U) \xrightarrow{\mathrm{d}} N(0, \lambda_U' V \lambda_U)$,
   where $V = \sum_{h=-\infty}^{\infty} \Gamma_h$, *and* $\lambda_U = \begin{pmatrix} \frac{1}{q} \\ -\frac{p}{q^2} \end{pmatrix}$

2. $\sqrt{T}(\hat{L}_T - L) \xrightarrow{\text{d}} N(0, \lambda_L' V \lambda_L)$,
where $\lambda_L = \begin{pmatrix} 1 \\ \frac{p-1}{(1-q)^2} \end{pmatrix}$

The proof is based on the central limit theorem. Appendix 1 contains the sketch of the proof.

3.3. Empirical Strategy and Data

In this section, I show the empirical strategy based on the theory outlined in section 3.2 that I use for the set-identification of the asymmetry parameter from analyst stock recommendations.

3.3.1 Empirical Strategy

To estimate the bounds given in section 3.2, we need to estimate p_{Z_t} and q_{Z_t}. If Z_t is an empty set, $\hat{p}_T$ and $\hat{q}_T$ are used to give the unconditional bounds for the asymmetry parameter.

If Z_t is non-empty, then p_{Z_t} and q_{Z_t} could be estimated using fitted values from the following logit regressions, using observations collected over time:

$$\hat{p}_{Z_t} = logit(Z_t'\hat{\beta}_p) = \frac{1}{1+e^{-Z_t'\hat{\beta}_p}}$$
$$\hat{q}_{Z_t} = logit(Z_t'\hat{\beta}_q) = \frac{1}{1+e^{-Z_t'\hat{\beta}_q}},$$

One can use the time series $\hat{p}_{Z_t}$ and $\hat{q}_{Z_t}$ (t=1, 2, ..., T) to derive L_t and U_t for every t. We could use different definitions for the overall bounds for c. We can either take max L_t and min U_t to be lower and upper bounds, respectively, or we could choose the minimum range min $(U_t - L_t)$ and denote its bounds as the overall highest and lowest bound. I use the latter method in the empirical exercise.

3.3.2 Data

As forecast data, $\hat{Y}_t$, I use monthly analyst stock recommendations for shares of Goldman Sachs and 3M Company. I have chosen these Blue Chip stocks because they are highly liquid and I have access to many individual analyst recommendations on them[5]. Analyst stock recommendations are usually published using similar rating scales, categorized into three to five levels. I standardize the different scales and binarize the recommendations in the following way: take $\hat{Y}_t = 1$ if the recommendation is strong buy, buy, or equivalent, and take $\hat{Y}_t = 0$ for sell, and strong sell recommendations. I impute missing observations with the previous recommendation. The categorization of hold recommendations is not straightforward, I use three different ways for treating these observations: imputing by zero (equivalent to sell), imputing by one (equivalent to buy), and imputing with the recommendation from the previous period. Imputing with the previous recommendation can be argued for if we treat 'holds' similarly to missing observations; I assume that an analyst issues a hold recommendation if she does not have any new information or expectation on the future behavior of stock price.

The time series I compare the forecasts to is Y_t, called the actual or realized series. I define Y_t to be one if the price growth[6] of Goldman or 3M Co. is positive and higher than the growth of the Dow Jones Industrial Average in one months from making the forecast:

$Y_t = 1$ if $\frac{P_{D,t+1}}{P_{D,t}} < \frac{P_{G,t+1}}{P_{G,t}}$ and $\frac{P_{G,t+1}}{P_{G,t}} > 1$

$Y_t = 0$ if $\frac{P_{D,t+1}}{P_{D,t}} \geq \frac{P_{G,t+1}}{P_{G,t}}$ or $\frac{P_{G,t+1}}{P_{G,t}} < 1$

where G: Goldman or 3M co., D: Dow Jones index

I compare the two stocks to the Dow Jones index, as Goldman Sachs and 3M Co. stocks are classic Blue Chip stocks. The length of the time series varies from analyst to analyst: it starts in 2003 the earliest (but in most cases, only after 2009), end ends in November 2016.

I present unconditional results along with conditional bounds, for which I include explanatory variables in the logit regressions. The included variable is a proxy for the public part of the analyst's information set used to make the recommendation. I follow Campbell and Thompson (2008), and use the smooth P/E ratio as a proxy for the analyst's

[5]I use a Bloomberg terminal and Reuters Eikon for data collection.

[6]Price is taken to be the end-of-month closing price of Goldman and 3M Co. stocks. Analyst recommendations are also published at the end of each month.

10.14754/CEU.2023.02

information set. The data I use was accessed using Bloomberg[7] and Reuters Eikon.

[7] I have access to Bloomberg Terminals in Corvinus University of Budapest's Financial Laboratory.

3.4. Empirical Results

In this section, I show and interpret the results from the empirical analysis. The estimation gives an upper and a lower bound for the asymmetry parameter of each analyst. The unconditional bounds are the estimates based on the sample averages $\hat{p}_T$ and $\hat{q}_T$. In the conditional case, upper and lower bounds are estimated based on the logit regression for every period t. Then, the largest lower bound L_t and smallest upper bound U_t are presented as the conditional bounds for the sample period.

3.4.1 Interpretation

How could we interpret the results; e.g. what does a $[0, 0.25]$ result mean? Ruling out the highest values for the asymmetry parameter means that the representative analyst is not extremely risk-averse in proposing a buy strategy. In this case, let us assume that the asymmetry parameter takes its highest estimated value, 0.25. Then, by writing up the definition for c:

$$0 \leq \frac{1}{1 + \frac{\ell(0,1)}{\ell(1,0)}} \leq 0.25$$

$$\downarrow$$

$$3 \leq \frac{\ell(0,1)}{\ell(1,0)}$$

$$\downarrow$$

$$3 \times \ell(1,0) \leq \ell(0,1),$$

This means that a 'false sell' is at least three times as costly as 'false buy'. This would make the analyst reluctant to propose a sell strategy. If the upper bound is below 0.5, the analyst has asymmetric loss: she is more inclined to overpredict the target than to underpredict it. On the other hand, when the lower bound is above 0.5, the analyst is more likely to issue more pessimistic recommendations than overly optimistic ones.

It is important to analyze the relationship between the variation in the time series and their consequences on c in more detail. Let me show the consequences on c, when there is absolutely no variation in the recommendation series. If the analyst recommends to sell the stock and the recommendation stays the same ($\hat{Y}_t = 0$) throughout the entire time series, then $p \in]0, 1[$ and $q = 0$. We assume that there is some variation in the binarized actual series.

$$\hat{L}_T = p$$

$$\hat{U}_T = \frac{p}{0} \rightarrow \infty$$

Similarly, if the analyst recommends to 'buy' the stock and the recommendation stays the same ($\hat{Y}_t = 1$) throughout the entire time series, then $p \in]0, 1[$ and $q = 0$. We assume that there is some variation in the binarized actual series.

$$\hat{L}_T = \frac{p-1}{0} \rightarrow -\infty$$

$$\hat{U}_T = p$$

3.4.2 Results

Table 1 shows the results for Goldman Sachs stocks, analyzed by fifteen brokerage houses in the sample. When we categorize hold as zero (hold is the same as a sell), we see that in eight cases, the lowest c's are ruled out. This suggests that for these eight analysts, a 'false buy' is likely costlier than a 'false sell'. We cannot conclude that these analysts have undoubtedly asymmetric loss functions, as the lower bonds are below 0.5. These results are in line with the argument for high risk aversion in making buy side recommendations: it is less costly for the analyst to suggest a sell (or hold), as he expects the client not to observe the stock's price performance after taking it out from the portfolio. If there are many hold recommendations in the time series, observing high asymmetry parameters might be due to the categorization of 'holds' as 'sells'.

The unconditional bounds for Oppenheimer's analyst are uninformative. This is because there are exactly as many ones in the binary actual series than in the binary recommendation series. Therefore, $\hat{p}_T = \hat{q}_T$, and hence $\hat{L}_T = 0$ and $\hat{U}_T = 1$. For the rest of the sample (six analysts out of the fifteen), the highest c's are ruled out: a 'false sell' is likely to be costlier than a 'false buy'.

The conditional bound intervals are narrower in all cases for hold=0 (column 2). This suggests that the smooth P/E ratio bears some forecasting power for stock price performance. In three cases (Wells Fargo, Macquarie and Oppenheimer) the estimated upper bound is lower than the estimated lower bound. In these cases, the estimated bounds are not informative.

When categorizing 'holds' as 1 (buy), the results change significantly. In all but one case, the highest asymmetry parameters are ruled out, suggesting that a 'false sell' is costlier than a 'false buy'[8]. This is in line with the argument for low risk aversion in making buy side recommendations: analysts might be biased towards optimistic recommendations. Analysts who are relatively more optimistic in their stock recommendations than the consensus can expect better career prospects, as it was shown by Hong and Kubik (2003).

The upper bounds are around 0.5 in most cases, suggesting certain asymmetry for c[9]. The conditional logit regressions produce results where the intervals for c become even narrower. E.g., we can conclude that Vining Sparks analysts are at least 5.25 times more likely to produce a 'false buy' than a 'false sell', when making recommendations for Goldman Sachs stocks.

In the last specification, we treat 'holds' similar to missing values and impute them with the previous recommendation. Depending on the exact time series, i.e. the typical recommendation and number of 'holds', this produces similar bounds as the hold=0 or the hold=1 categorization: in ten cases, the bounds are the same as in columns 1-2 (hold=0), and in five cases, they are equivalent to treating 'hold' as 1.

Figures 1, 2, 3 and 4 illustrate the sensitivity of the results on the categorization of 'holds'. In Figure 1, in UBS's case we see that the hold=previous specification gives the same bounds as the hold=1 (buy) specification. However, for Morgan Stanley (Figure 2), the bounds for hold=previous are the same as the bounds for hold=0 (sell). It can also happen that all three specifications produce different bounds (see Figure 3 for JMP), or in the absence of 'holds', all three pairs of estimates are the same (as for Credit Suisse, Figure 4).

The estimates on the other Blue Chip stock, 3M Company are quite similar to the results on Goldman Sachs. When hold is categorized as zero, c is relatively high in six cases (meaning that analysts are not too reluctant to propose a sell strategy). The lower bounds in the unconditional hold=sell case are on average lower than for Goldman Sachs

[8]The lower asymmetry parameters are ruled out in the estimated bounds for Societe Generale. This time series does not contain any hold recommendations, only 'sells'.

[9]I have not yet calculated the confidence intervals for the unconditional bounds. However, taking into consideration that in most cases, the upper bound or the lower bound is uninformative (i.e. $\hat{L}_T = 0$ or $\hat{U}_T = 1$) , it appears that the confidence intervals will be wide. This might change the interpretation of the results.

	hold=0		hold=1		hold=prev	
analyst's firm	unconditional	conditional	unconditional	conditional	unconditional	conditional
Wells Fargo	[0.14, 1]	[0.67, 0.5]	[0, 0.44]	[0, 0.29]	[0, 0.44]	[0, 0.29]
Nomura	[0.22, 1]	[0.34, 0.76]	[0, 0.47]	[0, 0.35]	[0.22, 1]	[0.34, 0.76]
Morgan Stanley	[0.27, 1]	[0.65, 0.68]	[0, 0.44]	[0, 0.3]	[0.27, 1]	[0.65, 0.68]
JMP	[0.38, 1]	[0.43, 1]	[0, 0.4]	[0, 0.35]	[0, 0.63]	[0, 0.44]
Barcleys	[0.41, 1]	[0.49, 1]	[0, 0.41]	[0, 0.32]	[0.41, 1]	[0.49, 1]
Macquarie	[0.43, 1]	[0.62, 0.23]	[0, 0.44]	[0, 0.23]	[0.43, 1]	[0.62, 0.23]
Societe Generale	[0.44, 1]	[0.53, 1]	[0.44, 1]	[0.53, 1]	[0.44, 1]	[0.53, 1]
RBC	[0.46, 1]	[0.52, 1]	[0, 0.53]	[0.28, 0.43]	[0.46, 1]	[0.52, 1]
Oppenheimer	[0,1]	[0.68, 0.36]	[0, 0.45]	[0, 0.31]	[0, 0.45]	[0, 0.31]
Atlantic	[0, 0.28]	[0, 0.02]	[0, 0.28]	[0, 0.02]	[0, 0.28]	[0, 0.02]
Credit Suisse	[0, 0.37]	[0, 0.18]	[0, 0.37]	[0, 0.18]	[0, 0.37]	[0, 0.18]
Vining Sparks	[0, 0.48]	[0.15, 0.23]	[0, 0.46]	[0, 0.16]	[0, 0.46]	[0, 0.16]
Rafferty	[0, 0.62]	[0, 0.34]	[0, 0.4]	[0, 0.25]	[0, 0.4]	[0, 0.25]
UBS	[0, 0.71]	[0.17, 0.33]	[0, 0.24]	[0, 0.2]	[0, 0.24]	[0, 0.2]
Evercore	[0, 0.93]	[0.1, 0.8]	[0, 0.41]	[0, 0.24]	[0, 0.94]	[0.1, 0.8]

Table 3.1: Bounds for the asymmetry parameter, Goldman Sachs stocks

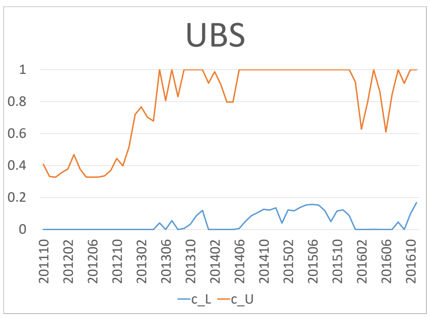
The forecast value is the binarized analyst recommendation for Goldman stocks made in t (strong buy, buy: 1; sell, strong sell: 0). Missing values in the recommendation series are imputed by the previous value. Hold recommendations are categorized as 0 (column 1 and 2), 1 (column 3 and 4), and imputed by the previous value (columns 5 and 6). The actual value is one if price growth for Goldman stocks is positive and outperforms the DJI one month from making the forecast, and zero otherwise. The explanatory variable Z is the smooth P/E ratio of Goldman Sachs in t.

10.14754/CEU.2023.02

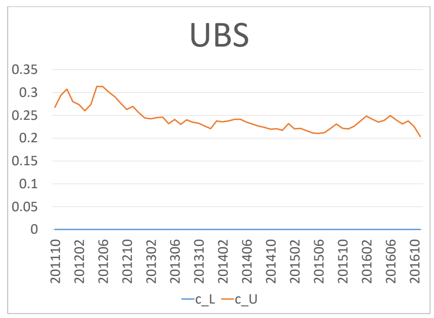
10.14754/CEU.2023.02

Figure 3.1: Bounds for c based on conditional probability estimates, UBS analyst recommendation for Goldman stocks

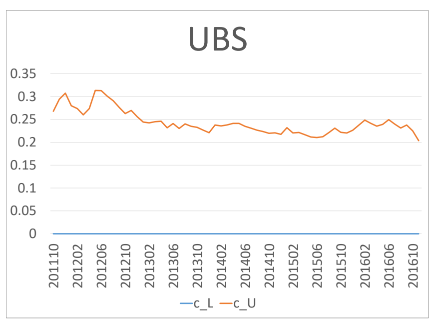
(a) hold=0



(b) hold=1



(c) hold=previous

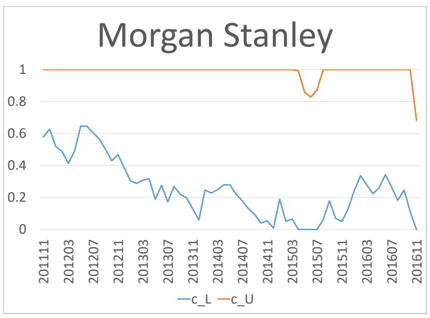


Bounds for c estimated using conditional logit regression. The forecast value is the binarized analyst recommendation for Goldman stocks made in t (strong buy, buy: 1; sell, strong sell: 0). Hold recommendations are categorized as 0 (a), 1 (b), and imputed by the previous value (c). The actual value is one if price growth for Goldman stocks is positive and outperforms the DJI one month from making the forecast, and zero otherwise. The explanatory variable Z is the smooth P/E ratio of Goldman Sachs in t.

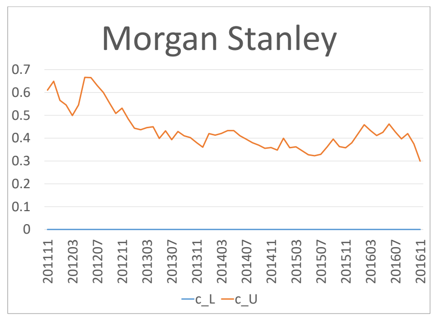
10.14754/CEU.2023.02

Figure 3.2: Bounds for c based on conditional probability estimates, Morgan Stanley analyst recommendation for Goldman stocks

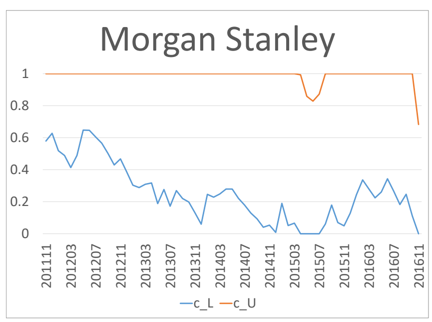
(a) hold=0



(b) hold=1



(c) hold=previous

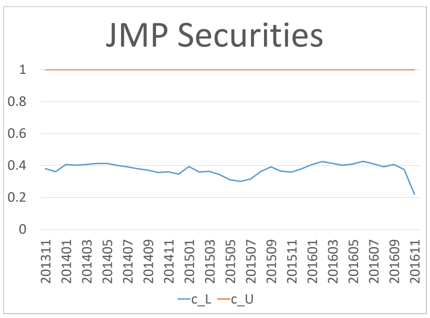


Bounds for c estimated using conditional logit regression. The forecast value is the binarized analyst recommendation for Goldman stocks made in t (strong buy, buy: 1; sell, strong sell: 0). Hold recommendations are categorized as 0 (a), 1 (b), and imputed by the previous value (c). The actual value is one if price growth for Goldman stocks is positive and outperforms the DJI one month from making the forecast, and zero otherwise. The explanatory variable Z is the smooth P/E ratio of Goldman Sachs in t.

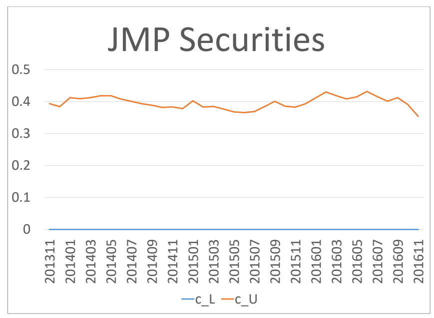
10.14754/CEU.2023.02

Figure 3.3: Bounds for c based on conditional probability estimates, JMP analyst recommendation for Goldman stocks

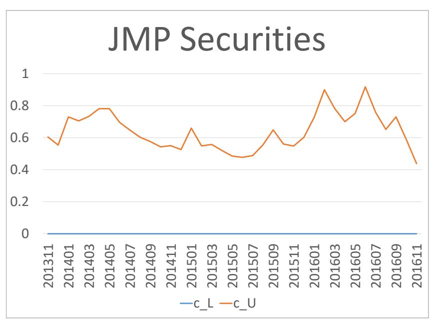
(a) hold=0



(b) hold=1



(c) hold=previous

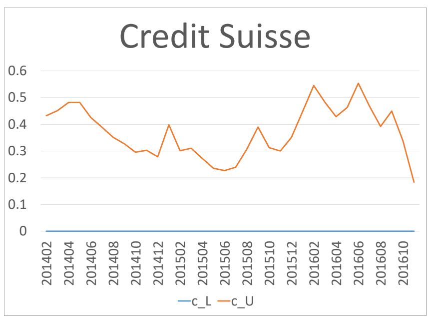


Bounds for c estimated using conditional logit regression. The forecast value is the binarized analyst recommendation for Goldman stocks made in t (strong buy, buy: 1; sell, strong sell: 0). Hold recommendations are categorized as 0 (a), 1 (b), and imputed by the previous value (c). The actual value is one if price growth for Goldman stocks is positive and outperforms the DJI one month from making the forecast, and zero otherwise. The explanatory variable Z is the smooth P/E ratio of Goldman Sachs in t.

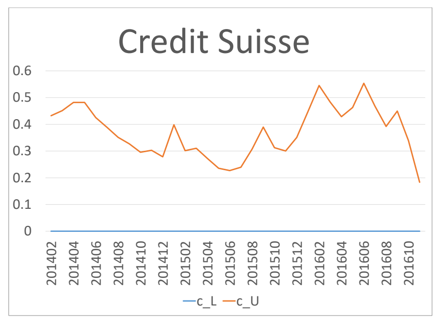
10.14754/CEU.2023.02

Figure 3.4: Bounds for c based on conditional probability estimates,Credit Suisse analyst recommendation for Goldman stocks

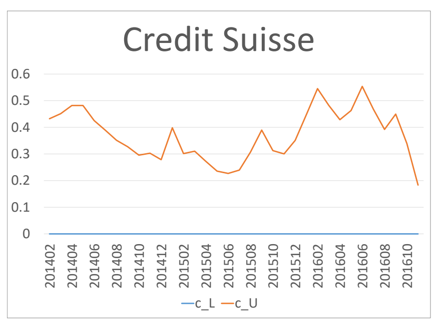
(a) hold=0



(b) hold=1



(c) hold=previous



Bounds for c estimated using conditional logit regression. The forecast value is the binarized analyst recommendation for Goldman stocks made in t (strong buy, buy: 1; sell, strong sell: 0). Hold recommendations are categorized as 0 (a), 1 (b), and imputed by the previous value (c). The actual value is one if price growth for Goldman stocks is positive and outperforms the DJI one month from making the forecast, and zero otherwise. The explanatory variable Z is the smooth P/E ratio of Goldman Sachs in t.

estimates, all six are under 0.5. Therefore, these estimates do not rule out symmetric loss. The four remaining analyst have relatively low asymmetry parameters. For Jefferies and Credit Suisse, we can rule out symmetric loss as the upper bound is below 0.5. The conditional bound intervals become narrower than the unconditional intervals.

In columns 3 and 4 in Table 2, we see that apart from RBC, the highest values are ruled out for c. This is similar to what I have found for Goldman stocks. The result is in line with the argument for low risk aversion towards buy strategies. Column 5 and 6 show the results for hold=previous. Here, in four of the cases the lowest values are ruled out, while in the other six cases $\hat{c}$ is relatively low.

We can see that the results are highly sensitive to the categorization of hold recommendations. If we take the hold=previous specification as baseline, we find that in the majority of cases, the highest values for c are ruled out.

	hold=0		hold=1		hold=prev	
analyst's firm	unconditional	conditional	unconditional	conditional	unconditional	conditional
Bernstein	[0.11, 1]	[0.33, 0.45]	[0, 0.38]	[0.39, 0.36]	[0, 0.41]	[0, 0.36]
RBC	[0.13, 1]	[0.31, 1]	[0.13, 1]	[0.31, 1]	[0.13, 1]	[0.31, 1]
Morgan Stanley	[0.15, 1]	[0.3, 0.4]	[0, 0.86]	[0.32, 0.34]	[0.05, 1]	[0.28, 0.4]
Barcleys	[0.31, 1]	[0.39, 1]	[0, 0.33]	[0, 0.29]	[0.31, 1]	[0.39, 1]
Goldman Sachs	[0.34, 1]	[0.48, 1]	[0, 0.35]	[0, 0.13]	[0.34, 1]	[0.48, 1]
William Blair	[0.39, 1]	[0.42, 0.65]	[0, 0.42]	[0, 0.4]	[0, 0.42]	[0, 0.4]
Edward Jones	[0, 0.5]	[0, 0.44]	[0, 0.46]	[0, 0.33]	[0, 0.46]	[0, 0.33]
Jefferies	[0, 0.32]	[0, 0.19]	[0, 0.32]	[0, 0.19]	[0, 0.32]	[0, 0.19]
Credit Suisse	[0, 0.33]	[0, 0.16]	[0, 0.29]	[0, 0.16]	[0, 0.33]	[0, 0.16]
J.P. Morgan	[0, 0.79]	[0, 0.61]	[0, 0.3]	[0, 0.26]	[0, 0.67]	[0, 0.52]

Table 3.2: Bounds for the asymmetry parameter, 3M Co. stocks

The forecast value is the binarized analyst recommendation for 3M Co. stocks made in t (strong buy, buy: 1; sell, strong sell: 0). Hold recommendations are categorized as 0 (column 1 and 2), 1 (column 3 and 4), and imputed by the previous value (columns 5 and 6). The actual value is one if price growth for 3M Co. stocks is positive and outperforms the DJI one month from making the forecast, and zero otherwise. The explanatory variable Z is the smooth P/E ratio of 3M Co. in t.

10.14754/CEU.2023.02

10.14754/CEU.2023.02

3.5. Conclusion

In a binary variable forecasting environment, I carry out an empirical analysis to estimate bounds for the parameter characterizing the forecaster's loss function. I use analyst stock recommendations as forecast data, and compare it to the one-month-ahead relative price performance of the analyzed stock. In the conditional logit regressions, I include a proxy for the publicly observed part of the forecaster's information set as an explanatory variable. Using a theoretical result from Lieli and Stinchcombe (2013), I set-identify the parameter that captures the analyst's cost of over- versus underpredicting the target (asymmetry parameter). Another novelty of this chapter is the derivation of confidence intervals for the bounds of the loss function asymmetry parameter introduced by Lieli and Stinchcombe (2013).

Previous research suggests that incorporating positive bias in stock analyst's forecasts is a rational action (Lim (2001)). It is also shown that controlling for accuracy, analysts who frequently issue optimistic forecasts are rewarded: they are much more likely to be offered higher prestige positions, with higher wages (Hong and Kubik (2003)). Therefore, we can expect analysts to issue overly optimistic forecasts more easily than pessimistic ones.

The reverse side of the argument can also be supported by intuitive claims. Consider that if an analyst issues a buy recommendation, then in the case of underperformance of the stock, her client will lose money for sure. However, if the analyst recommends a sell strategy, then her client might not even observe if the stock indeed outperforms the market. This suggests that analysts should avoid proposing overly optimistic recommendations.

I find that the results are highly sensitive to the categorization of hold recommendations. When we assume that 'hold' means 'sell', the estimated asymmetry parameters are relatively high. This suggests that analysts are not very reluctant to propose a 'sell'. However, when categorizing 'hold' into the buy category, the reverse is found: in almost all cases the highest possible values for the asymmetry parameter are ruled out. When imputing 'hold' with the previous recommendation, again the highest values are ruled out in more than half of the cases. Developing additional empirical applications (i.e. other binary forecasting problems) for the identification of the loss function's asymmetry parameter would be an interesting area for further research.

10.14754/CEU.2023.02

References

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

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance, 68*(3), 929-985.

Barber, B., Lehavy, R., McNichols, M., & Trueman, B. (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *The Journal of Finance, 56*(2), 531-563.

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

Bradshaw, M. T., Drake, M. S., Myers, J. N., & Myers, L. A. A re-examination of analysts' superiority over time-series forecasts of annual earnings. *Review of Accounting Studies, 17*, 944–968.

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

Brown, L. D., & Rozeff, M. S. (1978). The superiority of analyst forecasts as measures of expectations: Evidence from earnings. *The Journal of Finance, 33*(1), 1-16.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average?. *The Review of Financial Studies, 21*(4), 1509-1531.

Capistrán, C. (2008). Bias in Federal Reserve Inflation Forecasts: Is the Federal Reserve Irrational or Just Cautious?. *Journal of Monetary Economics, Volume 55*(8), 1415-1427.

Capistrán, C., & Timmermann, A. (2009). Disagreement and Biases in Inflation Expectations. *Journal of Money, Credit and Banking, 41* 365-396.

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

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

Christoffersen, P. F., & Diebold, F.X. (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.

Davies, A., & Lahiri, K. (1995). A New Framework for Analyzing Survey Forecasts using Three-Dimensional Panel Data. *Journal of Econometrics, 68*(1), 205-227.

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

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

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

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

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

Döpke, J., Fritsche, U., & Siliverstovs, B. (2010). Evaluating German Business Cycle

Forecasts under an Asymmetric Loss Function. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, vol. 2010/1.

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

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

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

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

Elliott, G., Komunjer, I., & Timmermann, A. (2008). Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss. *Journal of European Economic Association, 6*, 122–157.

Fissler, T. (2017). On Higher Order Elicitability and Some Limit Theorems on the Poisson and Wiener Space. Ph.D. thesis, University of Bern.

Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics, 44*(4), 1680–1707.

Franses, P. H. (2021). Testing bias in professional forecasts. *Journal of Forecasting, 40.* 1086– 1094.

Granger, C. (1969). Prediction with a Generalized Cost of Error Function. *Journal of the Operational Research Society, 20*, 199–207.

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

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

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

Green, T. C. (2006). The value of client access to analyst recommendations. *Journal of Financial and Quantitative Analysis, 41*(01), 1-24.

Grolmusz, V. M. (2016). Recovering Stock Analysts' Loss Functions from Buy/Sell Recommendations. *Spring Wind II. Conference Volume, 386-399.*

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

Hong, H., & Kubik, J. D. (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *The Journal of Finance, 58*(1), 313-351.

Hong, H., Kubik, J. D., & Solomon, A. (2000). Security analysts' career concerns and herding of earnings forecasts. *The Rand Journal of Economics*, 121-144.

Jegadeesh, N., Kim, J., Krische, S. D., & Lee, C. (2004). Analyzing the analysts: When do recommendations add value?. *The Journal of Finance, 59*(3), 1083-1124.

Lieli, R. P., & Stinchcombe, M. B. (2013). On the Recoverability of Forecasters' Preferences. *Econometric Theory, 29*(03), 517-544.

Lieli, R. P., Stinchcombe, M. B., & Grolmusz, V. M. (2019). Unrestricted and Controlled Identification of Loss Functions:
Possibility and Impossibility Results. *International Journal of Forecasting, 35*(03), 878–890.

Lim, T. (2001). Rationality and analysts' forecast bias. *The Journal of Finance, 56*(1), 369-385.

Loh, R. K., & Mian, G. M. (2006). Do accurate earnings forecasts facilitate superior investment recommendations?. *Journal of Financial Economics, 80*(2), 455-483.

de Mendonça, H. F., Garcia, P. M., & Vicente, J. V. M. (2021). Rationality and Anchoring of Inflation Expectations: An Assessment from Survey-Based and Market-Based Measures. *Journal of Forecasting, 40*, 1027– 1053.

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

Mishkin, F. S. (1981). Are Market Forecasts Rational?. *American Economic Review, 71*(3), 295-306.

Osband, K. H. (1985). Providing Incentives for Better Cost Forecasting. Unpublished Ph.D. thesis, University of California, Berkeley.

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

Pierdzioch, C., Rülke, J. C., & Stadtmann, G. (2012). On the loss function of the Bank of Canada: A note, *Economics Letters, 115*(2), 155-159.

Ramnath, S., Rock, S., & Shane, P. (2008). The financial analyst forecasting literature: A taxonomy with suggestions for further research. *International Journal of Forecasting, 24*(1), 34-75.

Rapoza, K. (2013, March 13). How Much Do Wall Streeters Really Earn?. *Forbes Investing*, Retrieved from http://www.forbes.com/sites/kenrapoza/2013/03/13/how-much-do-wall-streeters-really-earn/#3a92a01c7f08

Steinwart, I., Pasin, C., Williamson, R. C., Zhang, S. (2014). Elicitation and Identification of Properties. *JMLR: Workshop and Conference Proceedings, 35*, 1-45.

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

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

Stock, J.H., Watson, M.W. (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., Granger, C., & Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Elsevier.

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

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value?. *The Journal of Finance, 51*(1), 137-167.

Zarnowitz, V. (1985). Rational Expectations and Macroeconomic Forecasts. *Journal of Business & Economic Statistics, 3*(4), 293–311.

10.14754/CEU.2023.02

A Appendix for Chapter 1

Appendix A.1: Derivation of the general expected loss function using μ_e and σ_e

In this appendix, I show the derivation of the general expected loss function in equation 1.7 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:

$$E\{L(e_{t+1})|I_t, s_{t+1}\} = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} E\{((\alpha - (2\alpha-1)\mathbb{1}_{e_{s_{t+1}}>0})(e^2_{s_{t+1}}))|I_t\} \rightarrow \min \tag{3.1}$$

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:

$$E\{L(e_{t+1})|I_t, s_{t+1}\} = \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} E\{(\alpha - (2\alpha-1)\mathbb{1}_{e_{s_{t+1}}>0})[e^2_{s_{t+1}}]\} = \tag{3.2}$$

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

$$\underset{\downarrow}{\scriptstyle E(z^2_{s_{t+1}})=1;\ E(z_{s_{t+1}})=0}$$

$$= \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t}\ \alpha\ E[\mu_e^2 + \sigma_e^2] - (2\alpha-1) \sum_{s_{t+1}=1}^{k} \pi_{s_{t+1},t} \int_0^\infty e^2_{s_{t+1}}\ \mathrm{d}F(e_{s_{t+1}}) =$$

$$\underset{\downarrow}{\scriptstyle \text{changing variables in the integral}}$$

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

Appendix A.2: 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:

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

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

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

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

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

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

$$\left.+P_{12}\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}(-\omega_{0,s1}+\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}z)^2\mathrm{d}F(z)\right] \tag{3.3}$$

When the starting state is assumed to be s2, we get a similar expected loss function to equation 3.3, but the transition probabilities 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 3.3 yields the following first order conditions:

$$\frac{\partial E\{L(e)|s,I\}}{\partial\omega_0}=0:$$

$$\alpha\Big[P_{11}(\omega_{0,s1}+\omega_{1,s1}\mu_{f1,s1})+P_{12}(\omega_{0,s1})\Big]-$$

$$-(2\alpha-1)\left\{P_{11}\left[\int\limits_{\frac{\omega_{0,s1}+\omega_{1,s1}\,\mu_{f1,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}\omega_{0,s1}+\omega_{1,s1}\,\mu_{f1,s1}-\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}\,z\,\mathrm{d}F(z)\right]+\right.$$
$$\left.P_{12}\left[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}\omega_{0,s1}-\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}\,z\,\mathrm{d}F(z)\right]\right\} \tag{3.4}$$

The optimal weights for the two individual forecasts, ω_1 and ω_2 are determined by solving first order conditions 3.5 and 3.6:

$$\frac{\partial E\{L(e)|s,I\}}{\partial\omega_{1,s1}}=0:$$

$$\alpha\Big[P_{11}[(\omega_{0,s1}+\omega_{1,s1}\mu_{f1,s1})\mu_{f1,s1}+\omega_{1,s1}-1]+P_{12}(\omega_{1,s1}-1)\Big]-$$

$$-(2\alpha-1)\left\{P_{11}\left[\int\limits_{\frac{\omega_{0,s1}+\omega_{1,s1}\,\mu_{f1,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}c_{s1}\Big(-\mu_{f1,s1}+z\Big(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{1,s1}-1)\Big)\mathrm{d}F(z)\right]+\right.$$
$$\left.P_{12}\left[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}c_{s2}\Big(z\Big(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{1,s1}-1)\Big)\mathrm{d}F(z)\right]\right\} \tag{3.5}$$

where

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

$$\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{2,s1}} = 0:$$

$$\begin{aligned}
&\alpha[P_{11}(\omega_{2,s1}-1)+P_{12}(\omega_{2,s1}-1)]- \\
&-(2\alpha-1)\left\{P_{11}\left[\int\limits_{\frac{\omega_{0,s1}+\omega_{1,s1}\,\mu_{f1,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s1}\Big(z\Big(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{2,s1}-1)\Big)\mathrm{d}F(z)\right]+\right. \\
&\left. P_{12}\left[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty} c_{s2}\Big(z\Big(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{2,s1}-1)\Big)\mathrm{d}F(z)\right]\right\}
\end{aligned} \tag{3.6}$$

where

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

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 A.3: Equations for Special Case 2

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

$$
\begin{aligned}
&E\{L(e)|I,s1\} = \alpha[P_{11}(\mu_{e,s1}^2+\sigma_{e,s1}^2)+P_{12}(\mu_{e,s2}^2+\sigma_{e,s2}^2)]- \\
&-(2\alpha-1)\left[P_{1,1}\int\limits_{-\frac{\mu_{e,s1}}{\sigma_{e,s1}}}^{\infty}(\mu_{e,s1}+\sigma_{e,s1}z)^2\mathrm{d}F(z)+P_{1,2}\int\limits_{-\frac{\mu_{e,s2}}{\sigma_{e,s2}}}^{\infty}(\mu_{e,s2}+\sigma_{e,s2}z)^2\mathrm{d}F(z)\right]= \\
&=\alpha\left\{P_{11}[(-\omega_{0,s1})^2+(3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1}))]+\right. \\
&\qquad\left.+P_{12}[(-\omega_{0,s1})^2+(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1}))]\right\}- \\
&-(2\alpha-1)\left[P_{11}\int\limits_{\frac{\omega_{0,s1}}{\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}(-\omega_{0,s1}+\right. \\
&+\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}z\Big)^2\mathrm{d}F(z)+ \\
&\left.+P_{12}\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}(-\omega_{0,s1}+\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}z)^2\mathrm{d}F(z)\right]
\end{aligned}
\tag{3.7}
$$

When the starting state is assumed to be s2, we get a similar expected loss function to equation 3.7, 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 3.7 yields the following first order conditions:

$$\frac{\partial E\{L(e)|s,I\}}{\partial \omega_0}=0:$$

$$\begin{aligned}
&\alpha[P_{11}(\omega_{0,s1})+P_{12}(\omega_{0,s1})]-\\
&-(2\alpha-1)\Bigg\{P_{11}\Bigg[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}\omega_{0,s1}-\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}\,z\,\mathrm{d}F(z)\Bigg]+\\
&\qquad P_{12}\Bigg[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}\omega_{0,s1}-\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}\,z\,\mathrm{d}F(z)\Bigg]\Bigg\}
\end{aligned} \tag{3.8}$$

The optimal weights for the two individual forecasts, ω_1 and ω_2 are determined by solving first order conditions 3.5 and 3.9:

$$\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{1,s1}}=0:$$

$$\begin{aligned}
&\alpha[P_{11}[2\omega_{1,s1}-2]+P_{12}(\omega_{1,s1}-1)]-\\
&-(2\alpha-1)\Bigg\{P_{11}\Bigg[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}c_{s1}\Big(z\Big(3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(2\omega_{1,s1}-2)\Big)\mathrm{d}F(z)\Bigg]+\\
&\qquad P_{12}\Bigg[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}c_{s2}\Big(z\Big(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{1,s1}-1)\Big)\mathrm{d}F(z)\Bigg]\Bigg\}
\end{aligned} \tag{3.9}$$

where

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

$$\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{2,s1}}=0:$$

$$\alpha[P_{11}(\omega_{2,s1}-1)+P_{12}(\omega_{2,s1}-1)]-$$

$$-(2\alpha-1)\left\{P_{11}\left[\frac{\int\limits_{\omega_{0,s1}}^{\infty}}{\sqrt{3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})}}c_{s1}\Big(z\Big(3+(2\omega_{1,s1}^2+\omega_{2,s1}^2)-2(2\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{2,s1}-1)\Big)\mathrm{d}F(z)\right]+\right.$$

$$\left.P_{12}\left[\frac{\int\limits_{\omega_{0,s1}}^{\infty}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}c_{s2}\Big(z\Big(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{2,s1}-1)\Big)\mathrm{d}F(z)\right]\right\} \tag{3.10}$$

where

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

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.

10.14754/CEU.2023.02

Appendix A.4: Equations for Special Case 3

DGP:

$$y = f2 + e \tag{3.11}$$

forecast:

$$\hat{y} = \omega_0 + \omega_1 f1 + \omega_2 f2 \tag{3.12}$$

$$f1 = f + \epsilon f2 = f + \nu \tag{3.13}$$

$$Cov(f1, f2) = Var(f) + Cov(\epsilon, \nu) = 1 \tag{3.14}$$

The expected loss takes the following form:

$$\begin{aligned}
E\{L(e)|I, s1\} &= \alpha[P_{11}(\mu_{e,s1}^2 + \sigma_{e,s1}^2) + P_{12}(\mu_{e,s2}^2 + \sigma_{e,s2}^2)] - \\
&- (2\alpha - 1)\left[P_{1,1}\int_{-\frac{\mu_{e,s1}}{\sigma_{e,s1}}}^{\infty}(\mu_{e,s1} + \sigma_{e,s1}z)^2 \mathrm{d}F(z) + P_{1,2}\int_{-\frac{\mu_{e,s2}}{\sigma_{e,s2}}}^{\infty}(\mu_{e,s2} + \sigma_{e,s2}z)^2 \mathrm{d}F(z)\right] = \\
&= \alpha\left\{P_{11}[(-\omega_{0,s1})^2 + 1.2 + 2\omega_{1,s1}^2 + 1.2\omega_{2,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1}))] + \right. \\
&\quad \left. + P_{12}[(-\omega_{0,s1})^2 + (2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1}))]\right\} - \\
&- (2\alpha - 1)\left[P_{11}\int_{\frac{\omega_{0,s1}}{\sqrt{1.2 + 2\omega_{1,s1}^2 + 1.2\omega_{2,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1}}}}^{\infty}(-\omega_{0,s1} + \right. \\
&+ \sqrt{1.2 + 2\omega_{1,s1}^2 + 1.2\omega_{2,s1}^2 + 2\omega_{1,s1}\omega_{2,s1} - 2(\omega_{1,s1} + 1.2\omega_{2,s1}}z\Big)^2 \mathrm{d}F(z) + \\
&\left. + P_{12}\int_{\frac{\omega_{0,s1}}{\sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}}}^{\infty}\Big(-\omega_{0,s1} + \sqrt{2 + (\omega_{1,s1}^2 + \omega_{2,s1}^2) - 2(\omega_{1,s1} + \omega_{2,s1})}z)^2 \mathrm{d}F(z)\Big]\right]
\end{aligned} \tag{3.15}$$

When the starting state is assumed to be s2, we get a similar expected loss function to equation 3.15, 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 3.15 yields the following first order conditions:

$$\frac{\partial E\{L(e)|s,I\}}{\partial \omega_0}=0:$$

$$\begin{aligned}
&\alpha[P_{11}(\omega_{0,s1})+P_{12}(\omega_{0,s1})]-\\
&-(2\alpha-1)\Bigg\{P_{11}\Bigg[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{(1.2+2\omega_{1,s1}^2+1.2\omega_{2,s1}^2+2\omega_{1,s1}\omega_{2,s1}-2(\omega_{1,s1}+1.2\omega_{2,s1})}}}^{\infty}\omega_{0,s1}-\\
&-\sqrt{1.2+2\omega_{1,s1}^2+1.2\omega_{2,s1}^2+2\omega_{1,s1}\omega_{2,s1}-2(\omega_{1,s1}+1.2\omega_{2,s1})}\,z\,\mathrm{d}F(z)\Bigg]+\\
&P_{12}\Bigg[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}\omega_{0,s1}-\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}\,z\,\mathrm{d}F(z)\Bigg]\Bigg\}
\end{aligned} \tag{3.16}$$

The optimal weights for the two individual forecasts, ω_1 and ω_2 are determined by solving first order conditions 3.17 and 3.18:

$$\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{1,s1}}=0:$$

$$\begin{aligned}
&\alpha[P_{11}(2\omega_{1,s1}+\omega_{2,s1}-1)+P_{12}(\omega_{1,s1}-1)]-\\
&-(2\alpha-1)\Bigg\{P_{11}\Bigg[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{1.2+2\omega_{1,s1}^2+1.2\omega_{2,s1}^2+2\omega_{1,s1}\omega_{2,s1}-2(\omega_{1,s1}+1.2\omega_{2,s1})}}}^{\infty}c_{s1}\Big(z\Big(1.2+2\omega_{1,s1}^2+1.2\omega_{2,s1}^2+2\omega_{1,s1}\omega_{2,s1}-2(\omega_{1,s1}+1.2\omega_{2,s1})\Big)^{-1}\times\\
&\times(2\omega_{1,s1}+\omega_{2,s1}-1)\Big)\,\mathrm{d}F(z)\Bigg]+\\
&P_{12}\Bigg[\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}}^{\infty}c_{s2}\Big(z\Big(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{1,s1}-1)\Big)\mathrm{d}F(z)\Bigg]\Bigg\}
\end{aligned} \tag{3.17}$$

where

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

$$\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{2,s1}} = 0:$$

$$\alpha[P_{11}(\omega_{1,s1}+1.2\omega_{2,s1}-1.2)+P_{12}(\omega_{2,s1}-1)]-$$
$$-(2\alpha-1)\Bigg\{P_{11}\Bigg[\frac{\int_{\omega_{0,s1}}^{\infty}}{\sqrt{1.2+2\omega_{1,s1}^2+1.2\omega_{2,s1}^2+2\omega_{1,s1}\omega_{2,s1}-2(\omega_{1,s1}+1.2\omega_{2,s1})}} c_{s1}\Big(z\Big(1.2+2\omega_{1,s1}^2+1.2\omega_{2,s1}^2+2\omega_{1,s1}\omega_{2,s1}-2(\omega_{1,s1}+1.2\omega_{2,s1})\Big)^{-1}\times$$
$$\times(\omega_{1,s1}+1.2\omega_{2,s1}-1.2)\Big)\,\mathrm{d}F(z)\Bigg]+$$
$$P_{12}\Bigg[\frac{\int_{\omega_{0,s1}}^{\infty}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}} c_{s2}\Big(z\Big(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{2,s1}-1)\Big)\mathrm{d}F(z)\Bigg]\Bigg\} \tag{3.18}$$

where

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

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.

10.14754/CEU.2023.02

Appendix A.5: Equations for Special Case 4

DGP:

$$y = f1 + f2 + f3 + e \tag{3.19}$$

forecast:

$$\hat{y} = \omega_0 + \omega_1 f1 + \omega_2 f2 \tag{3.20}$$

$$\begin{aligned} f1 &= f3 + \epsilon \\ f2 &= f3 + \nu \end{aligned} \tag{3.21}$$

$$Cov(f1, f2) = Var(f3) + Cov(\epsilon, \nu) = 1 \tag{3.22}$$

$$\begin{aligned} Cov(y, f1) &= Var(f1) + Cov(f1, f2) + Cov(f1, f3) = 1.1 + 1 + 1 = 3.1 \\ Cov(y, f2) &= Var(f1) + Cov(f1, f2) + Cov(f2, f3) = 10 + 1 + 1 = 12 \end{aligned} \tag{3.23}$$

The expected loss takes the following form:

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

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

$$= \alpha\left\{P_{11}[(-\omega_{0,s1})^2 + (16.1 + (1.1\omega^2_{1,s1} + 2\omega_{1,s1}\omega_{2,s1} + 10\omega^2_{2,s1}) - 2(3.1\omega_{1,s1} + 12\omega_{2,s1}))] + \right.$$

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

$$-(2\alpha - 1)\left[P_{11}\int\limits_{\frac{\omega_{0,s1}}{\sqrt{16.1 + (1.1\omega^2_{1,s1} + 2\omega_{1,s1}\omega_{2,s1} + 10\omega^2_{2,s1}) - 2(3.1\omega_{1,s1} + 12\omega_{2,s1})}}}^{\infty}(-\omega_{0,s1} + \right.$$

$$+\sqrt{16.1 + (1.1\omega^2_{1,s1} + 2\omega_{1,s1}\omega_{2,s1} + 10\omega^2_{2,s1}) - 2(3.1\omega_{1,s1} + 12\omega_{2,s1})}z\Big)^2 \mathrm{d}F(z) +$$

$$\left. + P_{12}\int\limits_{\frac{\omega_{0,s1}}{\sqrt{2 + (\omega^2_{1,s1} + \omega^2_{2,s1}) - 2(\omega_{1,s1} + \omega_{2,s1})}}}^{\infty}\Big(-\omega_{0,s1} + \sqrt{2 + (\omega^2_{1,s1} + \omega^2_{2,s1}) - 2(\omega_{1,s1} + \omega_{2,s1})}z)^2 \mathrm{d}F(z)\Big]\right]$$

10.14754/CEU.2023.02

$$\tag{3.24}$$

When the starting state is assumed to be s2, we get a similar expected loss function to equation 3.24, 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 3.24 yields the following first order conditions:

$$\frac{\partial E\{L(e)|s,I\}}{\partial\omega_0}=0:$$

$$\begin{gathered}\alpha[P_{11}(\omega_{0,s1})+P_{12}(\omega_{0,s1})]- \\ -(2\alpha-1)\left\{P_{11}\left[\frac{\int\limits_{\omega_{0,s1}}^{\infty}}{\sqrt{(16.1+(1.1\omega_{1,s1}^2+2\omega_{1,s1}\omega_{2,s1}+10\omega_{2,s1}^2)-2(3.1\omega_{1,s1}+12\omega_{2,s1}))}}\omega_{0,s1}-\right.\right. \\ \left.-\sqrt{16.1+(1.1\omega_{1,s1}^2+2\omega_{1,s1}\omega_{2,s1}+10\omega_{2,s1}^2)-2(3.1\omega_{1,s1}+12\omega_{2,s1})}\,z\,\mathrm{d}F(z)\right]+ \\ \left.P_{12}\left[\frac{\int\limits_{\omega_{0,s1}}^{\infty}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}\omega_{0,s1}-\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}\,z\,\mathrm{d}F(z)\right]\right\}\end{gathered}\tag{3.25}$$

The optimal weights for the two individual forecasts, ω_1 and ω_2 are determined by solving first order conditions 3.26 and 3.27:

$$\frac{\partial E\{L(e)|s,I\}}{\partial\omega_{1,s1}}=0:$$

$$\begin{gathered}\alpha[P_{11}(1.1\omega_{1,s1}+\omega_{2,s1}-3.1)+P_{12}(\omega_{1,s1}-1)]- \\ -(2\alpha-1)\left\{P_{11}\left[\frac{\int\limits_{\omega_{0,s1}}^{\infty}}{\sqrt{16.1+(1.1\omega_{1,s1}^2+2\omega_{1,s1}\omega_{2,s1}+10\omega_{2,s1}^2)-2(3.1\omega_{1,s1}+12\omega_{2,s1})}}c_{s1}\Big(z\Big(16.1+(1.1\omega_{1,s1}^2+2\omega_{1,s1}\omega_{2,s1}+10\omega_{2,s1}^2)-\right.\right. \\ \left.-2(3.1\omega_{1,s1}+12\omega_{2,s1})\Big)^{-1}(1.1\omega_{1,s1}+\omega_{2,s1}-3.1)\Big)\,\mathrm{d}F(z)\right]+ \\ \left.P_{12}\left[\frac{\int\limits_{\omega_{0,s1}}^{\infty}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}}c_{s2}\Big(z\Big(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{1,s1}-1)\Big)\mathrm{d}F(z)\right]\right\}\end{gathered}\tag{3.26}$$

where

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

$$\frac{\partial E\{L(e)|s,I\}}{\partial \omega_{2,s1}} = 0:$$

$$\alpha[P_{11}(\omega_{1,s1}+10\omega_{2,s1}-12)+P_{12}(\omega_{2,s1}-1)]-$$
$$-(2\alpha-1)\Bigg\{P_{11}\Bigg[\frac{\int_{\omega_{0,s1}}^{\infty}}{\sqrt{16.1+(1.1\omega_{1,s1}^2+2\omega_{1,s1}\omega_{2,s1}+10\omega_{2,s1}^2)-2(3.1\omega_{1,s1}+12\omega_{2,s1})}} c_{s1}\Big(z\Big(16.1+(1.1\omega_{1,s1}^2+2\omega_{1,s1}\omega_{2,s1}+10\omega_{2,s1}^2)-$$
$$-2(3.1\omega_{1,s1}+12\omega_{2,s1})\Big)^{-1}(\omega_{1,s1}+10\omega_{2,s1}-12)\Big)\,\mathrm{d}F(z)\Bigg]+$$
$$P_{12}\Bigg[\frac{\int_{\omega_{0,s1}}^{\infty}}{\sqrt{2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})}} c_{s2}\Big(z\Big(2+(\omega_{1,s1}^2+\omega_{2,s1}^2)-2(\omega_{1,s1}+\omega_{2,s1})\Big)^{-1}(\omega_{2,s1}-1)\Big)\mathrm{d}F(z)\Bigg]\Bigg\} \quad (3.27)$$

where

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

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.

10.14754/CEU.2023.02

Appendix 6: Optimal weights from case 4: full tables

α	optimal weights		
	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 3.3: Optimal weights from case 4, only one state (s1)

α	optimal weights						ET optimal weights					
	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 3.4: Optimal weights from case 4, symmetric transition probabilities

α	optimal weights						ET optimal weights					
	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

Table 3.5: Optimal weights from case 4, asymmetric transition probabilities

10.14754/CEU.2023.02

B Appendix for Chapter 2

Appendix B.1. Generating observationally equivalent losses

This section gives a more detailed definition of Osband's principle (after Osband (1985)), and outlines the method on generating losses that lead to observationally equivalent forecasts. The section is based on subsection 4.1. in Lieli et al. (2019).

It follows from Osband's principle that it is not just scalar multiples of a loss function that are forecast equivalent to it, but one can generate a set of loss functions that are consistent for the same statistical functional as the original loss function, ℓ. Given an initial loss function, $\ell(\hat{y}, y)$, the idea is to generate observationally equivalent losses via the integral

$$\ell^{\dagger}(\hat{y}, y) := \int_a^{\hat{y}} \ell_{\hat{y}}(t, y) w(t) dt, \tag{3.28}$$

where $w(t) > 0$ is a continuously differentiable weight function. When $\ell(\cdot, y)$ is convex, so is $\int \ell(\cdot, y)\, dp(y)$, but $\ell^{\dagger}(\cdot, y)$ is need not be, unless $w(\cdot)$ satisfies further conditions (c.f. Fissler 2017, Ch. 4). The next set of arguments show that for any $w(\cdot)$, the first order condition $\frac{d}{d\hat{y}} \int \ell^{\dagger}(\cdot, y)\, dp(y) = 0$ has the same unique solution as the corresponding condition for ℓ, and the second order condition for a minimum is also satisfied at the unique solution.

- Integrating $\ell^{\dagger}$ with respect to a distribution $p(y)$ and interchanging the order of integration yields

$$\int \ell^{\dagger}(\hat{y}, y)\, dp(y) = \int_a^{\hat{y}} \left[\int \ell_{\hat{y}}(t, y)\, dp(y) \right] w(t) dt, \tag{3.29}$$

  and, therefore,

$$\frac{d}{d\hat{y}} \int \ell^{\dagger}(\hat{y}, y)\, dp(y) = w(\hat{y}) \int \ell_{\hat{y}}(\hat{y}, y)\, dp(y). \tag{3.30}$$

- Pulling the derivative out of the integral on the r.h.s. (see Lemma 2. in Appendix A od Lieli et al. (2019)) gives, for all $\hat{y} \in (a, b)$,

$$\frac{d}{d\hat{y}} \int \ell^{\dagger}(\hat{y}, y)\, dp(y) = w(\hat{y}) \frac{d}{d\hat{y}} \int \ell(\hat{y}, y)\, dp(y). \tag{3.31}$$

Equation (3.31) shows that the first order condition for the expected loss minimization of $\ell^{\dagger}$ is uniquely satisfied at the unique solution to the first order condition for the expected loss minimization of ℓ.

- To check that the second order condition holds, take the derivative in (3.31),

$$\tfrac{d}{d\hat{y}} w(\hat{y}) \int \ell_{\hat{y}}(\hat{y}, y)\, dp(y) = w'(\hat{y}) \int \ell_{\hat{y}}(\hat{y}, y)\, dp(y) + w(\hat{y}) \int \ell_{\hat{y},\hat{y}}(\hat{y}, y)\, dp(y).$$

When the first order condition holds, the first term is zero, and the second term is strictly positive.

By varying the weight function $w(\cdot)$, one can generate an entire class of forecast equivalent loss functions to ℓ, a class in which the first order conditions uniquely determine the unrestricted optimal forecast. For example, starting from square loss $(\hat{y} - y)^2$, Bregman losses can be generated by integrating $2w(t)(t - y)$. Working in a more general setting, Steinwart et al. (2014) demonstrate that all order-sensitive unrestrictedly forecast equivalent loss functions can actually be generated this way.

Appendix B.2. Table 2 of EKT (2005), included for comparison reasons

		IMF							OECD			
		Canada	France	Germany	Italy	Japan	UK	US	France	Germany	Italy	UK
						Current year						
Inst=1	alpha	0.60	0.52	0.40	0.16	0.24	0.24	0.20	0.27	0.22	0.29	0.48
	s.e	0.10	0.10	0.10	0.07	0.09	0.09	0.08	0.09	0.08	0.09	0.10
	p-value	0.31	0.84	0.31	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.85
Inst=2	alpha	0.58	0.54	0.40	0.14	19.00	0.20	0.18	0.28	0.22	0.28	0.50
	s.e	0.10	0.10	0.10	0.07	0.08	0.08	0.08	0.09	0.08	0.09	0.10
	p-value	0.39	0.67	0.33	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.98
Inst=3	alpha	0.59	0.54	0.42	0.15	0.24	0.24	0.19	0.13	0.12	0.29	0.50
	s.e	0.10	0.10	0.10	0.07	0.09	0.09	0.08	0.07	0.06	0.09	0.10
	p-value	0.39	0.68	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.99
Inst=4	alpha	0.59	0.54	0.40	0.13	0.20	0.19	0.17	0.11	0.11	0.26	0.49
	s.e	0.10	0.10	0.10	0.07	0.08	0.08	0.08	0.06	0.06	0.09	0.10
	p-value	0.39	0.67	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.92
						1-year ahead						
Inst=1	alpha	0.54	0.48	0.44	0.30	0.40	0.24	0.24	0.44	0.32	0.46	0.52
	s.e	0.10	0.10	0.10	0.10	0.10	0.09	0.09	0.10	0.09	0.10	0.10
	p-value	0.68	0.84	0.55	0.04	0.31	0.00	0.00	0.55	0.05	0.68	0.84
Inst=2	alpha	0.54	0.50	0.45	0.27	0.18	0.14	0.14	0.41	0.33	0.43	0.55
	s.e	0.10	0.10	0.10	0.09	0.08	0.07	0.07	0.10	0.09	0.10	0.10
	p-value	0.67	0.96	0.62	0.01	0.00	0.00	0.00	0.38	0.08	0.52	0.59
Inst=3	alpha	0.57	0.50	0.46	0.27	0.37	0.24	0.24	0.39	0.27	0.43	0.54
	s.e	0.10	0.10	0.10	0.09	0.10	0.09	0.09	0.10	0.09	0.10	0.10
	p-value	0.50	1.00	0.66	0.01	0.18	0.00	0.00	0.24	0.01	0.52	0.66
Inst=4	alpha	0.57	0.50	0.44	0.27	0.26	0.17	0.13	0.35	0.24	0.43	0.57
	s.e	0.10	0.10	0.10	0.09	0.09	0.08	0.07	0.10	0.09	0.10	0.10
	p-value	0.48	0.98	0.57	0.01	0.01	0.00	0.00	0.13	0.00	0.51	0.45

Note: Table 2 of EKT (2005), included for comparison reasons. Sample period: 1975-2001; The four instrument sets are the following from inst=1 to inst=4: (i) constant; (ii) constant, lagged forecast error; (iii) constant, lagged budget deficit; (iv) constant, lagged forecast error and lagged budget deficit.

Table 3.6: EKT's Estimated α parameters for $b = 1$

Appendix B.3. Estimated α parameters for $b = 0.25$

		IMF							OECD				
		Canada	France	Germany	Italy	Japan	UK	US	US	France	Germany	Italy	UK
		Current year											
Inst=1	alpha	0.34	0.47	0.51	0.38	0.37	0.22	0.24	0.31	0.47	0.40	0.38	0.35
	s.e	0.14	0.12	0.13	0.12	0.12	0.09	0.13	0.11	0.11	0.11	0.12	0.11
	p-value	0,23	0,77	0,95	0,34	0,26	0,00	0,04	0,09	0,76	0,35	0,32	0,19
Inst=2	alpha	0.35	0.49	0.51	0.37	0.20	0.20	0.30	0.35	0.49	0.37	0.36	0.39
	s.e	0.13	0.12	0.12	0.12	0.09	0.09	0.13	0.11	0.11	0.11	0.12	0.11
	p-value	0,28	0,92	0,94	0,27	0,00	0,00	0,12	0,17	0,94	0,25	0,21	0,32
Inst=3	alpha	0.35	0.46	0.41	0.43	0.28	0.22	0.26	0.30	0.39	0.28	0.34	0.34
	s.e	0.14	0.11	0.12	0.11	0.10	0.09	0.12	0.10	0.10	0.10	0.09	0.10
	p-value	0,26	0,71	0,44	0,54	0,03	0,00	0,05	0,05	0,29	0,02	0,10	0,11
Inst=4	alpha	0.36	0.46	0.41	0.42	0.18	0.22	0.27	0.29	0.33	0.28	0.34	0.35
	s.e	0.13	0.11	0.12	0.11	0.08	0.09	0.12	0.10	0.10	0.10	0.09	0.10
	p-value	0,29	0,71	0,44	0,49	0,00	0,00	0,06	0,04	0,10	0,02	0,10	0,12
		1-year ahead											
Inst=1	alpha	0.37	0.41	0.57	0.44	0.48	0.28	0.34	0.38	0.64	0.55	0.50	0.53
	s.e	0.13	0.14	0.16	0.15	0.21	0.14	0.13	0.13	0.13	0.11	0.13	0.13
	p-value	0,31	0,51	0,63	0,71	0,93	0,11	0,20	0,35	0,29	0,63	0,98	0,84
Inst=2	alpha	0.36	0.43	0.58	0.38	0.94	0.07	0.41	0.53	0.61	0.56	0.43	0.32
	s.e	0.13	0.14	0.16	0.14	0.05	0.07	0.13	0.11	0.14	0.11	0.12	0.11
	p-value	0,28	0,64	0,59	0,37	0,00	0,00	0,52	0,82	0,44	0,59	0,60	0,09
Inst=3	alpha	0.34	0.43	0.65	0.75	1.00	0.28	0.34	0.37	0.30	0.54	0.52	0.45
	s.e	0.13	0.14	0.14	0.11	0.13	0.14	0.13	0.12	0.17	0.11	0.11	0.11
	p-value	0,22	0,65	0,27	0,03	0,00	0,12	0,20	0,28	0,26	0,72	0,88	0,62
Inst=4	alpha	0.36	0.44	0.64	0.80	0.06	0.04	0.36	0.50	0.35	0.54	0.51	0.42
	s.e	0.13	0.14	0.14	0.10	0.07	0.06	0.13	0.11	0.16	0.11	0.11	0.11
	p-value	0,28	0,65	0,29	0,00	0,00	0,00	0,26	0,97	0,34	0,74	0,94	0,47

Note: Sample period: 1980-2017; The four instrument sets are based on EKT and are the following from inst=1 to inst=4: (i) constant; (ii) constant, lagged forecast error; (iii) constant, lagged budget deficit; (iv) constant, lagged forecast error and lagged budget deficit. The GMM weighting matrix $\hat{S}$ is specified as in EKT. p values refer to the null hypothesis of $\alpha = 0.5$.

Table 3.7: Estimated α parameters for $b = 0.25$

10.14754/CEU.2023.02

Appendix B.4. Estimated α parameters for $b = 0.5$

		IMF							OECD				
		Canada	France	Germany	Italy	Japan	UK	US	US	France	Germany	Italy	UK
Current year													
Inst=1	alpha	0.39	0.45	0.48	0.40	0.37	0.28	0.28	0.36	0.43	0.38	0.36	0.38
	s.e	0.13	0.11	0.12	0.11	0.11	0.10	0.12	0.10	0.10	0.10	0.10	0.10
	p-value	0,41	0,68	0,86	0,39	0,26	0,03	0,07	0,18	0,52	0,22	0,19	0,25
Inst=2	alpha	0.39	0.47	0.47	0.40	0.24	0.26	0.31	0.38	0.46	0.35	0.37	0.40
	s.e	0.13	0.11	0.12	0.11	0.10	0.10	0.12	0.10	0.11	0.10	0.10	0.10
	p-value	0,37	0,71	0,45	0,57	0,05	0,02	0,06	0,10	0,25	0,01	0,12	0,21
Inst=3	alpha	0.39	0.46	0.41	0.44	0.29	0.27	0.28	0.34	0.39	0.27	0.35	0.38
	s.e	0.13	0.11	0.11	0.11	0.10	0.10	0.12	0.10	0.10	0.09	0.10	0.10
	p-value	0.19	0.35	0.22	0.29	0.02	0.01	0.03	0.05	0.13	0.01	0.06	0.10
Inst=4	alpha	0.39	0.46	0.41	0.43	0.19	0.27	0.27	0.34	0.36	0.27	0.35	0.38
	s.e	0.13	0.11	0.11	0.11	0.09	0.10	0.12	0.10	0.10	0.09	0.10	0.10
	p-value	0,39	0,70	0,43	0,53	0,00	0,01	0,05	0,10	0,14	0,01	0,12	0,20
1-year ahead													
Inst=1	alpha	0.38	0.38	0.50	0.45	0.56	0.32	0.36	0.42	0.54	0.51	0.51	0.51
	s.e	0.13	0.13	0.14	0.14	0.17	0.13	0.13	0.11	0.12	0.10	0.11	0.11
	p-value	0,35	0,38	0,97	0,73	0,73	0,17	0,28	0,44	0,75	0,96	0,90	0,93
Inst=2	alpha	0.35	0.40	0.50	0.43	0.94	0.16	0.41	0.48	0.51	0.52	0.48	0.43
	s.e	0.13	0.14	0.15	0.14	0.05	0.10	0.13	0.11	0.12	0.10	0.11	0.11
	p-value	0,27	0,48	0,99	0,63	0,00	0,00	0,48	0,83	0,94	0,88	0,84	0,52
Inst=3	alpha	0.34	0.41	0.57	0.72	1.00	0.33	0.37	0.41	0.37	0.50	0.53	0.48
	s.e	0.13	0.14	0.14	0.12	0.11	0.14	0.13	0.11	0.12	0.10	0.11	0.10
	p-value	0,22	0,50	0,60	0,06	0,00	0,21	0,30	0,38	0,28	0,98	0,79	0,82
Inst=4	alpha	0.35	0.40	0.56	0.77	1.00	0.10	0.36	0.45	0.38	0.50	0.52	0.47
	s.e	0.13	0.14	0.14	0.11	0.00	0.09	0.13	0.11	0.12	0.10	0.11	0.10
	p-value	0,24	0,48	0,68	0,01	0,00	0,00	0,28	0,65	0,32	1,00	0,82	0,80

Note: Sample period: 1980-2017; The four instrument sets are based on EKT and are the following from inst=1 to inst=4: (i) constant; (ii) constant, lagged forecast error; (iii) constant, lagged budget deficit; (iv) constant, lagged forecast error and lagged budget deficit. The GMM weighting matrix $\hat{S}$ is specified as in EKT. p values refer to the null hypothesis of $\alpha = 0.5$.

Table 3.8: Estimated α parameters for $b = 0.5$

10.14754/CEU.2023.02

Appendix B.5. Estimated α parameters for $b = 2$

		IMF							OECD				
		Canada	France	Germany	Italy	Japan	UK	US	US	France	Germany	Italy	UK
Current year													
Inst=1	alpha	0.53	0.39	0.36	0.44	0.36	0.39	0.21	0.39	0.35	0.27	0.34	0.38
	s.e	0.13	0.11	0.11	0.12	0.12	0.13	0.10	0.10	0.11	0.09	0.10	0.11
	p-value	0.81	0.32	0.23	0.58	0.24	0.37	0.00	0.28	0.18	0.02	0.11	0.25
Inst=2	alpha	0.44	0.41	0.30	0.47	0.26	0.36	0.22	0.43	0.31	0.21	0.35	0.42
	s.e	0.13	0.12	0.11	0.12	0.10	0.12	0.10	0.11	0.10	0.08	0.10	0.11
	p-value	0.67	0.43	0.07	0.80	0.02	0.26	0.00	0.52	0.07	0.00	0.15	0.47
Inst=3	alpha	0.46	0.40	0.28	0.48	0.21	0.39	0.13	0.45	0.37	0.21	0.34	0.44
	s.e	0.13	0.11	0.10	0.12	0.09	0.13	0.08	0.11	0.10	0.08	0.10	0.11
	p-value	0.74	0.36	0.03	0.89	0.00	0.37	0.00	0.66	0.19	0.00	0.10	0.54
Inst=4	alpha	0.45	0.40	0.27	0.48	0.11	0.41	0.07	0.45	0.33	0.21	0.34	0.44
	s.e	0.13	0.11	0.10	0.12	0.07	0.12	0.06	0.11	0.09	0.08	0.10	0.11
	p-value	0.68	0.38	0.02	0.85	0.00	0.45	0.00	0.66	0.07	0.00	0.10	0.56
1-year ahead													
Inst=1	alpha	0.40	0.26	0.29	0.47	0.79	0.48	0.40	0.32	0.27	0.30	0.56	0.48
	s.e	0.14	0.12	0.12	0.14	0.10	0.17	0.13	0.10	0.09	0.09	0.11	0.12
	p-value	0.47	0.05	0.08	0.83	0.00	0.90	0.46	0.07	0.02	0.03	0.59	0.83
Inst=2	alpha	0.27	0.28	0.20	0.54	0.98	0.50	0.44	0.36	0.26	0.33	0.55	0.49
	s.e	0.12	0.13	0.11	0.14	0.03	0.17	0.14	0.10	0.09	0.10	0.11	0.11
	p-value	0.04	0.09	0.01	0.78	0.00	0.98	0.65	0.18	0.01	0.08	0.68	0.96
Inst=3	alpha	0.06	0.28	0.21	0.86	0.97	0.50	0.41	0.34	0.27	0.33	0.53	0.57
	s.e	0.06	0.13	0.10	0.11	0.04	0.17	0.14	0.10	0.09	0.10	0.11	0.11
	p-value	0.00	0.08	0.00	0.00	0.00	0.99	0.53	0.12	0.02	0.09	0.82	0.54
Inst=4	alpha	0.06	0.28	0.07	0.46	1.00	0.44	0.43	0.35	0.27	0.34	0.53	0.52
	s.e	0.06	0.13	0.05	0.14	0.00	0.16	0.14	0.10	0.09	0.10	0.11	0.11
	p-value	0.00	0.09	0.00	0.80	0.00	0.70	0.61	0.13	0.01	0.09	0.80	0.88

Note: Sample period: 1980-2017; The four instrument sets are based on EKT and are the following from inst=1 to inst=4: (i) constant; (ii) constant, lagged forecast error; (iii) constant, lagged budget deficit; (iv) constant, lagged forecast error and lagged budget deficit. The GMM weighting matrix $\hat{S}$ is specified as in EKT. p values refer to the null hypothesis of $\alpha = 0.5$.

Table 3.9: Estimated α parameters for $b = 2$

10.14754/CEU.2023.02

Appendix B.6. Estimated α parameters for $b = 3$

		IMF							OECD				
		Canada	France	Germany	Italy	Japan	UK	US	US	France	Germany	Italy	UK
Current year													
Inst=1	alpha	0.55	0.35	0.32	0.42	0.32	0.29	0.13	0.34	0.40	0.26	0.34	0.32
	s.e	0.14	0.13	0.12	0.12	0.15	0.14	0.08	0.12	0.15	0.12	0.11	0.13
	p-value	0.71	0.25	0.14	0.52	0.22	0.13	0.00	0.17	0.49	0.04	0.14	0.15
Inst=2	alpha	0.44	0.36	0.25	0.46	0.17	0.27	0.12	0.41	0.31	0.15	0.35	0.39
	s.e	0.14	0.14	0.11	0.13	0.08	0.13	0.06	0.12	0.14	0.07	0.11	0.14
	p-value	0.65	0.31	0.02	0.77	0.00	0.09	0.00	0.45	0.16	0.00	0.17	0.43
Inst=3	alpha	0.47	0.33	0.18	0.49	0.07	0.17	0.02	0.43	0.36	0.16	0.33	0.40
	s.e	0.14	0.11	0.09	0.12	0.05	0.11	0.05	0.12	0.12	0.07	0.10	0.12
	p-value	0.85	0.14	0.00	0.94	0.00	0.00	0.00	0.59	0.25	0.00	0.10	0.42
Inst=4	alpha	0.46	0.36	0.18	0.49	0.03	0.04	0.02	0.43	0.26	0.16	0.33	0.41
	s.e	0.14	0.11	0.09	0.12	0.04	0.07	0.03	0.12	0.09	0.07	0.10	0.12
	p-value	0.75	0.23	0.00	0.92	0.00	0.00	0.00	0.59	0.01	0.00	0.10	0.47
1-year ahead													
Inst=1	alpha	0.37	0.22	0.22	0.44	0.89	0.49	0.36	0.21	0.21	0.21	0.54	0.42
	s.e	0.15	0.15	0.12	0.15	0.07	0.22	0.14	0.09	0.10	0.09	0.12	0.15
	p-value	0.39	0.06	0.02	0.69	0.00	0.95	0.31	0.00	0.00	0.00	0.74	0.58
Inst=2	alpha	0.20	0.15	0.11	0.56	0.99	0.50	0.42	0.26	0.21	0.26	0.53	0.43
	s.e	0.10	0.12	0.08	0.15	0.02	0.21	0.14	0.10	0.10	0.09	0.12	0.14
	p-value	0.00	0.00	0.00	0.68	0.00	1.00	0.60	0.01	0.00	0.01	0.81	0.62
Inst=3	alpha	0.02	0.24	0.12	0.89	0.99	0.52	0.31	0.21	0.22	0.26	0.52	0.58
	s.e	0.04	0.14	0.08	0.11	0.02	0.23	0.14	0.09	0.09	0.10	0.12	0.15
	p-value	0.00	0.07	0.00	0.00	0.00	0.93	0.18	0.00	0.00	0.01	0.87	0.59
Inst=4	alpha	0.03	0.20	0.03	0.89	1.00	0.43	0.39	0.21	0.22	0.26	0.52	0.46
	s.e	0.04	0.13	0.03	0.11	0.00	0.19	0.14	0.09	0.09	0.09	0.12	0.13
	p-value	0.00	0.03	0.00	0.00	0.00	0.69	0.42	0.00	0.00	0.01	0.85	0.76

Note: Sample period: 1980-2017; The four instrument sets are based on EKT and are the following from inst=1 to inst=4: (i) constant; (ii) constant, lagged forecast error; (iii) constant, lagged budget deficit; (iv) constant, lagged forecast error and lagged budget deficit. The GMM weighting matrix $\hat{S}$ is specified as in EKT. p values refer to the null hypothesis of $\alpha = 0.5$.

Table 3.10: Estimated α parameters for $b = 3$

Appendix B.7. Further empirical examples: losses of different parametric families forecast equivalent to quad-quad

In this section, I am conducting a similar empirical exercise to that in subsection 2.4.1 and 2.4.2, however I am using different functional forms for the loss function. The loss functions and adherent moment conditions are the following:

1. $\mathbf{w(t) = |t|}$:

i $y < \hat{y}$

$$2(1-\alpha)\int_{y}^{\hat{y}} |t|(t-y)\,\mathrm{d}t = 2(1-\alpha)\left[\frac{1}{6}t^2 sgn(t)(2t-3y)\right]_y^{\hat{y}} = \frac{1}{3}(1-\alpha)\left[\hat{y}^2 sgn(\hat{y})(2\hat{y}-3y)+y^3 sgn(y)\right]$$

ii $y \geq \hat{y}$

$$-2\alpha\int_{y}^{\hat{y}} |t|(t-y)\,\mathrm{d}t = \frac{\alpha}{3}\left[\hat{y}^2 sgn(\hat{y})(2\hat{y}-3y)+y^3 sgn(y)\right]$$

2. $\mathbf{w(t) = t^2}$:

i $y < \hat{y}$

$$2(1-\alpha)\int_{y}^{\hat{y}} t^2(t-y)\,\mathrm{d}t = 2(1-\alpha)\left[\frac{t^4}{4} - \frac{t^3 y}{3}\right]_y^{\hat{y}} = 2(1-\alpha)\left[\frac{1}{4}(\hat{y}^4 - y^4) - \frac{y}{3}(\hat{y}^3 - y^3)\right]$$

ii $y \geq \hat{y}$

$$-2\alpha\int_{y}^{\hat{y}} t^2(t-y)\,\mathrm{d}t = -2\alpha\left[\frac{t^4}{4} - \frac{t^3 y}{3}\right]_{\hat{y}}^{y} = -2\alpha\left[\frac{1}{4}(y^4 - \hat{y}^4) - \frac{\hat{y}}{3}(y^3 - \hat{y}^3)\right]$$

3. $\mathbf{w(t) = |t|^3}$:

i $y < \hat{y}$

$$2(1-\alpha)\int_{y}^{\hat{y}} |t|^3(t-y)\,\mathrm{d}t = 2(1-\alpha)\left[\frac{1}{20}t^4 sgn(t)(4t-5y)\right]_y^{\hat{y}} = \frac{1}{10}(1-\alpha)\left[\hat{y}^4 sgn(\hat{y})(4\hat{y}-5y)+y^5 sgn(y)\right]$$

ii $y \geq \hat{y}$

$$-2\alpha\int_{y}^{\hat{y}} |t|^3(t-y)\,\mathrm{d}t = (-2\alpha)\left[\frac{1}{20}t^4 sgn(t)(4t-5y)\right]_{\hat{y}}^{y} = \frac{\alpha}{10}\left[\hat{y}^4 sgn(\hat{y})(4\hat{y}-5y)+y^5 sgn(y)\right]$$

4. $\mathbf{w(t) = t^4}$:

i $y < \hat{y}$

$$2(1-\alpha)\int_y^{\hat{y}} t^4(t-y)\,\mathrm{d}t = 2(1-\alpha)\left[\frac{t^6}{6} - \frac{t^5 y}{5}\right]_y^{\hat{y}} = 2(1-\alpha)\left[\frac{1}{6}(\hat{y}^6 - y^6) - \frac{y}{5}(\hat{y}^5 - y^5)\right]$$

ii $y \geq \hat{y}$

$$-2\alpha\int_y^{\hat{y}} t^4(t-y)\,\mathrm{d}t = -2\alpha\left[\frac{t^4}{4} - \frac{t^3 y}{3}\right]_{\hat{y}}^{y} = -2\alpha\left[\frac{1}{6}(y^6 - \hat{y}^6) - \frac{y}{5}(y^5 - \hat{y}^5)\right]$$

5. $\mathbf{w(t) = e^t}$:

i $y < \hat{y}$

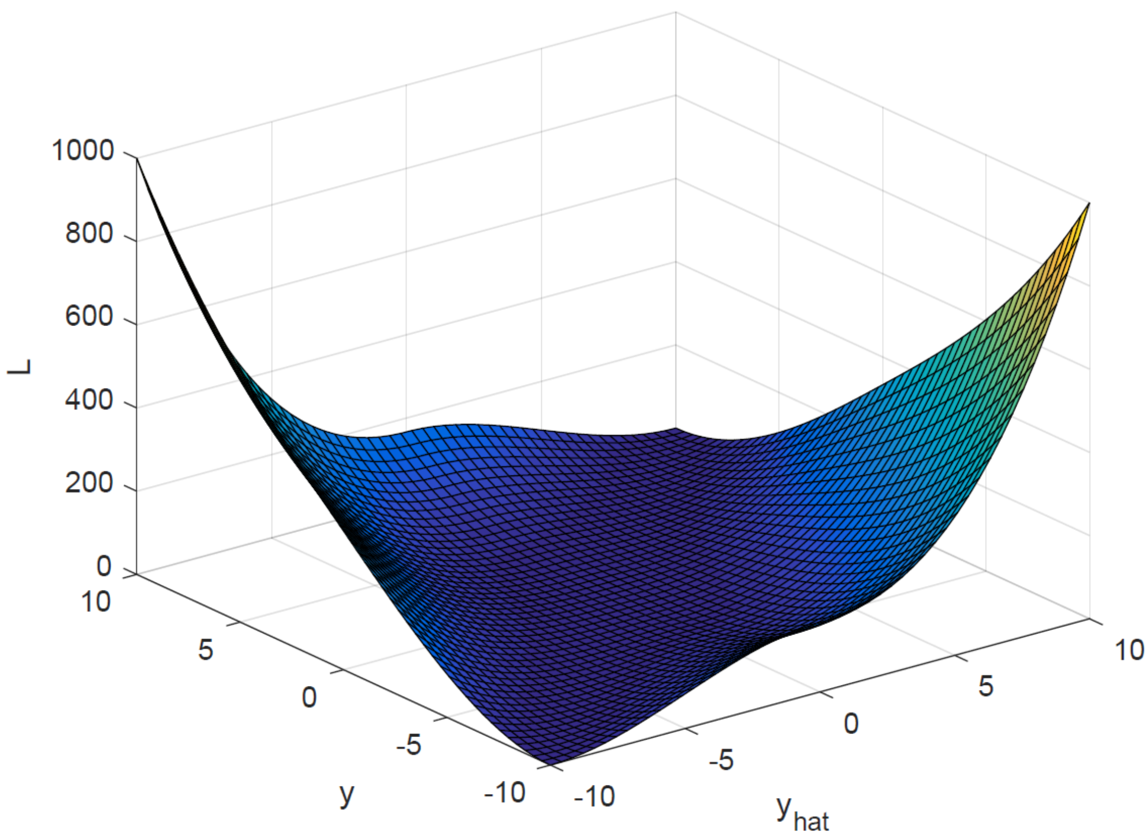
$$2(1-\alpha)\int_y^{\hat{y}} e^t(t-y)\,\mathrm{d}t = 2(1-\alpha)\left[e^t(t-y-1)\right]_y^{\hat{y}} = 2(1-\alpha)\left[e^{\hat{y}}(\hat{y}-y-1) + e^y\right]$$

ii $y \geq \hat{y}$

$$-2\alpha\int_y^{\hat{y}} e^t(t-y)\,\mathrm{d}t = -2\alpha\left[e^t(t-y-1)\right]_{\hat{y}}^{y} = 2\alpha\left[e^{\hat{y}}(\hat{y}-y-1) + e^y\right]$$

10.14754/CEU.2023.02

Figure 3.5: Loss function of y and $\hat{y}$, $w(t) = |t|$, alpha=0.5



Using different functional forms for the loss function results in a diverse set of estimated α parameters. Table 3.11 includes the results for the loss function where $w(t) = |t|$. As an illustration, this loss function is plotted in figure 3.5. The results show a limited rejection rate of the symmetry of $\hat{\alpha}$. For the end-of-year IMF forecasts, symmetry can be rejected for Japan's (Inst=2; 3; 4) and the United States' (Inst=3;4) estimates. This is in line with the results from the GLS loss functions under different b's (see tables 2.2 and 3.7 through 3.10), where symmetry was also rejected in many cases for Japan's and the United States' end-of-year IMF estimates. In the OECD end-of-year results, we can reject the symmetry of $\hat{\alpha}$ for Germany regardless of the instrument set used. For the next-year forecasts, rejection rates of the $w(t) = |t|$ case are similar to the GLS-results for the $b = 0.25$ case in table 3.7.

Different functional forms result in different rejection rates: while setting the weight function to $|t|$ and t^2 gives lower average rejection rates (0.16 and 0.22, respectively), choosing a higher order for t (t^3 and t^4) or the exponential function ($w(t) = exp(t)$) as

		IMF							OECD				
		Canada	France	Germany	Italy	Japan	UK	US	US	France	Germany	Italy	UK
		Current year											
Inst=1	alpha	0.65	0.45	0.37	0.34	0.27	0.51	0.22	0.46	0.31	0.18	0.31	0.38
	s.e	0.45	1.56	0.47	0.50	0.25	5.38	0.21	1.65	0.33	0.13	0.37	0.59
	p-value	0.74	0.98	0.79	0.76	0.36	1.00	0.18	0.98	0.57	0.01	0.61	0.84
Inst=2	alpha	0.60	0.52	0.32	0.25	0.12	0.48	0.23	0.50	0.26	0.12	0.29	0.47
	s.e	0.23	0.90	0.32	0.23	0.10	1.51	0.21	0.48	0.20	0.09	0.27	1.77
	p-value	0.67	0.98	0.57	0.28	0.00	0.99	0.19	0.99	0.24	0.00	0.43	0.99
Inst=3	alpha	0.58	0.48	0.30	0.42	0.13	0.50	0.15	0.46	0.30	0.11	0.31	0.45
	s.e	0.30	0.83	0.25	0.63	0.11	0.93	0.14	0.33	0.26	0.08	0.35	0.50
	p-value	0.78	0.98	0.42	0.89	0.00	1.00	0.01	0.91	0.42	0.00	0.59	0.93
Inst=4	alpha	0.59	0.51	0.29	0.22	0.02	0.50	0.06	0.46	0.20	0.11	0.25	0.45
	s.e	0.24	0.61	0.25	0.17	0.04	0.89	0.06	0.33	0.14	0.08	0.20	0.50
	p-value	0.70	0.99	0.42	0.11	0.00	1.00	0.00	0.91	0.04	0.00	0.22	0.92
		1-year ahead											
Inst=1	alpha	0.59	0.63	0.38	0.24	0.82	0.56	0.51	0.38	0.45	0.17	0.45	0.32
	s.e	0.14	0.46	0.16	0.18	1.86	0.61	0.17	0.15	0.27	0.11	0.50	0.24
	p-value	0.51	0.77	0.47	0.16	0.86	0.92	0.96	0.39	0.84	0.00	0.92	0.46
Inst=2	alpha	0.28	0.71	0.49	0.41	0.99	0.87	0.44	0.49	0.62	0.26	0.39	0.40
	s.e	0.23	0.26	2.02	0.19	0.04	0.23	0.24	0.37	0.30	0.22	0.34	0.31
	p-value	0.33	0.41	1.00	0.64	0.00	0.11	0.81	0.98	0.69	0.26	0.76	0.74
Inst=3	alpha	0.57	0.57	0.29	0.25	0.98	0.53	0.46	0.56	0.48	0.17	0.40	0.54
	s.e	0.19	0.26	0.42	0.19	0.05	0.25	0.24	0.25	0.31	0.20	0.54	0.23
	p-value	0.71	0.78	0.62	0.21	0.00	0.91	0.86	0.81	0.96	0.10	0.85	0.86
Inst=4	alpha	0.24	0.49	0.49	0.25	1.00	0.46	0.38	0.60	0.58	0.20	0.41	0.48
	s.e	0.11	0.28	0.42	0.17	0.00	0.26	0.23	0.26	0.23	0.20	0.25	0.25
	p-value	0.02	0.96	0.97	0.14	0.00	0.88	0.61	0.72	0.72	0.13	0.72	0.94

Note: Sample period: 1980-2017; The four instrument sets are based on EKT and are the following from inst=1 to inst=4: (i) constant; (ii) constant, lagged forecast error; (iii) constant, lagged budget deficit; (iv) constant, lagged forecast error and lagged budget deficit. $w(t) = |t|$

Table 3.11: Estimated α parameters for $w(t) = |t|$

weight functions yields higher rejection rates (see table 3.12).

For some countries, the estimated symmetry parameters are quite robust to the change in the loss' functional form. Canada is such a country: we can see from table 3.13 that the end-of-year α-estimates are close to 0.6 for all, except for the exponential, weighting functions.

w(t)	|t|	t^2	t^3	t^4	exp(t)	$\sum$
IMF end-of-year	0.18	0.25	0.43	0.46	0.36	0.34
IMF 1-year ahead	0.14	0.21	0.39	0.43	0.46	0.33
IMF $\sum$	0.16	0.23	0.41	0.45	0.41	0.33
OECD end-of-year	0.25	0.20	0.30	0.40	0.15	0.26
OECD 1-year ahead	0.05	0.20	0.30	0.30	0.35	0.24
OECD $\sum$	0.15	0.20	0.30	0.35	0.25	0.25
$\sum$	0.16	0.22	0.36	0.41	0.34	0.30

Table 3.12: Rejection rates across different functional forms and forecast subgroups

Canada, IMF forecast						
	w(t)	|t|	t^2	t^3	t^4	exp(t)
			current year			
Inst=1	alpha	0.65	0.68	0.68	0.67	0.65
	s.e	0.45	0.39	0.42	0.51	0.02
	p-value	0.74	0.65	0.67	0.73	0.00
Inst=2	alpha	0.60	0.64	0.62	0.59	0.08
	s.e	0.23	0.25	0.29	0.33	0.01
	p-value	0.67	0.58	0.67	0.78	0.00
Inst=3	alpha	0.58	0.60	0.61	0.62	0.06
	s.e	0.30	0.39	0.54	0.64	0.00
	p-value	0.78	0.80	0.83	0.85	0.00
Inst=4	alpha	0.59	0.69	0.93	0.98	0.00
	s.e	0.24	0.21	0.08	0.05	0.00
	p-value	0.70	0.38	0.00	0.00	0.00
			1-year ahead			
Inst=1	alpha	0.59	0.64	0.63	0.61	0.59
	s.e	0.14	0.55	0.61	0.84	0.42
	p-value	0.51	0.80	0.83	0.90	0.83
Inst=2	alpha	0.28	0.31	0.30	0.25	0.07
	s.e	0.23	0.17	0.15	0.13	0.06
	p-value	0.33	0.26	0.18	0.05	0.00
Inst=3	alpha	0.57	0.86	1.05	1.12	0.00
	s.e	0.19	0.20	0.18	0.23	0.02
	p-value	0.71	0.06	0.00	0.01	0.00
Inst=4	alpha	0.24	0.73	0.94	0.99	0.00
	s.e	0.11	0.20	0.09	0.05	0.01
	p-value	0.02	0.25	0.00	0.00	0.00

Table 3.13: Estimated α parameters for Canada's IMF forecasts using different functional forms and instruments

10.14754/CEU.2023.02

C Appendix for Chapter 3

Appendix C.1: Sketch of Proof of Theorem 1.

1. Using the delta-method, we can write $\hat{U}_t$ in the following linear form (assuming that the second and higher order parts of the Taylor-expansion are zero):
   $\sqrt{T}(\hat{U}_T - U) \approx \frac{1}{q}\sqrt{T}(\hat{p}_T - p) - \frac{p}{q^2}\sqrt{T}(\hat{q}_T - q)$.

   The central limit theorems for the univariate iid series $\hat{p}_t$ and $\hat{q}_t$ are the following:
   $E(\hat{p}_t) = p \quad Var(\hat{p}_t) = p(1-p) < \infty$, then $\sqrt{T}(\hat{p}_t - p) \xrightarrow{\text{d}} N(0, p(1-p))$
   $E(\hat{q}_t) = q \quad Var(\hat{q}_t) = q(1-q) < \infty$, then $\sqrt{T}(\hat{q}_t - q) \xrightarrow{\text{d}} N(0, q(1-q))$.

   The Cramer-Wold theorem states that $X_n \xrightarrow{\text{d}} X$ if and only if $a'X_n \xrightarrow{\text{d}} a'X$ for all $a \in \mathbb{R}^k$. Let $\begin{pmatrix} p \\ q \end{pmatrix} \xrightarrow{\text{d}} N_k(0, \Sigma)$ then we can take any vector $a \in \mathbb{R}^k$;(k=2 in this case) and show: $a'\left[\sqrt{T}\begin{pmatrix} \hat{p}_t \\ \hat{q}_t \end{pmatrix} - \begin{pmatrix} p \\ q \end{pmatrix}\right] \xrightarrow{\text{d}} a'\begin{pmatrix} p \\ q \end{pmatrix}$. In the case of the upper bound, a), $a = \lambda_U$.

2. Using the delta-method, we can write $\hat{L}_t$ in the following linear form (assuming that the second and higher order parts of the Taylor-expansion are zero):
   $\sqrt{T}(\hat{L}_T - L) \approx \sqrt{T}(\hat{p}_T - p) - \frac{p-1}{(1-q)^2}\sqrt{T}(\hat{q}_T - q)$.
   Then, we use the Cramer-Wold device as in point a) for the upper bound, but now $a = \lambda_L$.