Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions*

Graham Elliott
University of California San Diego

Allan Timmermann
University of California San Diego

April 29, 2002

Abstract

Existing results on the properties and performance of forecast combinations have been derived in the context of mean squared error loss. Under this loss function empirical studies have generally found that estimates of optimal forecast combination weights lead to higher losses than equally-weighted combined forecasts which in turn outperform the best individual predictions. We show that this and other results can be overturned when asymmetries are introduced in the loss function and the forecast error distribution is skewed. We characterize the optimal combination weights for the most commonly used alternatives to mean squared error loss and demonstrate how the degree of asymmetry in the loss function and skews in the underlying forecast error distribution can significantly change the optimal combination weights. We also propose estimation methods and investigate their small sample properties in simulations and in an inflation forecasting exercise.

*We wish to thank seminar participants at Bocconi and Emory University and the 2001 NBER conference on macroeconomics and forecasting, particularly Frank Diebold, Carlo Favero, Massimiliano Marcellino and Paul Ruud, for helpful comments and suggestions. Carlos Capistran provided excellent research assistance. We are grateful for financial support from the NSF under grant SES 0111238.

1. Introduction

Economic decision makers are often presented with numerous competing forecasts generated by different economic models or forecasting services with access to different information sets. A decision maker who needs a forecast could choose a single set of forecasts from a method that has historically performed well according to a particular loss function, or some combination of the available forecasts could be considered. If a combination of forecasts proves to be superior to its individual components, a better model should be available by pooling the information sets used to form the various forecasts. In practice all the basic information sets are not available, so this is not a feasible approach. Combining forecasts becomes an attractive strategy in this situation and combinations have indeed proven empirically to lead to substantial improvements over the forecasts produced by the best single model.[1]

Both the theoretical results on forecast combination weights and empirical findings on the performance of forecast combinations have almost exclusively assumed a symmetric, quadratic loss function. However, in economics and finance forecasting performance is increasingly evaluated under more general loss functions that account for asymmetries; see, e.g., Christofferson and Diebold (1997), Diebold (2001), Granger and Newbold (1986), Granger and Pesaran (2000) and West, Edison and Choi (1996). For example, financial forecasting performance is often measured through the Sharpe ratio which divides mean returns over the standard deviation of returns on a portfolio constructed on the basis of forecasting information.[2] Similarly, it is common to have a value at risk (VaR) objective to account for the disproportionately high costs associated with large losses. This leads to a loss function that puts particular emphasis on draws from the left tail of the forecast error distribution. Central Banks may also have asymmetric preferences which will

[1]See, e.g., the extensive survey by Clemen (1989). More recently Chen, Stock and Watson (1999), Dunis et al (2001), Marcellino (2002), Newbold and Harvey (2001) and Stock and Watson (1998, 1999) have found evidence that combined forecasts often outperform forecasts from the best individual model. Reasons offered for the empirical success of forecast combinations include model misspecification, changes in the underlying parameters, different models' use of heterogenous information and shrinkage estimation effects. See Hendry and Clements (2002) for a discussion of these points.

[2]Hence even if the mapping from forecasts to portfolio weights is linear, there will be a highly nonlinear mapping from the forecasts to the performance metric.

affect their optimal policies, c.f. Peel and Nobay (1998).

This paper offers a comprehensive treatment of forecast combination under general loss functions and forecast error distributions. We extend the existing literature on optimal forecast combinations by generalizing the set of loss functions from the standard symmetric class to allow for arbitrary asymmetries and continuously differentiable as well as non-differentiable functions. Similarly, we consider general forecast error densities and establish new results when this density is elliptically symmetric or generated by a general mixture distribution that can capture arbitrary shapes of the forecast error density and nests the normal distribution as a special case. We establish the factors that determine the optimal weights in population; numerical examples are then used to demonstrate the extent to which the degree of asymmetry in the loss function interacts with the skew and kurtosis of the underlying forecast errors from the individual models in determining optimal combination weights.

Under elliptical symmetry, our paper shows that the forecast combination weights (other than the constant term) are identical for (almost) all loss functions regardless of the values chosen to parameterize these. This invariance result follows since for any set of forecast combination weights, the constant can be adjusted to choose the bias that optimally trades off the bias and variance of the forecast error. The weights on the forecasts are left free to minimize the variance and can thus be found as the solution to a standard quadratic optimization problem. Under elliptical symmetry, for arbitrarily asymmetric loss functions the problem of optimally determining combination weights thus reduces to simply examining the optimal bias.

This result has strong practical implications. Since the forecast combination weights are the same as those under Mean Squared Error (MSE) loss, the estimation problem reduces to a simple two-stage procedure: Least squares estimation of the optimal combination weights followed by estimation of the constant term which controls the optimal bias. Only the latter depends on the shape of the loss function so the search is over a single parameter. This simplifies the estimation problem greatly since typical estimation procedures require search methods which can be difficult if the dimensions of the parameter vector is large.

An often quoted 'folk theorem' in the forecast combination literature is that using average as opposed to optimal weights often works better in practice.[3] The reason is of course that it can be difficult to precisely estimate the optimal forecast combination weights. Equal-weighted forecast combinations may be biased but they also reduce the forecast error variance by not relying on estimated combination weights that depend on second moments of forecast errors. This finding depends, however, on the trade-off between forecast error bias and variance (as well as higher order moments in the case of asymmetric forecast errors) which in turn depends on the shape of the loss function. Consequently this 'folk theorem' can be overturned under asymmetric loss functions and non-Gaussian forecast errors. We find that the larger the asymmetry in the loss function, the larger the gains from using optimally estimated combination weights over the equal-weighted combination.

We finally address estimation of optimal combination weights in a much more general context than the existing literature. For general differentiable loss functions such as linex we propose to estimate the forecast combination weights using M estimation. For loss functions such as linlin that are not everywhere differentiable, we show that estimation can be cast in the context of a quantile regression problem and we exploit the insights that this provides. For asymmetric quadratic loss, we propose an iterated weighted least squares estimation approach. The performance of these and other estimators is examined in Monte Carlo simulations.

The plan of the paper is as follows. Section 2 reviews results for the standard case with mean squared error loss. Section 3 provides theoretical results for the general case with arbitrary loss function and forecast error distribution. Section 4 considers three commonly used asymmetric loss functions and Section 5 investigates forecast combinations in the context of Gaussian mixture distributions. Section 6 discusses estimation of the optimal combination weights, while Section 7 presents Monte Carlo simulation results. Section 8 provides an empirical application to inflation forecasting and Section 9 concludes.

2. Forecast Combination Under Mean Squared Error Loss

Suppose a decision maker is interested in forecasting some variable, y_t, on the basis of an $m-$vector of forecasts of this variable $\hat{\mathbf{y}}_t$. Each element of $\hat{\mathbf{y}}_t$ is determined *ex ante* and is adapted to an expanding sequence of information sets, Ω_t, which

[3] See, e.g., Granger (1989), Stock and Watson (1999) and Fildes and Ord (2001).

constitutes a standard filtration. Hence $\hat{\mathbf{y}}_t$ is adapted to Ω_{t-1}, whereas y_t is not.[4] Ω_{t-1} comprises y_{t-1} in addition to other variables used to predict y_t. We will assume that y_t and $\hat{\mathbf{y}}_t$ have joint distribution $\mathcal{P}((y_t\ \hat{\mathbf{y}}_t')')$ with finite first and second moments

$$E\begin{pmatrix} y_t \\ \hat{\mathbf{y}}_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu} \end{pmatrix} \tag{1}$$

and

$$Var\begin{pmatrix} y_t \\ \hat{\mathbf{y}}_t \end{pmatrix} = \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}_{21}' \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \tag{2}$$

To keep the notation simple, we suppress the dependence of the joint distribution and moments on Ω_{t-1}, simply using $\mathcal{P}(.)$, $E[.]$ and $Var(.)$ in place of $\mathcal{P}(.|\Omega_{t-1})$, $E[.|\Omega_{t-1}]$ and $Var(.|\Omega_{t-1})$.

Motivated by the seminal paper of Bates and Granger (1969), the forecast combination literature has studied the class of linear forecast combinations, $\omega^c + \boldsymbol{\omega}'\hat{\mathbf{y}}_t$, where $\boldsymbol{\omega}$ is an $m-$vector of combination weights and ω^c is a scalar constant.[5] This gives rise to a forecast error, e_t, from the combination,

$$e_t = y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t. \tag{3}$$

Under the assumed moment structure, e_t has first and second moments

$$\begin{aligned} \mu_e &= \mu_y - \omega^c - \boldsymbol{\omega}'\boldsymbol{\mu} \\ \sigma_e^2 &= \sigma_y^2 + \boldsymbol{\omega}'\boldsymbol{\Sigma}_{22}\boldsymbol{\omega} - 2\boldsymbol{\omega}'\boldsymbol{\sigma}_{21}. \end{aligned} \tag{4}$$

Assuming the loss is quadratic and symmetric in the forecast error, $L(e_t) = e_t^2$, the objective is to minimize the following expression:

$$E\left[(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t)^2\right] = (\mu_y - \omega^c - \boldsymbol{\omega}'\boldsymbol{\mu})^2 + \sigma_y^2 + \boldsymbol{\omega}'\boldsymbol{\Sigma}_{22}\boldsymbol{\omega} - 2\boldsymbol{\omega}'\boldsymbol{\sigma}_{21}. \tag{5}$$

Differentiating with respect to ω^c and $\boldsymbol{\omega}$, we obtain the first order conditions

$$\begin{aligned} \frac{\partial E\left[(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t)^2\right]}{\partial \omega^c} &= -(\mu_y - \omega^c - \boldsymbol{\omega}'\boldsymbol{\mu}) = 0, \\ \frac{\partial E\left[(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t)^2\right]}{\partial \boldsymbol{\omega}} &= 2\boldsymbol{\Sigma}_{22}\boldsymbol{\omega} - 2\boldsymbol{\sigma}_{21} - 2\boldsymbol{\mu}(\mu_y - \omega^c - \boldsymbol{\omega}'\boldsymbol{\mu}) = \mathbf{0}. \end{aligned}$$

[4] Thus we are only ruling out the uninteresting case where y_t is perfectly predictable.

[5] Linearity in the combination weights is not a particularly restrictive assumption at least in the sense that higher order powers of the forecasts are easily included in this setup simply by expanding the space of forecasting signals that are being combined.

The optimal population values of the constant and the vector of combination weights, ω_0^c and $\boldsymbol{\omega}_0$, are thus

$$\begin{aligned} \omega_0^c &= \mu_y - \boldsymbol{\omega}'\boldsymbol{\mu} \\ \boldsymbol{\omega}_0 &= \Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}. \end{aligned} \quad (6)$$

The sample analog of these weights is of course the usual least squares estimator for the outcomes regressed on a constant and the vector of forecasts (Granger and Ramanathan (1984)). Notice the dichotomy of this solution: the optimal constant depends on the biases in the forecasts but not on the variance-covariance matrix. In contrast, the vector of combination weights depends only on the variance-covariance matrix of the outcome and predictions. The inclusion of a constant ensures that the combined forecast is unbiased since the choice of ω_0^c ensures that $\mu_e = 0$. This is clearly optimal under MSE loss.

3. Forecast Combination under General Loss Functions and Forecast Error Distributions

While the Mean Squared Error loss function has proved to be a useful approximation in many cases, a large volume of work has recently been interested in extending the results to allow for more general asymmetric loss functions. Tay and Wallis (2000) provide several references to this literature. Christoffersen and Diebold (1997) develop the theory for optimal bias adjustment in the case of commonly used asymmetric loss functions when the moments of the underlying error distribution vary over time, such as under volatility clustering, but maintain the assumption that (conditionally) the forecast errors follow a Gaussian distribution.

None of these results have been extended to the forecast combination literature. To fill out this gap, this section develops results for optimal forecast combinations under general loss functions and forecast error distributions. We then explore the significance of the factors determining the optimal combination weights for the most commonly used asymmetric loss functions such as linlin, linex and asymmetric quadratic loss.

3.1. *The Loss Function*

We will be concerned with loss functions that only depend on the forecast error and thus take the form $L(e_t)$. Granger (1999) provides an outline of the basic properties such loss functions are required to have:

1. $L(0) = 0$;
2. $\underset{e}{Min}\ L(e) = 0$, so $L(e) \geq 0$;
3. $L(e)$ is monotonic nondecreasing as e moves away from zero:

$$L(e_1) \geq L(e_2) \text{ if } e_1 > e_2 > 0 \text{ and if } e_1 < e_2 < 0.$$

In addition to these properties, the loss function may also be symmetric ($L(-e) = L(e)$), homogenous ($L(ae) = h(a)L(e)$ for some positive function $h(a)$) and differentiable up to some order.

The decision maker's problem is to find the optimal forecast combination weights and a constant that minimize expected loss:

$$Arg \min_{\omega^c, \boldsymbol{\omega}} \int L(e_t) dF(e_t). \tag{7}$$

where $F(e_t)$ is the cumulated density of e_t.

It is clear from this generic optimization problem that the optimal forecast combination weights generally depend on the shapes of both the loss function and the forecast error distribution. The joint effect of the moments of the forecast errors and the loss function can perhaps best be demonstrated in the context of a Taylor series expansion of the loss function around the bias of the forecast error, $\mu_e = E[e_t]$. Under general conditions we have the following result:

Proposition 1. *Suppose that*

1. *the expected loss is finite;*
2. *the loss function is analytic except for possibly at a finite number of points (occurring with probability zero) at which it is continuous;*
3. *all conditional moments of the forecast error distribution up to the highest non-zero derivative of the loss function with respect to the forecast combination weights exist.*

Then the optimal combination weights solve the expression

$$Arg \min_{\omega^c, \boldsymbol{\omega}} \left\{ L(\mu_e) + \frac{1}{2} L''_{\mu_e} E[(e_t - \mu_e)^2] + \sum_{m=3}^{\infty} L^m_{\mu_e} \sum_{i=0}^{m} \frac{1}{i!(m-i)!} E[e_t^{m-i} \mu_e^i] \right\}.$$

$L^k_{\mu_e}$ in the proposition represents the k^{th} derivative of $L(.)$ evaluated at μ_e, i.e. $\partial^k L(e_t)/\partial^k \boldsymbol{\omega}|_{e_t=\mu_e}$. The result follows from noting that, at the points where $L(.)$ is analytic, it lends itself to a Taylor-series expansion around μ_e:

$$L(e_t) = L(\mu_e) + L'_{\mu_e}(e_t - \mu_e) + \frac{1}{2} L''_{\mu_e}(e_t - \mu_e)^2 + \sum_{k=3}^{\infty} (\frac{1}{k!}) L^k_{\mu_e}(e_t - \mu_e)^k.$$

Taking expectations, the finite number of points where $L(.)$ is not analytic can be ignored since they are assumed to occur with probability zero. We therefore get the expected loss:

$$\begin{aligned}
E[L(e_t)] &= L(\mu_e) + \frac{1}{2} L''_{\mu_e} E[(e_t - \mu_e)^2] + \sum_{k=3}^{\infty} (\frac{1}{k!}) L^k_{\mu_e} E[(e_t - \mu_e)^k] \\
&= L(\mu_e) + \frac{1}{2} L''_{\mu_e} E[(e_t - \mu_e)^2] + \sum_{k=3}^{\infty} (\frac{1}{k!}) L^k_{\mu_e} \sum_{i=0}^{k} \binom{k}{i} E[e_t^{k-i} \mu_e^i] \\
&= L(\mu_e) + \frac{1}{2} L''_{\mu_e} E[(e_t - \mu_e)^2] + \sum_{k=3}^{\infty} L^k_{\mu_e} \sum_{i=0}^{k} \frac{1}{i!(k-i)!} E[e_t^{k-i} \mu_e^i].
\end{aligned}$$

Notice that the second assumption does not rule out loss functions such as lin-lin which are non-differentiable at a single point. Furthermore, when the loss function is not linear or quadratic, higher order moments of the forecast error distribution such as the skew generally matter. Finally for the expected loss to exist, the moment generating function of the forecast errors must be such that all moments exist for which the corresponding derivative of the loss function with respect to the forecast error is non-zero. This is a strong requirement and rules out many potential combinations of loss functions and forecast error distributions.[6]

[6]For example, combining a $t-$distribution with three degrees of freedom with a loss function whose fourth derivative is non-zero will result in a non-existing expected loss. Another example comes from the linex loss function, whose higher order derivatives are all non-zero. All moments of the forecast error distribution must therefore exist to ensure that the expected loss is well defined.

3.2. *Elliptically Symmetric Data*

An important special case arises when the expected loss function depends only on the first two moments of the forecast errors. For this case, only the constant (ω^c) depends on the shape of the loss function, while the forecast combination weights are unaltered from the case with MSE loss:

Proposition 2. *Suppose that the expected loss can be written as* $E[L(e_t)] = g(\mu_e, \sigma_e^2)$ and *let* μ_e^* *be the optimal value for* μ_e*. Then if the partial derivative* $\partial g(\mu_e^*, \sigma_e^2)/\partial \sigma_e^2$ *is nonzero,*

(i) $\boldsymbol{\omega}_0 = \Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$;

(ii) ω_0^c *is the solution to* $\partial g(\mu_e^*, \sigma_e^2)/\partial \mu_e = 0$.

Proposition 2 is straightforward to prove. Since ω^c only appears in μ_e, the optimal value for this parameter solves

$$\frac{\partial g(\mu_e, \sigma_e^2)}{\partial \omega^c} = \frac{\partial g(\mu_e, \sigma_e^2)}{\partial \mu_e}\frac{\partial \mu_e}{\partial \omega^c} = -\frac{\partial g(\mu_e, \sigma_e^2)}{\partial \mu_e} = 0.$$

This is the result in part (ii) of the proposition. This choice for ω^c sets μ_e at its optimal point (μ_e^*). The optimal value for $\boldsymbol{\omega}$ therefore solves

$$\frac{\partial g(\mu_e^*, \sigma_e^2)}{\partial \boldsymbol{\omega}} = \frac{\partial g(\mu_e^*, \sigma_e^2)}{\partial \sigma_e^2}\frac{\partial \sigma_e^2}{\partial \boldsymbol{\omega}} = \mathbf{0}.$$

Furthermore, $\frac{\partial \sigma_e^2}{\partial \boldsymbol{\omega}} = 2\left(\Sigma_{22}\boldsymbol{\omega} - \boldsymbol{\sigma}_{21}\right)$ so the solution is as stated so long as $\frac{\partial g(\mu_e^*, \sigma_e^2)}{\partial \sigma_e^2} \neq 0$. It is easy to show that the second order condition is satisfied as long as the loss function satisfies conditions (1) - (3) in Granger (1999). Section 4 shows how to verify the primitive conditions of Proposition 2 for a set of standard asymmetric loss functions.

There are numerous cases where the expected loss takes the form assumed in Proposition 2. Most obviously, if the forecast error distribution is Gaussian then the expected loss depends only on the mean and variance and the optimal forecast combination weights are identical to the MSE weights.[7] This is true regardless of the degree of asymmetry in the loss function.[8] In contrast, the constant term is

[7]Obviously the parameters of the loss function also matter but these are typically assumed to be known constants in applications.

[8]This implication is closely related to the result in finance that the distribution of a portfolio is determined by the mean and variance in these situations, see Chamberlain (1993).

determined to ensure that the optimal bias reflects the asymmetry of the loss function. The same is true for any marginal distribution of forecast errors that depends only on the first two moments of the forecast errors. The general class of distributions for which this is true is for $\mathcal{P}((y_t\ \hat{\mathbf{y}}_t')')$ elliptically symmetric. The multivariate normal is a special case of this family, as is the multivariate t-distribution. We shall for simplicity refer to distributions satisfying the assumptions of Proposition 2 as elliptically symmetric.

Proposition 2 has two important implications. First, estimation of forecast combination weights under asymmetric loss simplifies to a two-stage procedure of first employing the MSE combination weights to the vector of forecasts and then bias adjusting (which requires estimating a single parameter, ω_0^c). These weights are simple to compute regardless of the form of the loss function, e.g. by running a least squares regression and then estimating the mean conditional on the OLS estimates and forecasts that are to be combined.

The second implication is that departures from elliptical symmetry are needed to drive the forecast combination weights away from their solution under MSE loss. If, under elliptical symmetry, a forecast adds information (in the sense that it has a non-zero combination weight) under MSE loss, it will also add information under arbitrarily asymmetric loss functions. The converse also holds - if the forecast is not useful under MSE loss then it will not be useful for any other loss function:

Corollary 1. *Suppose that the conditions of Proposition 2 hold. Then a forecast gets assigned a non-zero weight in the combined forecast under a general loss function if and only if its weight under MSE loss is non-zero.*

The proof is immediate from Proposition 2, since the combination weights are identical irrespective of the assumed form of the loss function provided that the forecast errors are elliptically symmetric. This result stands in stark contrast to the results we obtain under general (non-elliptical) forecast error distributions. For sufficient deviation from elliptical symmetry (particularly when there is skew in the forecast errors) the weights on the forecasts will depend on the loss function. However, our result says that under elliptically symmetric forecast errors, decision makers agree on the value of prediction signals irrespective of their loss function.[9]

[9]Notice also the relationship between Corollary 1 and the result in Diebold, Gunther and Tay (1998) that decision makers with different loss functions cannot generally agree in their

They could well disagree about the need to include an intercept as this depends on the shape of the loss function.

When elliptical symmetry is abandoned, a particular forecast or subset of forecasts may contain information under MSE loss but not under asymmetric loss. The converse may also hold: a forecast could be assigned a non-zero weight under asymmetric loss but not under MSE loss. We state this result in Corollary 2 which is the natural converse of the result in Corollary 1 which was established under elliptical symmetry:

Corollary 2. *Suppose that the conditions of Proposition 2 hold. Then*

(i) a forecast may contain information under MSE loss but not under alternative (asymmetric) loss functions;

(ii) a forecast may contain information under asymmetric loss but not under MSE loss.

Corollary 2 is easy to prove by examining the first order conditions of an asymmetric loss function (e.g., linex loss) under, say, a Gaussian mixture distribution and noting that these generally do not simplify to the formulas in equation (6).

3.3. *Forecast Encompassing Tests*

An alternative to the strategy of forecast combination is to simply use a single forecast. This situation arises when the forecast produced by one model, B, does not add anything to the forecast from another model, A. In this situation, model A is said to forecast encompass model B. Under MSE loss usually this is tested by least squares estimation of the equation

$$y_{t+1} = \alpha + \beta_A \widehat{y}_{t+1}^A + \beta_B \widehat{y}_{t+1}^B + \varepsilon_{t+1}. \tag{8}$$

If the joint hypothesis that $\beta_A = 1$, $\beta_B = 0$ cannot be rejected, model A forecast encompasses model B.

An implication of the previous two propositions and corollaries is the following

ranking of misspecified predictive density models. Provided that the forecast errors are elliptically symmetric, our result says that decision makers can agree about what is regarded as a useful forecast even if the forecasts are individually misspecified.

Proposition 3. *Suppose that the assumptions of Proposition 2 hold. Then if model A forecast encompasses model B under MSE loss, model A also encompasses model B for any other loss function.*

Conversely, if these assumptions do not hold, it is possible that model A forecast encompasses model B under MSE loss but not under an alternatie loss function or that model A does not forecast encompass model B under MSE loss, but does so under the alternative loss function.

When forecast errors are not generated by an elliptically symmetric distribution, it follows that it is not possible to set up a universal encompassing test since the outcome will depend on the specific loss function. Effectively the comparison of the forecasting performance of the individual models depends on the parameters of the loss function.

4. Results for Specific Loss Functions

To demonstrate the general results from the previous section, this section considers three commonly entertained asymmetric loss functions, namely linex, lin-lin and asymmetric quadratic loss. To provide a first impression of the significance of adopting an asymmetric loss function, Figure 1 uses the lin-lin and asymmetric quadratic loss functions to show three very different types of loss aversion. There is a single parameter, θ, that controls the degree of asymmetry. In the first panel ($\theta = 0.1$) very little weight is put on negative forecast errors, while a large weight is put on positive forecast errors. The second panel covers the special case where the loss function is symmetric ($\theta = 0.5$). In the third panel, large weight is put on negative forecast errors and small weight on positive forecast errors.

4.1. *Linex Loss*

We first characterize the solution to the optimal (population) forecast combination weights for the most popular differentiable asymmetric loss function, namely linex loss. This loss function is given by

$$L(e_t) = \exp(ae_t) - ae_t - 1. \tag{9}$$

The parameter a controls the extent of asymmetry. If $a > 0$, there are large losses from positive forecast errors and the losses are higher the larger the value of a. If

$a < 0$, large losses result from negative forecast errors and the losses are larger, the smaller is a. The expected loss is

$$E[L(e_t)] = M_e(a) - a\mu_e - 1, \tag{10}$$

where $M_e(a)$ is the moment generating function of the forecast errors and thus depends on their marginal distribution. Equation (10) shows that the expected linex loss only exists when the forecast error distribution has an infinite number of moments.[10] Differentiating the expected loss with respect to ω^c and $\boldsymbol{\omega}$ yields the first order conditions

$$\begin{aligned} \frac{\partial E[L(e_t)]}{\partial \omega^c} &= \frac{\partial M_e(a)}{\partial \omega^c} + a = 0 \\ \frac{\partial E[L(e_t)]}{\partial \boldsymbol{\omega}} &= \frac{\partial M_e(a)}{\partial \boldsymbol{\omega}} + a\boldsymbol{\mu} = \mathbf{0}. \end{aligned} \tag{11}$$

These equations fully characterize the solution to the optimal weights as a function of the parameters of the underlying forecast error distribution and of the loss function. The parameter a which controls the asymmetry in the loss function plays a prominent role in the first order conditions.

In common with most other asymmetric loss functions, the expression is generally a highly nonlinear function in the forecast combination weights and simple closed form solutions are typically not available. An exception arises when the forecast errors are normally distributed. In this case $M_e(a) = \exp\left\{a\mu_e + \frac{a^2}{2}\sigma_e^2\right\}$, and the optimal constant is given by

$$\omega_0^c = \mu_y - \boldsymbol{\omega}_0'\boldsymbol{\mu} + \frac{a}{2}\sigma_e^2$$

Under normality of the forecast errors, the optimal population bias from the combined forecasts is $\mu_e = -\frac{a}{2}\sigma_e^2$. This is the optimal bias derived in Christoffersen and Diebold (1997) for the univariate "mean-only" linex forecasting problem. The expression has an intuitive interpretation: when $a > 0$, large losses follow from positive forecast errors and it is optimal to choose a positive constant so that

[10] This follows from Proposition 1 since all derivatives of the linex loss function are nonzero. The primitive condition in Proposition 2 requires that

$$\frac{\partial M_e(a)}{\partial \sigma_e} = a\sigma_e \exp\left\{a\mu_e + \frac{a^2}{2}\sigma_e^2\right\} \neq 0.$$

This is always satisfied, except in the trivial case with perfect forecasts ($\sigma_e = 0$).

the forecast errors have a negative mean. The forecast error distribution is hence biased towards negative losses to avoid the high penalties associated with positive losses. The reverse is true when the decision maker tries to avoid negative losses (i.e. $a < 0$). In both cases, the larger is a the larger the penalty so the larger the bias. The larger the variance of the forecast errors the higher the chance of large forecast errors and hence the more the combination gets biased to reduce the chance of large losses.

4.2. *Lin-lin Loss*

A popular way of capturing asymmetries in the loss function is to let losses take one form for forecast errors over and above a certain threshold, κ, and another form for forecast errors below this value:

$$L(e_t) = \begin{cases} L_1(e_t) & \text{if } e_t > \kappa \\ L_2(e_t) & \text{if } e_t \leq \kappa \end{cases},$$

where $L_1(\kappa) = L_2(\kappa)$ ensures continuity. Asymmetries are typically associated with different losses for positive and negative forecast errors so that $\kappa = 0$ is the threshold. If we further assume that $L(.)$ is piece-wise linear, we get the linlin loss function:

$$L(e_t) = \begin{cases} a\,|e_t| & \text{if } e_t > 0 \\ b\,|e_t| & \text{if } e_t \leq 0 \end{cases}, \tag{12}$$

where a and b are non-negative scalars. Defining $\theta = \frac{b}{a+b}$, this can be written as

$$L(e_t) = (a+b)\left\{-\theta + 1_{e_t>0}\right\} e_t. \tag{13}$$

Minimizing this expression over the constant and combination weights is equivalent to solving the problem

$$Arg \min_{\omega^c, \boldsymbol{\omega}} \left\{-\theta + 1_{e_t>0}\right\} e_t.$$

This differs from the original loss function only by a positive scaling factor that is independent of the weights. Hence the two loss functions are in the same homogenous class and the expected loss to be minimized, $L^*(.)$ can equivalently be written as

$$E\left[L^*(e_t)\right] = \int_0^\infty e_t dF(e_t) - \theta\mu_e,$$

where we recall that $F(e_t)$ is the cumulative density function of the forecast error. To evaluate the objective function, we express e_t as $e_t = \mu_e + \sigma_e z_t$, where z_t is simply the centered and standardized forecast error with density $f_z(.)$ and cumulative density $F_z(z)$. Using this transformation the expected loss can be written as

$$E\left[L^*(e_t)\right] = \mu_e \left(1 - \theta - F_z\left(-\frac{\mu_e}{\sigma_e}\right)\right) + \sigma_e \int_{-\mu_e/\sigma_e}^{\infty} z_t dF_z(z_t). \tag{14}$$

The form of this expected loss function means that the derivatives of $L^*(.)$ with respect to ω^c and $\boldsymbol{\omega}$ are difficult to interpret and not very useful since, in many cases, the marginal distribution $F_z(.)$ will itself depend on $(\omega^c, \boldsymbol{\omega})$. An exception to this occurs when e_t is Gaussian and, using Leibnitz' rule of integration, the first order condition for the constant simplifies to

$$1 - \theta - \Phi\left(-\frac{\mu_e}{\sigma_e}\right) = 0,$$

where $\Phi(.)$ is the cumulative density of a standard normal variate.

Recalling that $\mu_e = \mu_y - \omega^c - \boldsymbol{\omega}_0' \boldsymbol{\mu}$ and using symmetry of $\Phi(.)$ we obtain a closed-form solution for the scalar intercept term:[11]

$$\omega_0^c = \mu_y - \boldsymbol{\omega}_0' \boldsymbol{\mu} - \sigma_e \Phi^{-1}(\theta). \tag{15}$$

Again we can evaluate the direction of the bias, using that $\mu_e = \sigma_e \Phi^{-1}(\theta)$. For $\theta = 0.5$, the loss function is symmetric and the error from the combined forecast is unbiased. For $\theta < 1/2$, higher weight is placed on positive errors as b is small relative to a. $\Phi^{-1}(\theta)$ is therefore negative and so the optimal combination has a negative mean for all possible variance-covariances of the forecasts that enter into the combination. Again, to avoid the high weight placed on positive forecast

[11] The assumption of Proposition 2 can be verified under symmetry of the forecast errors by differentiating the expected loss with respect to σ_e, evaluated at μ_e^* :

$$\begin{aligned} \frac{\partial}{\partial \sigma_e} E\left[L(e_t)\right] &= \int_{-\mu_e^*/\sigma_e}^{\infty} z_t dF_z(z_t) \\ &= \int_{-\mu_e^*/\sigma_e}^{\mu_e^*/\sigma_e} z_t dF_z(z_t) + \int_{\mu_e^*/\sigma_e}^{\infty} z_t dF_z(z_t) \\ &= \int_{\mu_e^*/\sigma_e}^{\infty} z_t dF_z(z_t) > 0, \end{aligned}$$

by symmetry of $dF_z(.)$ around zero.

errors the optimal combination biases the mean forecast error to be negative. The reverse is true when $\theta > 0.5$. As with linex loss, the size of the bias increases as the variance of the forecasts errors rises, to better avoid high loss outcomes.

4.3. *Asymmetric Squared Loss*

The asymmetric squared loss function is given by

$$L(e_t) = \begin{cases} ae_t^2 \text{ if } e_t > 0 \\ be_t^2 \text{ if } e_t \leq 0 \end{cases} \tag{16}$$

where a, b are non-negative scalars. Alternatively, this can be written as

$$L(e_t) = (a+b)\left\{\theta - (2\theta - 1)1_{e_t>0}\right\} e_t^2,$$

where again $\theta = \frac{b}{a+b}$. Minimizing this expression over $\boldsymbol{\omega}$ is equivalent to solving

$$Arg \min_{\omega^c, \boldsymbol{\omega}} \left\{\theta - (2\theta - 1)1_{e_t>0}\right\} e_t^2. \tag{17}$$

This differs from the original loss function only by a positive scaling factor that is independent of ω^c and $\boldsymbol{\omega}$. Hence the two loss functions are in the same homogenous class.

Again we can write $e_t = \mu_e + \sigma_e z_t$ and change variables in the expression for the expected loss:

$$\begin{aligned} E[L(e_t)] &= \theta(\sigma_e^2 + \mu_e^2) - (2\theta - 1)\Big\{\mu_e^2\left(1 - F_z\left(-\mu_e/\sigma_e\right)\right) \\ &\quad + \sigma_e^2 \int_{-\mu_e/\sigma_e}^{\infty} z_t^2 dF_z(z_t) + 2\mu_e\sigma_e \int_{-\mu_e/\sigma_e}^{\infty} z_t dF_z(z_t)\Big\}. \end{aligned} \tag{18}$$

Taking derivatives of the expected loss with respect to $\boldsymbol{\omega}$ yields the first order condition[12]

$$\theta\left\{\frac{\partial\sigma_e^2}{\partial\boldsymbol{\omega}} + \frac{\partial\mu_e^2}{\partial\boldsymbol{\omega}}\right\} - (2\theta - 1)\left\{\frac{\partial\mu_e^2}{\partial\boldsymbol{\omega}}\left(1 - F_z\left(-\frac{\mu_e}{\sigma_e}\right)\right) - \mu_e^2 f_z\left(-\frac{\mu_e}{\sigma_e}\right)\frac{\partial}{\partial\boldsymbol{\omega}}\left(-\frac{\mu_e}{\sigma_e}\right)\right.$$

[12] Here we used the relations

$$\begin{aligned} \frac{\partial}{\partial m}\int_m^{\infty} z_t^2 dF_z(z_t) &= -m^2 f_z(m) \\ \frac{\partial}{\partial m}\int_m^{\infty} z_t dF_z(z_t) &= -m f_z(m). \end{aligned}$$

$$
\begin{aligned}
&+\frac{\partial \sigma_e^2}{\partial \boldsymbol{\omega}} \int z_t^2 dF_z(z_t)) + \sigma_e^2 \left(-\frac{\mu_e}{\sigma_e}\right)^2 f_z\left(-\frac{\mu_e}{\sigma_e}\right) \frac{\partial}{\partial \boldsymbol{\omega}}\left(-\frac{\mu_e}{\sigma_e}\right) \\
&+2\mu_e \frac{\partial \sigma_e}{\partial \boldsymbol{\omega}} \int z_t dF_z(z_t) + 2\sigma_e \frac{\partial \mu_e}{\partial \omega} \int z_t dF_z(z_t) - 2\sigma_e \mu_e \left(\frac{\mu_e}{\sigma_e}\right) f_z\left(-\frac{\mu_e}{\sigma_e}\right) \frac{\partial}{\partial \boldsymbol{\omega}}\left(-\frac{\mu_e}{\sigma_e}\right)\Bigg\} \\
= \ & \mathbf{0}.
\end{aligned}
$$

Canceling out the terms involving $f_z\left(-\frac{\mu_e}{\sigma_e}\right)$ and using that $\frac{\partial \sigma_e^2}{\partial \omega} = 2(\boldsymbol{\Sigma}_{22}\boldsymbol{\omega} - \boldsymbol{\sigma}_{21})$, $\frac{\partial \sigma_e}{\partial \omega} = \frac{1}{2\sigma_e}\frac{\partial \sigma_e^2}{\partial \omega}$, $\frac{\partial \mu_e}{\partial \omega} = -\boldsymbol{\mu}$ this simplifies to

$$
\begin{aligned}
\mathbf{0} \ = \ & \theta\left(\boldsymbol{\Sigma}_{22}\boldsymbol{\omega} - \boldsymbol{\sigma}_{21} - \mu_e \boldsymbol{\mu}\right) \qquad (19) \\
& -(2\theta - 1)\left\{-\boldsymbol{\mu} \int (\mu_e + \sigma_e z_t) dF_z(z_t) + \frac{1}{\sigma_e}(\Sigma_{22}\boldsymbol{\omega} - \boldsymbol{\sigma}_{21}) \int (\mu_e + \sigma_e z_t) z_t dF_z(z_t)\right\},
\end{aligned}
$$

where the integrals run from $(-\mu_e/\sigma_e)$ to ∞. The first order condition for the constant term takes the complicated form

$$
\omega_0^c = \mu_y - \boldsymbol{\omega}_0' \boldsymbol{\mu} - \frac{(2\theta - 1)\sigma_e \int_{-\mu_e/\sigma_e}^{\infty} z_t dF_z(z_t)}{\theta - (2\theta - 1)\left(1 - F_z\left(-\mu_e/\sigma_e\right)\right)}. \qquad (20)
$$

This term does not simplify nicely even in the Gaussian case. Only when $\theta = 0.5$, do we get the closed-form solution $\boldsymbol{\omega}_0 = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$ and $\omega_0^c = \mu_y - \boldsymbol{\omega}_0'\boldsymbol{\mu}$, which is identical to the MSE result from Section 2. For all other cases, the above formulas do not provide a closed-form solution to ω_0^c since μ_e is itself a function of this constant and one must again resort to solution methods.

5. Mixture Distributions

Section 3 showed that departures from elliptical symmetry are needed to get combination weights that differ from the standard MSE weights. To examine the effect of such departures, this section considers a setup similar to that studied by Marron and Wand (1992) where the joint distribution of the outcomes and the forecasts, $\mathcal{P}((y_t \ \hat{\mathbf{y}}_t')')$, is a mixture of k normals. Such mixture distributions are capable of closely approximating a very wide family of distributions.[13] The forecast error

[13] Marron and Wand (1992) show how such mixtures can be used to generate a wide assortment of distributional shapes. While in principle one could extend this class of Gaussian mixture densities to include mixtures of elliptical distributions (including t-distributions), these can in most cases be approximated by mixtures of normals so that not much generality is obtained this way.

density takes the form

$$\begin{aligned} f(e_t|\omega^c, \boldsymbol{\omega}) &= \sum_{i=1}^{k} p_i \phi_{\sigma_i}(e_t - \mu_i), \\ \sum_{i=1}^{k} p_i &= 1, \quad 0 < p_i \leq 1. \end{aligned} \tag{21}$$

where $\mu_i = \mu_{y_i} - \omega^c - \boldsymbol{\omega}'\boldsymbol{\mu}_i$, $\sigma_i^2 = \sigma_{yi}^2 + \boldsymbol{\omega}'\Sigma_{22i}\boldsymbol{\omega} - 2\boldsymbol{\omega}'\boldsymbol{\sigma}_{21i}$ are the state-specific mean and variance and ϕ_{σ_i} is the normal density with mean zero and standard deviation σ_i. For example, when $k = 2$, with probability p_1 $\mathcal{P}((y_t \ \hat{\mathbf{y}}_t')')$ takes the form

$$\begin{pmatrix} y_t \\ \hat{\mathbf{y}}_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{y1} \\ \boldsymbol{\mu}_1 \end{pmatrix}, \begin{pmatrix} \sigma_{y1}^2 & \boldsymbol{\sigma}_{211}' \\ \boldsymbol{\sigma}_{211} & \boldsymbol{\Sigma}_{221} \end{pmatrix} \right),$$

while with probability $(1 - p_1)$, $\mathcal{P}((y_t \ \hat{\mathbf{y}}_t')')$ takes the form

$$\begin{pmatrix} y_t \\ \hat{\mathbf{y}}_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{y2} \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \sigma_{y2}^2 & \boldsymbol{\sigma}_{212}' \\ \boldsymbol{\sigma}_{212} & \boldsymbol{\Sigma}_{222} \end{pmatrix} \right).$$

From the moment generating function of a normal variable, the Taylor series expansion in Proposition 1 takes a particularly simple form under the Gaussian mixture distribution:

$$Arg \min_{\omega^c, \boldsymbol{\omega}} \sum_{i=1}^{k} p_i \left\{ L(\mu_{e_i}) + \sum_{m=2}^{\infty} L^m(\mu_{e_i}) \frac{1}{(m/2)!} \frac{\sigma_{e_i}^m}{2^{m/2}} \right\}. \tag{22}$$

The summation over m only includes even powers.

There are many advantages to considering this class of Gaussian mixtures. First, in many economic applications, the underlying mixtures correspond to economic states such as recessions and expansions with the benefit that this offers in terms of interpretation of the optimal combination weights in different states. Second, mixture distributions are easy to generalize to the multivariate case, since only the mixing probabilities, means and covariances need to be extended when additional forecasts are considered. Third, it is easy to introduce skew and kurtosis while controlling the first and second moments. Finally, many elliptically symmetric distributions can be obtained as a special case of the mixtures. Let $\mathbf{X}_t$ be an $m \times 1$ vector which has a multivariate normal mixture distribution $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with $i = 1, ..., k$ states. Then if $\boldsymbol{\mu}_i = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_i = \gamma_i \boldsymbol{\Sigma}$ for some scalar γ_i then $\mathbf{X}_t$ is elliptical. To see this, notice that for an elliptical distribution we need to be able to write the joint density as

$$f(\mathbf{x}) = h(|\mathbf{V}|)^{-1/2} g((\mathbf{x}-\mathbf{a})'\mathbf{V}^{-1}(\mathbf{x}-\mathbf{a})). \tag{23}$$

The joint density for $\mathbf{X}_t$ is

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^{k} p_i (2\pi)^{-n/2} |\mathbf{\Sigma}_i| \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)'\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right\} \\ &= (2\pi^n |\mathbf{\Sigma}|)^{-1/2} \sum_{i=1}^{k} p_i \gamma_i \exp\left\{-\frac{\gamma_i}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \end{aligned}$$

thus the result is shown for $\mathbf{V} = \mathbf{\Sigma}, \mathbf{a} = \boldsymbol{\mu},\ h(\mathbf{V}) = 2\pi^n |\mathbf{V}|,\ g(s) = \sum_{i=1}^{k} p_i \gamma_i \exp\left\{-\frac{\gamma_i}{2} s\right\}$. When there exists at least one pair, $i \neq j$, for which $\mu_i \neq \mu_j$, the mixture generally falls outside the elliptically symmetric family.

5.1. *An Example*

To illustrate the effect of skewness and kurtosis on the optimal forecast combination weights under asymmetric loss, we follow Marron and Wand (1992) in separately considering a model with skew and a model with kurtosis. In both cases we mix models with variance-covariance matrices that are of different scales. For the skew model there is also a difference in the means between the two states.

The precise specification of the mixture models is as follows. Both models have two states. The skew model assumes the first state has a probability of 0.6 with distribution

$$\begin{pmatrix} y_t \\ \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 & 0.15 \\ 0.2 & 0.25 & 0.125 \\ 0.15 & 0.125 & 0.2 \end{pmatrix}\right].$$

In state two the means are all assumed to be 0.5 and the variance-covariance matrix is identical to the one in state 1 but divided by 10:

$$\begin{pmatrix} y_t \\ \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} \sim N\left[\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0.1 & 0.02 & 0.015 \\ 0.02 & 0.025 & 0.0125 \\ 0.015 & 0.0125 & 0.02 \end{pmatrix}\right].$$

Effectively the second state is shifted to the right and more tightly distributed around its mean.

The kurtosis mixture assumes that the probability of the first state is 0.2 with a joint distribution

$$\begin{pmatrix} y_t \\ \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} \sim N \left[\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 & 0.15 \\ 0.2 & 0.25 & 0.125 \\ 0.15 & 0.125 & 0.25 \end{pmatrix} \right].$$

In state two the means are the same as for state one and the variance covariance matrix is scaled by dividing by 15:

$$\begin{pmatrix} y_t \\ \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} \sim N \left[\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.067 & 0.013 & 0.01 \\ 0.013 & 0.017 & 0.008 \\ 0.01 & 0.008 & 0.017 \end{pmatrix} \right].$$

This ensures a zero skew as is clear from the following table which gives the moments of each of the variables:

	outcome		1st forecast		2nd forecast	
	skew	kurtosis	skew	kurtosis	skew	kurtosis
model 1	-0.574	3.722	-0.901	2.466	-0.942	2.213
model 2	0.000	9.515	0.000	9.515	0.000	9.515

Model 1 has negative skew for all three variables but little or no excess kurtosis. Model 2 has strong kurtosis but no skew and is thus an example of a non normal model that is nonetheless elliptically symmetric.

For the skewed mixture model the effect on the population weights of varying θ is shown in Figure 2 (for asymmetric quadratic and lin-lin loss, respectively). The extent of the asymmetry clearly matters for the optimal combination weights. In contrast and consistent with Proposition 2, the combination weights derived under the kurtotic mixture model shown in Figure 3 are independent of θ.

It is not surprising that asymmetry in the loss function interacts with asymmetry in the underlying distributions of the outcomes and forecasts that are being combined. Suppose that $\theta < 0.5$, so the decision maker dislikes positive forecast errors more than negative ones. Now consider moving mass of the distribution to the right, so the underlying distribution is positively skewed. This skewness must come from either the outcome variable or the forecasts, presumably from both in practice. The loss function prefers lower probability of positive forecast errors when choosing the combination weights. The extent of the asymmetry in the loss function drives the size of this effect. Hence the combination weights are clearly a function of the parameters of the loss function. Figures 2 and 3 also show that θ has a large effect on the optimal constant. When θ is very small and negative forecast errors are preferred to positive ones, the constant is large and positive, guaranteeing that the forecast error distribution has a negative bias and is centered to the left of zero. In contrast, for large positive values of θ, positive forecast errors are preferred to negative ones and the constant is large and negative, ensuring that the forecast error distribution is centered to the right of zero.

6. Estimation of Forecast Combination Weights

So far we have considered the population values of the optimal forecast combination weights. In practice, forecast combination weights must be estimated from past data. This involves assumptions as to the stability over time of the joint distribution of the predicted variable and prediction signals. A survey of the empirical evidence of employing estimation methods based largely on least squares is provided in Clemen (1989) and further discussed by Diebold and Lopez (1996).

Here we consider two strategies for estimation. If it is known that the forecast errors are elliptically symmetric, or suspected that they are close to elliptically symmetric, Proposition 2 showed that forecast combination weights are invariant under a large set of loss functions. We show in this section that this insight greatly simplifies estimation. The remainder of the section treats the general estimation problem when elliptical symmetry does not hold.

6.1. *Estimation under Elliptical Symmetry*

When the forecast errors are elliptically symmetric, a simple two stage procedure is suggested by our theoretical results in Section 3:

(i) Estimate $\widehat{\boldsymbol{\omega}}$ by OLS, regressing y_t on a constant and $\hat{\mathbf{y}}_t$

(ii) Use the constructed variable $y_t - \widehat{\boldsymbol{\omega}}'\hat{\mathbf{y}}_t$ to estimate $\widehat{\omega}^c$ based on the relevant loss function.

Under linex and lin-lin loss closed form solutions are available for the second stage estimates when the forecast errors have a Gaussian distribution in which case the estimation problem is extremely simple. From the first stage we obtain estimates $\{\widehat{\mu}_y, \widehat{\boldsymbol{\mu}}, \ \widehat{\boldsymbol{\omega}}, \widehat{\sigma}_e^2\}$. These are used in the second stage estimation. For example, in the case of linex loss,

$$\widehat{\omega}^c = \widehat{\mu}_y - \widehat{\boldsymbol{\omega}}'\widehat{\boldsymbol{\mu}} + \frac{a}{2}\widehat{\sigma}_e^2,$$

while under lin-lin loss

$$\widehat{\omega}^c = \widehat{\mu}_y - \widehat{\boldsymbol{\omega}}'\widehat{\boldsymbol{\mu}} - \widehat{\sigma}_e\Phi^{-1}(\theta).$$

The results for asymmetric quadratic loss are a little more difficult as there is no closed form solution for $\widehat{\omega}^c$. But even in this case the estimation problem is much simpler since the optimization procedure only has to search over a one dimensional parameter space regardless of the number of forecasts we are combining.

When $\mathcal{P}((y_t \ \hat{\mathbf{y}}_t')')$ is not believed to be elliptically symmetric, it is more complicated to estimate the combination weights. We first discuss estimators for general loss functions and then show how these apply to the three loss functions considered in Section 4.

6.2. *Moment Estimators*

In general we are interested in choosing the constant and the combination weights $\omega^c, \boldsymbol{\omega}$ that minimize the expected loss. The realized loss in period t, which we denote by $Q_t(\omega^c, \boldsymbol{\omega})$, can be written as

$$Q_t(\omega^c, \boldsymbol{\omega}) = L(\omega^c, \boldsymbol{\omega}|y_t, \widehat{\mathbf{y}}_t, \boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ are the parameters of the loss function. ω^c and $\boldsymbol{\omega}$ can be obtained as an M-estimator based on the sample analog of $E[Q_t]$ using n observations $\{y_t, \widehat{\mathbf{y}}_t\}_{t=1}^n$:[14]

$$Q(\omega^c, \boldsymbol{\omega}) = n^{-1}\sum_{t=1}^{n} Q_t(\omega^c, \boldsymbol{\omega}).$$

An alternative estimator that is available when we can interchange the expectation operator and the derivative of Q_t with respect to ω^c and $\boldsymbol{\omega}$,[15] $\partial Q_t/\partial((\omega^c$

[14]Other examples of estimators of this form include maximum likelihood estimators, where Q is replaced by minus one times the log-likelihood.

[15]This assumption obviously imposes a restriction on the distribution of the forecast error.

$\boldsymbol{\omega}')') = \mathbf{Q}_t'$, selects $\omega^c, \boldsymbol{\omega}$ to satisfy the first order condition

$$E[\mathbf{Q}_t'(\omega^c, \boldsymbol{\omega})] = \mathbf{0}.$$

The corresponding sample analog is the average over the n observations. One problem with this method is that it can give rise to multiple solutions even when the M-statistic is regular.

A second method is related to instrumental variables estimation of a nonlinear regression and can therefore be handled in a GMM framework. The GMM approach estimates $\boldsymbol{\omega}$ by minimizing the quadratic form

$$V = \left(\sum_{t=1}^{n} \mathbf{Q}_t'(\omega^c, \boldsymbol{\omega})\right)' A^{-1} \left(\sum_{t=1}^{n} \mathbf{Q}_t'(\omega^c, \boldsymbol{\omega})\right) \tag{24}$$

where A is some positive definite matrix. As with any GMM problem, we can easily determine the optimal weighting matrix.

An advantage of the M-estimator is that there are a wealth of results available for consistency and asymptotic normality of the estimated weights. A similar result is true for the GMM estimator.

6.2.1. Linex Loss

These estimation methods can be demonstrated for the linex loss function. For this case, $Q_t(\omega^c, \boldsymbol{\omega})$ takes the form

$$Q_t(\omega^c, \boldsymbol{\omega}) = \exp\left(a(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t)\right) - a(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t) - 1.$$

GMM estimation is based on the derivative, $\mathbf{Q}_t'$:

$$\begin{aligned} \mathbf{Q}_t' &= \left(-a \exp\left(a(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t)\right) + a\right) \begin{pmatrix} 1 \\ \hat{\mathbf{y}}_t \end{pmatrix} \\ &= a\left(1 - \exp\left(a(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t)\right)\right) \begin{pmatrix} 1 \\ \hat{\mathbf{y}}_t \end{pmatrix}. \end{aligned} \tag{25}$$

The solution to this is equivalent to the IV estimator obtained from running the nonlinear regression

$$1 = \exp\{a(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t)\} + u_t$$

using $(1, \hat{\mathbf{y}}_t')'$ as instruments.

6.2.2. Lin-lin Loss

Estimation of the combination weights for the lin-lin loss function can be transformed into a quantile regression problem through use of homogenous classes of loss functions. Recall that the lin-lin loss can be written

$$\begin{aligned} L(e_t) &= a|e_t|1_{e_t>0} + b|e_t|1_{e_t\leq 0} \\ &= (a+b)\{(1-\theta)|e_t|1_{e_t>0} + \theta|e_t|1_{e_t\leq 0}\}, \end{aligned}$$

where $\theta = b(a+b)^{-1}$. Since $(a+b) > 0$, the loss function

$$L^*(e_t) = \{(1-\theta)|e_t|1_{e_t>0} + \theta|e_t|1_{e_t\leq 0}\} \tag{26}$$

is in the same homogenous class as $L(e)$ and the vector $(\omega^c, \boldsymbol{\omega}')'$ minimizing either is the same. We can therefore focus on $L^*(e_t)$. This is precisely the loss function associated with the quantile regression problem (see, e.g., Koenker and Basset (1978), Buchinsky (1992)).

Assuming independent and identically distributed forecast errors, the function to be minimized becomes

$$Q_t(\omega^c, \boldsymbol{\omega}) = \frac{1}{T}\left\{(1-\theta)\sum_{\{t:y_t>\omega^c+\boldsymbol{\omega}'\hat{\mathbf{y}}_t\}} |y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t| + \theta \sum_{\{t:y_t\leq\omega^c+\boldsymbol{\omega}'\hat{\mathbf{y}}_t\}} |y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t|\right\}.$$

Buchinsky (1992) shows that this can be rewritten to fit in a method of moments framework. Define the sign function $s(\lambda) = 1_{\lambda>0} - 1_{\lambda\leq 0}$. Then the objective function takes the form

$$Q_t(\boldsymbol{\omega}) = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{2} - \theta + \frac{1}{2}s(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t)\right\}(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t). \tag{27}$$

Differentiating with respect to $(\omega^c, \boldsymbol{\omega})$, the first order conditions are

$$\mathbf{Q}'_t = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{2} - \theta + \frac{1}{2}s(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t)\right\} \\ \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{2} - \theta + \frac{1}{2}s(y_t - \omega^c - \boldsymbol{\omega}'\hat{\mathbf{y}}_t)\right\}\hat{\mathbf{y}}_t \end{pmatrix} = \mathbf{0}, \tag{28}$$

which again fits into the GMM Framework. Buchinsky (1992) shows that the parameters can efficiently be solved for using the simplex method.

6.2.3. Asymmetric Squared Loss

The asymmetric squared loss function can be written as

$$\begin{aligned} L(e_t) &= ae_t^2 1_{e_t>0} + be_t^2 1_{e_t\leq 0} \\ &= (a+b)\left\{(1-\theta)e_t^2 1_{e_t>0} + \theta e_t^2 1_{e_t\leq 0}\right\}. \end{aligned}$$

Since $(a+b)>0$, the loss function,

$$L^*(e_t) = |1-\theta-1_{e_t\leq 0}|e_t^2 \tag{29}$$

is in the same homogenous class as $L(e_t)$ and the solution $(\omega^c, \boldsymbol{\omega}')'$ minimizing the two is the same. $L^*(e_t)$ is the loss function associated with an observation for the "expectile" regression problem examined in Newey and Powell (1987).

We choose the constant and forecast combination weights $(\omega^c, \boldsymbol{\omega}')'$ to minimize

$$Q_t(\omega^c, \boldsymbol{\omega}) = \frac{1}{T}\sum_{t=1}^{T}|1-\theta-1_{y_t\leq\omega^c+\boldsymbol{\omega}'\hat{\mathbf{y}}_t}|(y_t-\omega^c-\boldsymbol{\omega}'\hat{\mathbf{y}}_t)^2. \tag{30}$$

This function is continuously differentiable and iterated weighted least squares can therefore be used to estimate the forecast combination weights. The estimator at iteration n becomes

$$\hat{\boldsymbol{\omega}}_n = \left(\sum_{t=1}^{T}k_{t,n-1}\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t'\right)^{-1}\left(\sum_{t=1}^{T}k_{t,n-1}\tilde{\mathbf{y}}_t y_t\right) \tag{31}$$

where $\tilde{\mathbf{y}}_t = (1\ \hat{\mathbf{y}}_t')'$ and the scalar weights $k_{t,n-1}$ are constructed according to

$$k_{t,n-1} = |1-\theta-1_{y_t<\hat{\boldsymbol{\omega}}_{n-1}'\tilde{\mathbf{y}}_t}|. \tag{32}$$

6.3. *OLS Weights with Bias Correction*

An alternative strategy to estimating all of the parameters of the model along the lines of the preceding subsections is to estimate the forecast combination weights by OLS and only use the optimal loss function for estimating the constant term. The justification for this is that under elliptical symmetry, this would yield a consistent estimator of the optimal population weights and also greatly simplifies the estimation (and testing) problem. In practice, most of the methods above require iterative techniques which raise problems of potential lack of convergence of the parameter estimates.

This strategy would also make sense if the data does not diverge too far from elliptical symmetry or if the dimension of the forecast combination problem is large as in all cases the search procedure gets reduced to a one dimensional search. We will analyze this strategy along with the ones discussed above in a Monte Carlo simulation experiment.

7. Monte Carlo Results

This section reports the performance of the various estimation procedures applied to the linex, lin-lin and asymmetric quadratic loss functions under the two mixture models. In each case we combine two forecasts and assume that a sample of 100 observations is available to estimate the combination weights. All results are based on 5,000 Monte Carlo simulations. We study both the biases in the various estimators as well as the average losses associated with using these estimators.

7.1. *Biases in the Estimators*

To assess the performance of the estimators, Tables 1 through 3 show both population values and the bias in the estimates arising from various estimation methods. The first column shows the value assumed for the asymmetry parameter in the loss function (a in the case of linex loss, θ for lin-lin and asymmetric quadratic loss). The next three columns show the population values of the optimal weights. Then follows the bias in the estimates from the preferred estimation method and the bias arising from ordinary least squares estimation. The final column reports the bias in the constant term when the forecast combination weights are estimated by OLS followed by GMM estimation of the constant.

For the linex loss function Table 1 shows that the biases in the M estimator are very small for all values of a. Under the skewed mixture model the OLS estimates of combination weights are sometimes heavily biased and the bias increases in the degree of asymmetry in the loss function. Using the OLS forecast combination weights and estimating the constant term by nonlinear IV also leads to estimates of the constant that are quite biased. Under the kurtotic mixture model, consistent with Proposition 2 the biases in the OLS estimates of the forecast combination weights are small. However, OLS estimation leads to a large bias in the constant term.

For the lin-lin loss function, the preferred estimation method is quantile regression. Table 2 shows that this estimation method only introduces very minor biases in the estimated forecast combination weights and the constant term, irrespective of whether the data generating process is the skewed or kurtotic mixture model and even under very large degrees of asymmetry in the loss function. The OLS estimator, optimal when $\theta = 0.5$, is quite heavily biased even for small departures from symmetric loss. The biases in the constant are mitigated to a large extent when estimating the forecast combination weights by OLS followed by quantile estimation of the constant.[16] Nevertheless, the estimates have a bias of around 10% for the most asymmetric loss functions and skewed data.

Under asymmetric quadratic loss (Table 3), the iterative WLS estimates are largely unbiased for both the skewed and kurtotic data generating processes. Under the skewed mixture model the OLS estimates of the weights are not consistent and this is reflected in the biases in the Monte Carlo experiment. Estimating the forecast combination weights by OLS followed by WLS estimation of the constant term reduces but does not eliminate the bias in the constant. In the elliptically symmetric mixture model, the OLS estimates are again consistent for the forecast combination weights, but not for the constant term. When the constant term is estimated by WLS, once again we obtain estimates with small biases.

7.2. *Expected Losses*

Tables 4 through 6 examine the expected loss associated with these estimation results. We report the ratio of the expected loss under the estimated weights relative to the loss under the optimal population weights:

$$\frac{E[L(e|\widehat{\boldsymbol{\omega}})]}{E[L(e|\boldsymbol{\omega}^{opt})]}.$$

Results are shown for each of the estimation methods considered in Tables 1-3. As in these tables the first column reports the value of the asymmetry parameter of the loss function. The second column gives the expected loss using the population weights. The third column evaluates the expected loss when the forecast combination weights and the constant is based on the preferred estimation method, OLS estimation and OLS estimation of the weights followed by estimation of the con-

[16]This reduces to a simple LAD estimation problem.

stant term, respectively.[17] We also show the expected loss under an equal-weighted combination. Once again we report separate results for the skewed and kurtotic mixture models.

Results for the linex loss function are contained in Table 4. Under the skewed mixture the M estimator, which provided good estimates of the weight parameters, leads to an increase in the expected loss betwen 5 and 8 percent over the loss under the population parameters. Losses using the OLS estimates are much larger and grow when the asymmetry in the loss function rises. If the forecast combination weights are estimated by OLS while the constant term is estimated by nonlinear IV, the expected loss is significantly lower than that under the 'full' OLS procedure, but it still exceeds the expected loss from M estimation. These results suggest that using suboptimal OLS estimates of the forecast combination weights (which are simple to compute) and then estimating the constant using the optimal procedure is a relatively efficient way to go.

In the elliptically symmetric model the OLS estimator is consistent for the weights on the forecasts but not for the constant. Compared with the results for the skewed mixture, the increase in expected loss due to using OLS estimates is much smaller for the kurtotic mixture model even for values of θ reflecting large asymmetries in the loss function. Using the OLS combination weights and the optimal estimator for the constant term leads to only slightly higher average loss than under M estimation.

When the asymmetry in the loss function is small ($|a| \approx 0$), using equal-weighted forecast combination weights lead to losses that are quite similar to those resulting from OLS estimation. However, as the asymmetry in the loss function rises, the forecast performance of the equal-weighted combination clearly deteriorates relative to the other estimation methods.

Under linlin and linex loss (Tables 5 and 6), we obtain results that are qualitatively very similar to those from the linex case. Under both the skewed and kurtotic mixture models, the quantile regression and WLS estimators produce expected losses that are 3-16% higher than the expected loss under population parameter values. In the case of the skewed mixture model, use of the suboptimal OLS weights can entail an increase in average loss of more than 100 percent relative to the loss under population parameter values. This additional loss grows

[17]The expectation is taken with respect to the true distribution of the forecast errors.

as the degree of asymmetry in the loss function increases. The loss from using OLS estimates remains large in the elliptically symmetric model due to the use of an inconsistent estimator of the constant term. However, use of the OLS weights with the LAD or weighted least squares estimate of the constant entails a fairly small loss over using the optimal estimator and sometimes even has a better small sample performance.[18]

For completeness, and to demonstrate that these results also carry over to data that simultaneously displays skew and kurtosis, Table 7 shows results for a mixture model that sets $p_1 = p_2 = 0.5$ has the same covariance matrices as in the kurtosis mixture model, zero means in state 1 and means of -0.25 in state 2. This model matches more closely the moments of the empirical data in the next section. To preserve space we only report results for the asymmetric quadratic loss function but the results are very similar for linex and lin-lin loss. Under asymmetric loss OLS continues to generate substantial biases in the parameter estimates and this leads to a large increase in the expected loss. In contrast the weighted least squares estimator or the bias-adjusted least squares method produce only small biases and relatively modest values of the expected loss.

The following general conclusions emerge from these results. For linex loss, the M estimator works well even in small samples. For linlin loss the quantile regression estimator is suggested. For the asymmetric quadratic loss function the weighted least squares approach of Newey and Powell (1987) works well. For all three loss functions these estimation methods significantly outperform OLS estimation. The two-stage bias correction method (OLS followed by optimal estimation of the constant) works very well for the kurtotic mixture but typically leads to higher losses than the optimal estimation methods under the skewed mixture.

Finally the average losses under different estimation methods can be compared to those reported for the equal-weighted forecast combination. While the simple OLS weights lead to higher average losses than even weights, the optimal estimators

[18]Part of the reason for this is that the 'optimal bias' in the constant term adjusts for the non-optimal estimation of the combination weights. Note that under the 'constant adjustment' strategy the estimated forecast combination weights, and hence the relevance of the individual forecasts in the construction of the combined forecast, are the same. This means that if one uses this method (which is generally simpler in practice and also less dependent on the exact form of the loss function and parameter chosen) then the usefullness of the individual forecasts rests entirely on their usefulness in the mean square error loss forecast combination.

that account for the shape of the loss function work much better than even weights. There is only one exception to this, namely when the loss function is symmetric ($\theta = 0.5$). The 'folk theorem' on the superiority of using even weights seems to mainly hold in the context of symmetric loss.

8. Empirical Application to Inflation Forecasting

To demonstrate the theoretical ideas developed in the previous sections, we briefly consider an application that combines predictions of changes in the consumer price index (CPI) from a simple autoregressive time-series model with survey predictions. The source of the latter is the Livingston Survey data base maintained by the Federal Reserve Bank of Philadelphia.[19] This provides a time series of predictions of the consumer price index (CPI) six months ahead in time over the period 1946:1 to 2001:1. From this we define the inflation rate as the log-difference, ΔCPI. The forecasts are likely to be based on diverse information sets that comprise a much larger set of public and private information than is typically considered in econometric models.

We combine the aggregate survey predictions[20] with predictions from a simple autoregressive forecasting model. Schwarz's information criterion supported a simple first-order autoregressive (AR(1)) model which we use to predict inflation one-step, or six-months, ahead. One-step-ahead forecasts from the autoregressive model are computed on the basis of the recursively updated parameters estimated from the autoregressive model.[21]

Figure 4 plots the time series of forecast errors from the autoregressive model and the survey data. Clearly the two sets of forecast errors share a common component even though they are based on very different information sets. The estimated correlation between the two series is 0.63 over the full sample. While high, this estimate only reflects the linear correlation between the two sets of forecast errors. A more complete picture of the difference between the two sets of forecast errors is provided in Figure 5 which (in the upper window) plots the density of the forecast

[19] This survey data has previously been used in numerous studies. For a comprehensive list, see the web site maintained by the Federal Reserve Bank of Philadelphia.

[20] These are computed as the arithmetic mean of the individual forecasters' predictions.

[21] Two observations are used to define the first-difference of the logarithm of the CPI and its lagged value and we use the first 10 observations as an initial sample for parameter estimation.

errors from the AR(1) model and (in the lower window) from the Livingston survey data against the normal distribution curve with the same mean and spread. Relative to the normal curve, the two forecast error distributions have more probability mass in the center and in the tails, thus indicating leptokurtosis. However, the shapes of the forecast error densities from the two models are very different. The density generated by the autoregressive model has more probability mass in the right shoulder and tail than the density associated with the Livingston data which has a thicker left shoulder, particularly for forecast errors between -2.5 and -3.5 percent.

Table 8 presents another indication that the two sets of forecasts contain very different information. Using asymmetric quadratic loss, this table examines the expected losses for various values of θ. For small values of θ the Livingston forecasts strongly outperform the AR(1) forecasts. As θ gets larger, this difference declines. When θ is 0.7 or larger the AR(1) forecasts outperform the Livingston forecasts in terms of the average loss over the sample.

To explore the importance the loss function plays in determining the optimal combination weights, we varied the asymmetry parameter, θ, between 0.1 and 0.9. This moves us from strong aversion against large positive forecast errors via a symmetric loss function towards strong aversion against large negative forecast errors. The optimal combination weights as a function of θ are shown in the upper window of Figure 6. Under standard, symmetric loss, the optimal weight on the AR(1) forecast is close to zero. However as the loss function becomes increasingly asymmetric (in either direction), it becomes optimal to put a non-zero weight on the time-series forecast. For small values of θ a positive weight is put on both the AR(1) forecasts and the density forecasts. For large values of θ, the optimal weight on the AR(1) forecasts becomes negative while the weight on the Livingston data increases. Strong aversion against positive forecast errors thus means assigning positive combination weights to both the autoregressive and Livingston predictions. Under strong aversion against negative forecast errors the combination weight on the autoregressive model is strongly negative while the weight on the Livingston prediction is strongly positive. The lower window of Figure 6 repeats this exercise for the asymmetric quadratic loss function. The actual estimates for the weights differ somewhat but the story is basically the same as under lin-lin loss.

To better understand these results, we plot in Figure 7 the forecast error densities corresponding to the optimal combined forecasts for three values of θ, namely $\theta = 0.1$ (strong aversion to large positive forecast errors), $\theta = 0.5$ (symmetric loss) and $\theta = 0.9$ (strong aversion to large negative forecast errors). When $\theta = 0.1$, the forecast error density is centered to the left of zero, corresponding to a negative bias, a result of aversion against large positive forecast errors. In addition the left tail of the forecast error distribution is quite fat while, conversely, the right shoulder and tails of this distribution are thin. Under symmetric loss, the forecast error density is also more symmetric and centered closer to zero. When $\theta = 0.9$, the forecast error distribution is centered to the right of zero and now has a very thin left shoulder and tail but a thick right shoulder and tail. These results clearly demonstrate the effect of asymmetric loss, not just on the forecast error bias, but on the shape of the entire forecast error distribution. Even simple linear combinations of two forecasts can thus give rise to a complete change in the shape of the forecast error density as the loss asymmetry is varied.

Since the forecast combination weights are very sensitive to the parameters of the loss function, we expect that the distribution of the forecast errors (and indeed the individual forecasts as well as the change in inflation) are far from symmetric. Figure 5 already suggested this finding. In Table 9 we provide formal results based on Jarque-Bera tests of normality for each of the underlying variables as well as the forecast errors for three values of θ. As expected we strongly reject normality for the individual forecasts and the actual time series while normality is only rejected for the forecast error distribution when $\theta = 0.1$. Notice also the shift from left to right skew as θ increases from 0.1 to 0.9.

Finally, an implication of the Monte Carlo results in Section 7 is that simply getting the constant correct (imposing the correct bias) is a major part of the battle in terms of obtaining linearly combined forecasts that perform relatively well. In the numerical work we confirmed that simply using the OLS weights on the actual forecasts with the 'optimal' method for estimating the constant entailed only minor additional losses. Table 10 examines this possibility with the inflation rate data using out-of-sample forecast errors under asymmetric quadratic loss. Except when $\theta = 0.9$, the average loss from using OLS forecast combination weights and WLS to estimate the constant is very similar to the loss based on the iterative WLS estimates. When $\theta = 0.9$ the weighted least squares estimation method does lead to better results.

8.1. *Disaggregated Survey Predictions*

The Livingston survey data lists predictions according to the forecasters' affiliation in broad categories. The most frequent affiliations are nonfinancial business ('industry'), academic institution ('academic') and commercial and investment banking ('banking'). Figure 8 shows the forecast error densities for each of these groups. Although the three forecasts are strongly correlated, there are also some significant differences. Notably, the forecast error density associated with the 'banking' professionals has a wider, flatter peak than those produced by academics and industry professionals.

Forecasters from the same area of business are perhaps more likely to use the same information sets and/or models, so we next consider how the weights on these disaggregated forecasts depend on the loss function. We compute averages of the optimal combination weights on the AR(1) predictions and the three survey predictions as a function of θ. All predictions are out-of-sample and the superscripts AR, A, B, I refer to the autoregressive, academic, banking and industry predictions. Table 11 shows the outcome. When the loss function assigns a very large weight to large negative forecast errors ($\theta = 0.1$), the academic forecasts get a zero weight. Otherwise this weight is large and positive. The weight on the 'banking' inflation forecasts is always negative and tends to be small. Most weight is put on the forecasts produced by professionals in nonfinancial business. These consistently receive a weight above 0.5 irrespective of the shape of the loss function. The weight on the autoregressive forecast does not appear to change much by using disaggregated survey data. It continues to be large and positive when $\theta = 0.1$ and large and negative when $\theta = 0.9$.

We finally investigated whether the often quoted finding that simple averages of forecasts outperform estimated optimal weights depends on the shape of the loss function. The right columns in Tables 10 and 11 present the average out-of-sample loss associated with the estimated optimal weights and the equal weights, respectively. The results confirm the frequent finding that equal-weights outperform estimated optimal weights under MSE loss. However, they also show very clearly that this result is overturned under asymmetric loss where use of estimated optimal weights leads to far smaller average losses out-of-sample.

9. Conclusion

Several conclusions emerge from our analysis. If the forecast error distribution is elliptically symmetric then the forecast combination weights are identical for a wide range of loss functions, including the popular mean square error loss function. This implies that if a particular forecast is informative under MSE loss, then this forecast will be informative for a very wide range of loss functions. If there is little departure from elliptical symmetry, then the optimal forecast combination weights are near the standard weights derived under MSE loss. This has practical implications since such weights are often much easier to compute.

Under departures from elliptically symmetric distributions, the optimal combination weights under asymmetric loss can be very different than under mean squared error loss. For example, if negative forecast errors are associated with much lower losses than positive ones, then it will be optimal to select combination weights that give rise to a much larger upward bias than under MSE loss.

Irrespective of the distribution of the forecast errors, the constant term in the forecast combination is affected by the loss function. Our simulations suggest that even with heavily skewed forecast errors, the loss from using MSE combination weights (which are suboptimal) and optimally adjusting the constant results in fairly small increases in expected loss. Since this strategy is simple and has good small sample properties, for many applications this approach will perhaps work well.

Our application showed that the optimal weights in a combination of inflation survey forecasts and forecasts from a simple autoregressive model strongly depend on the degree of asymmetry in the loss function. In the absence of loss asymmetry, the autoregressive forecast does not add much information. However, under asymmetric loss (in either direction), both sets of forecasts appear to contain information in the sense that they have non-zero weights in the combined forecast.

Our inflation forecasting experiment confirmed the frequent finding that equal-weights outperform estimated optimal weights under MSE loss. However, it also showed very clearly that this result can be overturned under asymmetric loss where use of estimated optimal weights led to much smaller average losses out-of-sample.

There are many directions for generalizing our results. To keep things simple, we have not dealt with serial correlation in the forecast errors. However, as shown by Diebold (1988) under MSE loss, this can be done by allowing the forecast errors from the combination model to follow a finite-order ARMA process. Another interesting generalization is to allow the optimal combination weights to vary over time, an idea that is further considered in Elliott and Timmermann (2002) in the context of regime switching models. Finally, it would also be interesting to explore the ideas in this paper in the context of multi-step-ahead forecasting. Some GARCH models imply that the one-step-ahead forecast errors are elliptically symmetric, while multi-step-ahead forecast errors are not. This raises the possibility of letting the estimation method depend on the forecast horizon.

References

Bates, J.M. and C.W.J. Granger, 1969, The Combination of Forecasts. Operations Research Quarterly 20, 451-468.

Buchinsky, M., 1992, Methodological Issues in Quantile Regression. Manuscript, Brown University.

Chamberlain, G., 1993, A Characterization of the Distributions that Imply Mean-Variance Utility Functions. Journal of Economic Theory, 29, 185-201.

Chen, J. Stock and M. W. Watson, 1999, A Dynamic Factor Model Framework for Forecast Combination. Spanish Economic Review 1, 91-121.

Christofferson, P. and F.X. Diebold, 1997, Optimal Prediction under Asymmetrical Loss. Econometric Theory 13, 806-817.

Clemen, R.T., 1989, Combining Forecasts: A Review and Annotated Bibliography. International Journal of Forecasting 5, 559-581.

Diebold, F.X., 1988, Serial Correlation and the Combination of Forecasts. Journal of Business and Economic Statistics 6, 105-111.

Diebold, F.X., 2001, Elements of Forecasting, Second Edition. South-Western Publishing.

Diebold, F. X. and J. A. Lopez, 1996, Forecast Evaluation and Combination. In Maddala and Rao (eds.) Handbook of Statistics. Elsevier: Amsterdam.

Diebold, F.X., T.A. Gunther and A.S. Tay, 1998, Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review 39, 863-883.

Dunis, C. J. Laws and S. Chauvin, 2001, The Use of Market Data and Model Combinations to Improve Forecast Accuracy. In Dunis, C., Timmermann, A. and Moody, J. (eds), Developments in Forecast Combination and Portfolio Choice. Chichester: Wiley.

Elliott, G. and A. Timmermann, 2002, Optimal Forecast Combination Weights Under Regime Switching. Manuscript, University of California San Diego.

Fildes R. and K. Ord, 2001, Forecasting Competitions - Their Role in Improving Forecasting Practice and Research. In Clements and Hendry (eds), 253-322.

Granger, C.W.J., 1989, Combining Forecasts - Twenty Years Later. Journal of Forecasting 8, 167-173.

Granger, C.W.J., 1999, Outline of Forecast Theory Using Generalized Cost Functions. Spanish Economic Review 1, 161-173.

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

Granger, C.W.J. and M.H. Pesaran, 2000, Economic and Statistical Measures of Forecast Accuracy. Journal of Forecasting 19, 537-560.

Granger, C.W.J. and R. Ramanathan, 1984, Improved Methods of Combining Forecasts. Journal of Forecasting 3, 197-204.

Hendry, D.F. and M.P. Clements, 2002, Pooling of Forecasts. Econometrics Journal 5, 1-26.

Koenker, R. and G. Basset, 1978, Regression Quantiles. *Econometrica*, 46, 33-50.

Marcellino, M., 2002, Forecast Pooling for Short Time Series of Macroeconomic Variables. Mimeo, IGIER, Bocconi University.

Marron, J.S. and M.P. Wand, 1992, Exact Mean Integrated Squared Error. Annals of Statistics 20, 712-736.

Newbold, P. and D.I. Harvey, 2001, Forecast Combination and Encompassing. In Clements, M.P. and D.F. Hendry (eds), A Companion to Economic Forecasting. Oxford: Blackwells.

Newey, W. and J. Powell, 1987, Asymmetric Least Squares Estimation and Testing. *Econometrica* 55, 819-847.

Peel, D.A. and A.R. Nobay, 1998, Optimal Monetary Policy in a Model of Asymmetric Central Bank Preferences. FMG discussion paper 0306.

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

Stock J.H. and M. W. Watson, 1999, Forecasting Inflation. Journal of Monetary Economics 44, 293-375.

Tay, A. and K. Wallis, 2000, Density Forecasting: A Survey. Journal of Forecasting 19, 235-254.

West, K.D., H.J. Edison and D. Cho, 1993, A Utility-based Comparison of Some Models of Exchange Rate Volatility. Journal of International Economics 35, 23-46.

Table 1: Small Sample Results for Linex Loss Function

	Population Weights			M Estimates			OLS Estimates			Optimal Constant
a	ω_1	ω_2	ω^c	ω_1	ω_2	ω^c	ω_1	ω_2	ω^c	ω^c
					Skewed Mixture					
-1.25	0.73	0.55	-0.41	0.00	-0.01	0.03	-0.11	-0.19	0.41	0.07
-1.00	0.71	0.52	-0.32	0.00	-0.01	0.02	-0.09	-0.16	0.32	0.06
-0.75	0.69	0.49	-0.24	0.00	-0.01	0.01	-0.07	-0.12	0.24	0.04
0.75	0.55	0.25	0.24	0.00	0.00	-0.01	0.07	0.12	-0.24	-0.04
1.00	0.53	0.21	0.32	0.00	0.01	-0.02	0.09	0.15	-0.32	-0.06
1.25	0.51	0.18	0.41	0.00	0.01	-0.03	0.11	0.18	-0.41	-0.07
					Kurtotic Mixture					
-1.25	0.67	0.27	-0.09	-0.01	-0.01	0.04	-0.01	0.00	0.17	0.03
-1.00	0.67	0.27	-0.05	-0.01	-0.01	0.03	-0.01	0.00	0.13	0.02
-0.75	0.67	0.27	-0.02	0.00	-0.01	0.02	-0.01	0.00	0.10	0.02
0.75	0.67	0.27	0.15	-0.01	0.00	0.00	-0.01	0.00	-0.07	0.00
1.00	0.67	0.27	0.18	-0.01	0.00	-0.01	-0.01	0.00	-0.11	0.00
1.25	0.67	0.27	0.22	-0.01	0.00	-0.02	-0.01	0.00	-0.15	-0.01

Note: The first column reports the value of the parameter, a, that controls the asymmetry of the loss function. The second through fourth columns give the population values of the optimal forecast combination weights and the constant. All other entries are bias estimates based on a sample of 100 observations and 5000 Monte Carlo replications.

Table 2: Small Sample Results for Lin-lin Loss Function.

	Population Weights			Quantile Estimates			OLS Estimates			Optimal Constant
θ	ω_1	ω_2	ω^c	ω_1	ω_2	ω^c	ω_1	ω_2	ω^c	ω^c
					Skewed Mixture					
0.1	0.39	-0.02	0.95	0.00	0.03	0.00	0.23	0.39	-0.95	-0.12
0.2	0.45	0.08	0.59	0.00	0.01	0.00	0.17	0.28	-0.59	-0.10
0.3	0.51	0.19	0.36	0.00	0.01	0.00	0.11	0.18	-0.36	-0.10
0.4	0.57	0.28	0.17	0.00	0.01	0.00	0.05	0.08	-0.17	-0.05
0.5	0.62	0.37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.6	0.67	0.46	-0.17	0.00	0.00	0.00	-0.05	-0.09	0.17	0.05
0.7	0.73	0.55	-0.36	0.00	0.00	0.00	-0.11	-0.18	0.36	0.10
0.8	0.79	0.65	-0.59	0.00	-0.01	0.00	-0.17	-0.29	0.59	0.13
0.9	0.85	0.76	-0.95	0.00	-0.04	0.00	-0.23	-0.39	0.95	0.11
					Kurtotic Mixture					
0.1	0.67	0.27	0.47	-0.01	0.00	0.02	-0.01	0.00	-0.39	0.01
0.2	0.67	0.27	0.31	-0.01	0.00	0.01	-0.01	0.00	-0.23	0.02
0.3	0.67	0.27	0.21	-0.01	0.00	0.01	-0.01	0.00	-0.13	0.01
0.4	0.67	0.27	0.14	-0.01	0.00	0.01	-0.01	0.00	-0.06	0.01
0.5	0.67	0.27	0.07	0.00	0.00	0.01	-0.01	0.00	0.01	0.01
0.6	0.67	0.27	0.00	0.00	-0.01	0.01	-0.01	0.00	0.08	0.01
0.7	0.67	0.27	-0.08	0.00	-0.01	0.01	-0.01	0.00	0.16	0.01
0.8	0.67	0.27	-0.18	0.00	0.00	0.01	-0.01	0.00	0.25	0.01
0.9	0.67	0.27	-0.34	0.00	-0.01	0.00	-0.01	0.00	0.42	0.01

Note: The first column reports the value of the parameter, θ, that controls the asymmetry of the loss function. The second through fourth columns give the population values of the optimal forecast combination weights and the constant. All other entries are bias estimates based on a sample of 100 observations and 5000 Monte Carlo replications.

Table 3: Small Sample Results for Asymmetric Quadratic Loss.

	Population Weights			WLS Estimates			OLS Estimates			Optimal Constant
θ	ω_1	ω_2	ω^c	ω_1	ω_2	ω^c	ω_1	ω_2	ω^c	ω_1
					Skewed Mixture					
0.1	0.50	0.17	0.67	0.00	0.00	-0.02	0.12	0.20	-0.67	-0.05
0.2	0.54	0.23	0.42	0.00	0.00	-0.01	0.09	0.14	-0.42	-0.04
0.3	0.57	0.28	0.26	0.00	0.00	-0.01	0.06	0.09	-0.25	-0.03
0.4	0.59	0.32	0.12	0.00	0.00	0.00	0.03	0.04	-0.12	-0.02
0.5	0.62	0.37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.6	0.65	0.41	-0.12	0.00	0.00	0.00	-0.02	-0.05	0.12	0.01
0.7	0.67	0.46	-0.25	0.00	-0.01	0.00	-0.05	-0.09	0.25	0.03
0.8	0.70	0.51	-0.42	0.00	-0.01	0.01	-0.08	-0.14	0.42	0.04
0.9	0.74	0.56	-0.67	0.00	-0.01	0.02	-0.12	-0.20	0.67	0.05
					Kurtotic Mixture					
0.1	0.67	0.27	0.44	-0.01	0.00	-0.01	-0.01	0.00	-0.36	0.00
0.2	0.67	0.27	0.28	-0.01	0.00	0.00	-0.01	-0.01	-0.20	0.01
0.3	0.67	0.27	0.19	-0.01	0.00	0.01	-0.01	0.00	-0.12	0.01
0.4	0.67	0.27	0.13	-0.01	0.00	0.01	-0.01	0.00	-0.05	0.01
0.5	0.67	0.27	0.07	-0.01	0.00	0.01	-0.01	0.00	0.01	0.01
0.6	0.67	0.27	0.01	-0.01	-0.01	0.01	-0.01	0.00	0.07	0.01
0.7	0.67	0.27	-0.06	-0.01	-0.01	0.02	-0.01	0.00	0.14	0.01
0.8	0.67	0.27	-0.15	-0.01	-0.01	0.02	-0.01	0.00	0.23	0.02
0.9	0.67	0.27	-0.30	-0.01	-0.01	0.03	-0.01	0.00	0.38	0.02

Note: The first column reports the value of the parameter, θ, that controls the asymmetry of the loss function. The second through fourth columns give the population values of the optimal forecast combination weights and the constant. All other entries are bias estimates based on a sample of 100 observations and 5000 Monte Carlo replications.

Table 4: Average Loss for Linex Loss Function.

	Skewed Mixture Model					Kurtotic Mixture Model				
a	Loss	M-est	OLS	OLSc	Ave	Loss	M-est	OLS	OLSc	Ave
-1.25	0.44	1.08	1.36	1.08	1.30	0.20	1.17	1.21	1.11	1.12
-1.00	0.27	1.06	1.23	1.07	1.18	0.12	1.13	1.16	1.10	1.08
-0.75	0.15	1.05	1.14	1.05	1.10	0.06	1.10	1.12	1.09	1.05
0.75	0.15	1.05	1.14	1.06	1.10	0.06	1.11	1.12	1.09	1.05
1.00	0.27	1.06	1.23	1.07	1.19	0.12	1.13	1.16	1.10	1.08
1.25	0.44	1.07	1.36	1.08	1.31	0.20	1.17	1.21	1.11	1.12

Note: The first column reports the value of the parameter, *a*, controlling asymmetry. The second column gives the population expected loss. Other entries give average loss over the 5000 replications based on estimates of the forecast combination weights and the constant from 100 observations. These losses are scaled by the population loss. *OLSc* refers to estimation of the combination weights by OLS followed by optimal estimation of the constant. *Ave* refers to the equal-weighted forecast combination.

Table 5: Average Loss for Lin-lin Loss Function

	Skewed Mixture Model					Kurtotic Mixture Model				
θ	Loss	Quantile	OLS	OLSc	Ave	Loss	Quantile	OLS	OLSc	Ave
0.1	0.13	1.06	2.08	1.08	2.04	0.08	1.08	1.85	1.05	1.79
0.2	0.19	1.03	1.39	1.06	1.36	0.11	1.05	1.35	1.04	1.30
0.3	0.23	1.03	1.15	1.04	1.13	0.13	1.04	1.15	1.04	1.11
0.4	0.26	1.03	1.05	1.02	1.03	0.14	1.03	1.07	1.04	1.03
0.5	0.26	1.03	1.02	1.02	1.00	0.15	1.03	1.04	1.04	1.00
0.6	0.26	1.03	1.05	1.03	1.03	0.14	1.03	1.07	1.04	1.03
0.7	0.23	1.03	1.15	1.04	1.13	0.13	1.04	1.15	1.04	1.11
0.8	0.19	1.03	1.38	1.06	1.35	0.11	1.05	1.35	1.04	1.30
0.9	0.13	1.06	2.06	1.08	2.03	0.08	1.08	1.85	1.05	1.79

Note: The first column reports the value of the parameter, θ, controlling asymmetry. The second column gives the population expected loss. Other entries give average loss over the 5000 replications based on estimates of the forecast combination weights and the constant from 100 observations. These losses are scaled by the population loss. *OLSc* refers to estimation of the combination weights by OLS followed by optimal estimation of the constant. *Ave* refers to the equal-weighted forecast combination.

Table 6: Average Loss for Asymmetric Quadratic Loss Function

	Skewed Mixture Model					Kurtotic Mixture Model				
θ	Loss	WLS	OLS	OLSc	Ave	Loss	WLS	OLS	OLSc	Ave
0.1	0.14	1.06	1.89	1.08	1.82	0.07	1.16	1.65	1.09	1.55
0.2	0.20	1.05	1.34	1.06	1.30	0.09	1.11	1.29	1.08	1.22
0.3	0.24	1.04	1.15	1.05	1.11	0.10	1.09	1.16	1.08	1.09
0.4	0.26	1.04	1.06	1.04	1.03	0.10	1.08	1.09	1.08	1.03
0.5	0.26	1.04	1.04	1.04	1.00	0.10	1.08	1.08	1.08	1.01
0.6	0.26	1.04	1.06	1.04	1.03	0.10	1.08	1.09	1.08	1.03
0.7	0.24	1.04	1.15	1.05	1.11	0.10	1.09	1.16	1.08	1.09
0.8	0.20	1.05	1.34	1.06	1.29	0.09	1.11	1.29	1.08	1.22
0.9	0.15	1.07	1.88	1.07	1.82	0.07	1.16	1.65	1.09	1.55

Note: The first column reports the value of the parameter, θ, controlling asymmetry. The second column gives the population expected loss. Other entries give average loss over the 5000 replications based on estimates of the forecast combination weights and the constant from 100 observations. These losses are scaled by the population loss. *OLSc* refers to estimation of the combination weights by OLS followed by optimal estimation of the constant. *Ave* refers to the equal-weighted forecast combination.

Table 7: Small Sample Results for Asymmetric Quadratic Loss: Mixture with skew and kurtosis matching empirical data.

	Population Weights			WLS Estimates			OLS Estimates			Optimal Constant	Average loss				
θ	ω_1	ω_2	ω^c	ω_1	ω_2	ω^c	ω_1	ω_2	ω^c	ω_1	Loss	WLS	OLS	OLSc	Ave
0.1	0.71	0.52	0.60	0.00	0.00	-0.02	-0.09	-0.16	-0.60	-0.04	0.13	1.08	1.79	1.07	1.71
0.2	0.69	0.48	0.36	0.00	-0.01	-0.01	-0.07	-0.12	-0.36	-0.03	0.18	1.06	1.31	1.06	1.26
0.3	0.67	0.44	0.21	0.00	-0.01	-0.01	-0.05	-0.08	-0.21	-0.02	0.20	1.05	1.14	1.05	1.09
0.4	0.64	0.40	0.10	0.00	-0.01	0.00	-0.02	-0.04	-0.10	-0.01	0.21	1.05	1.07	1.05	1.02
0.5	0.62	0.37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.05	1.05	1.05	1.00
0.6	0.60	0.33	-0.10	0.00	0.00	0.00	0.02	0.04	0.10	0.01	0.21	1.05	1.07	1.05	1.02
0.7	0.57	0.29	-0.21	0.00	0.00	0.00	0.05	0.08	0.21	0.02	0.20	1.05	1.14	1.05	1.10
0.8	0.55	0.25	-0.36	0.00	0.00	0.01	0.07	0.12	0.36	0.02	0.17	1.06	1.31	1.06	1.26
0.9	0.53	0.21	-0.60	0.00	-0.01	0.02	0.09	0.15	0.59	0.03	0.13	1.08	1.78	1.07	1.72

Note: The first column reports the value of the parameter, θ, that controls the asymmetry of the loss function. The second through fourth columns give the population values of the optimal forecast combination weights and the constant. Columns five to eleven are bias estimates based on a sample of 100 observations and 5000 Monte Carlo replications. Column twelve gives the population expected loss. Subsequent entries give average loss over the 5000 replications based on estimates of the forecast combination weights and the constant from 100 observations. These loses are scaled by the population loss.

Table 8. Average Loss from AR(1) and Livingston Forecasts

θ	AR(1)	Livingston	Difference
0.1	0.079	0.038	0.041
0.2	0.074	0.040	0.034
0.3	0.070	0.042	0.028
0.4	0.065	0.044	0.021
0.5	0.060	0.046	0.014
0.6	0.055	0.049	0.006
0.7	0.050	0.051	-0.001
0.8	0.045	0.053	-0.008
0.9	0.040	0.055	-0.015

Note: This table reports the average out-of-sample loss under the asymmetric quadratic loss function when either the AR(1) or the Livingston forecasts are used separately.

Table 9: Jarque-Bera Tests for Normality of Forecast Errors

Series	JB test	Skew	Kurtosis
Actual	25.621	1.119	4.238
AR(1)	36.217	1.237	4.769
Livingston	9.348	0.727	3.524
θ=0.1	10.107	-1.028	4.080
θ=0.5	3.164	-0.567	3.635
θ=0.9	1.064	0.305	3.441

Note: Reported in the first column are Jarque-Bera tests for normality. The 95% critical value is 5.99 and the test rejects for larger values. The second and third columns are measures of skew (zero is no skew) and kurtosis (a value of three is no excess kurtosis), respectively. The first three rows are for the realized series (actual) and forecasts, respectively.
Values in rows marked θ refer to forecast errors estimated using weights under the asymmetric quadratic loss function.

Table 10: Average Losses from Combined and Equal Weighted Forecast

WLS Estimates						Losses	
θ	ω_1	ω_2	constant	Loss	constant	OLSc	Equal weights
0.1	0.594	0.595	0.006	0.028	0.013	0.028	0.034
0.5	-0.001	0.824	0.004	0.053	0.004	0.053	0.046
0.9	-0.428	0.926	0.002	0.027	-0.003	0.037	0.058

Note: WLS estimates are estimates of the parameters using the WLS method described for the asymmetric quadratic loss function. The loss reported is average out of sample one step ahead loss. In the OLSc case OLS weights are used for the forecast combination weights but the constant is estimated using WLS. The final column gives average loss for even weights on the forecasts and a zero constant.

Table 11: Average Losses from Combined and Equal-Weighted Disaggregate Forecasts

WLS Estimates						Losses	
θ	ω	ω^{AR}	ω^{A}	ω^{B}	ω^{I}	WLS^{I}	Equal weights
0.1	0.53	0.56	0.00	-0.27	0.87	0.027	0.033
0.5	0.61	-0.03	0.30	-0.09	0.61	0.052	0.045
0.9	0.23	-0.51	0.48	-0.10	0.60	0.026	0.057

Note: This table is based on combination of the autoregressive (AR), academic (A), bank (B) and industry (I) forecasts. Loss is assumed to be asymmetric quadratic with asymmetry parameter θ.

Figure 1: Lin-Lin and Asymmetric Quadratic loss functions

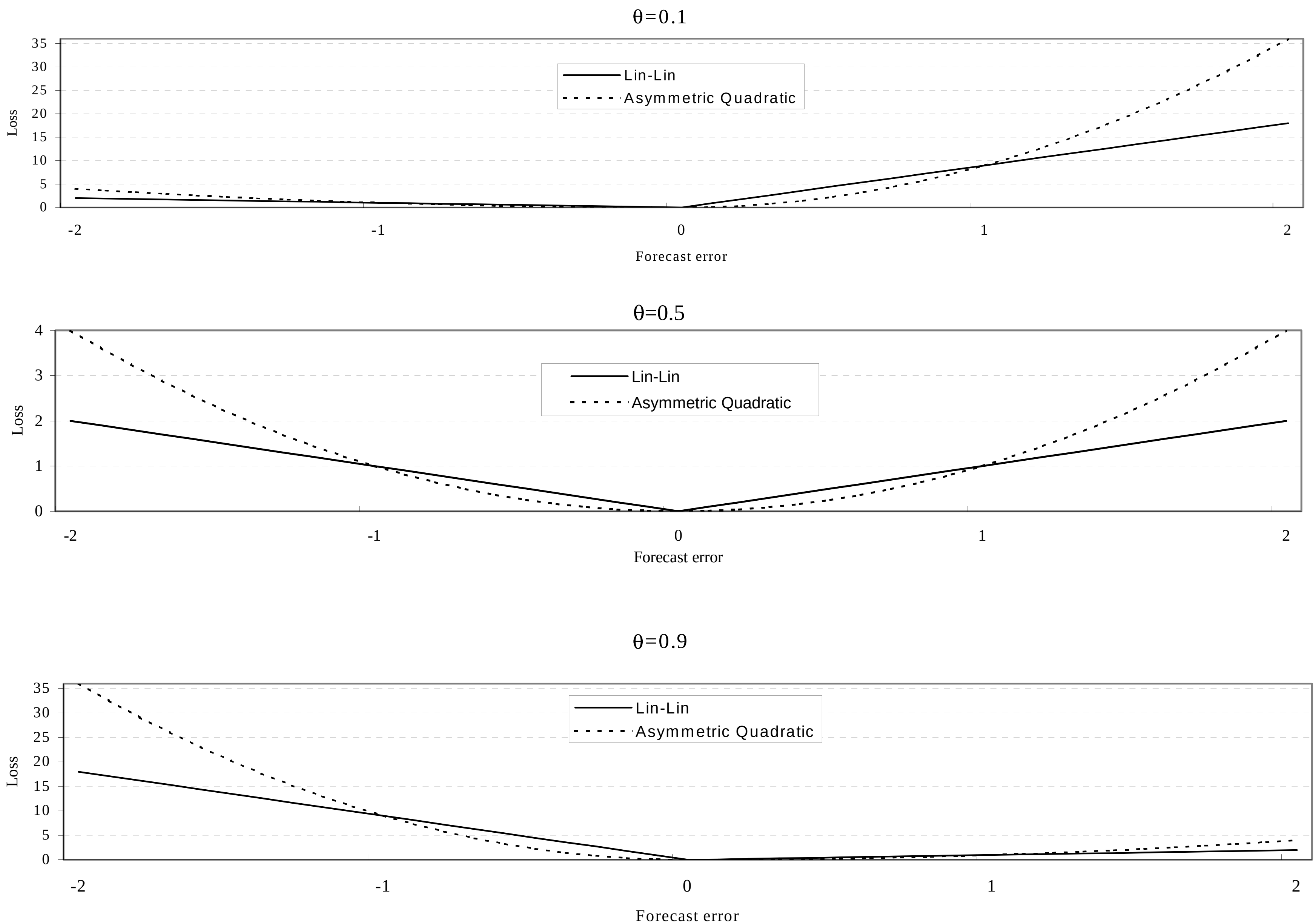
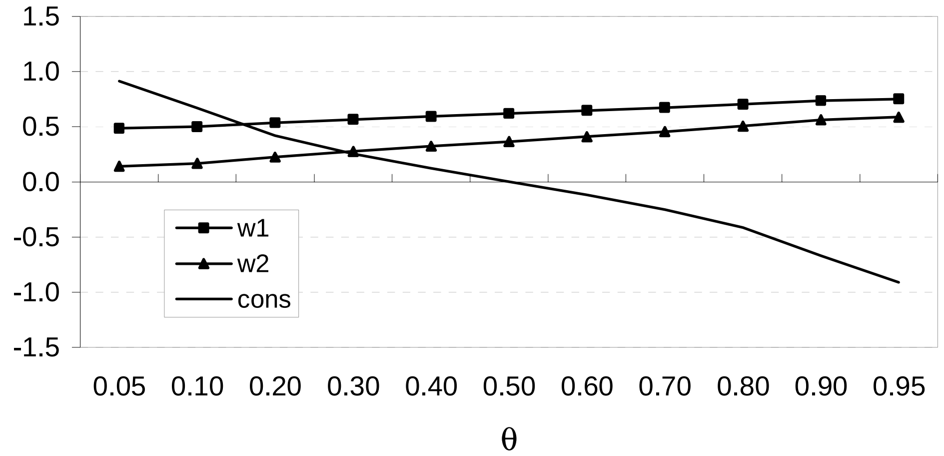


Figure 2: Optimal combination weights and constant

Asymmetric quadratic loss (skewed mixture)



Lin-lin loss (skewed mixture)

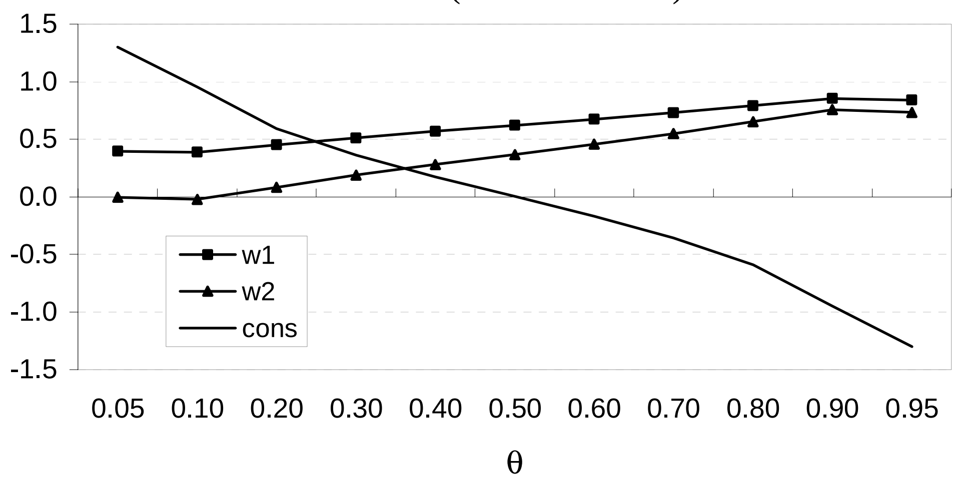


Figure 3: Optimal combination weights and constant

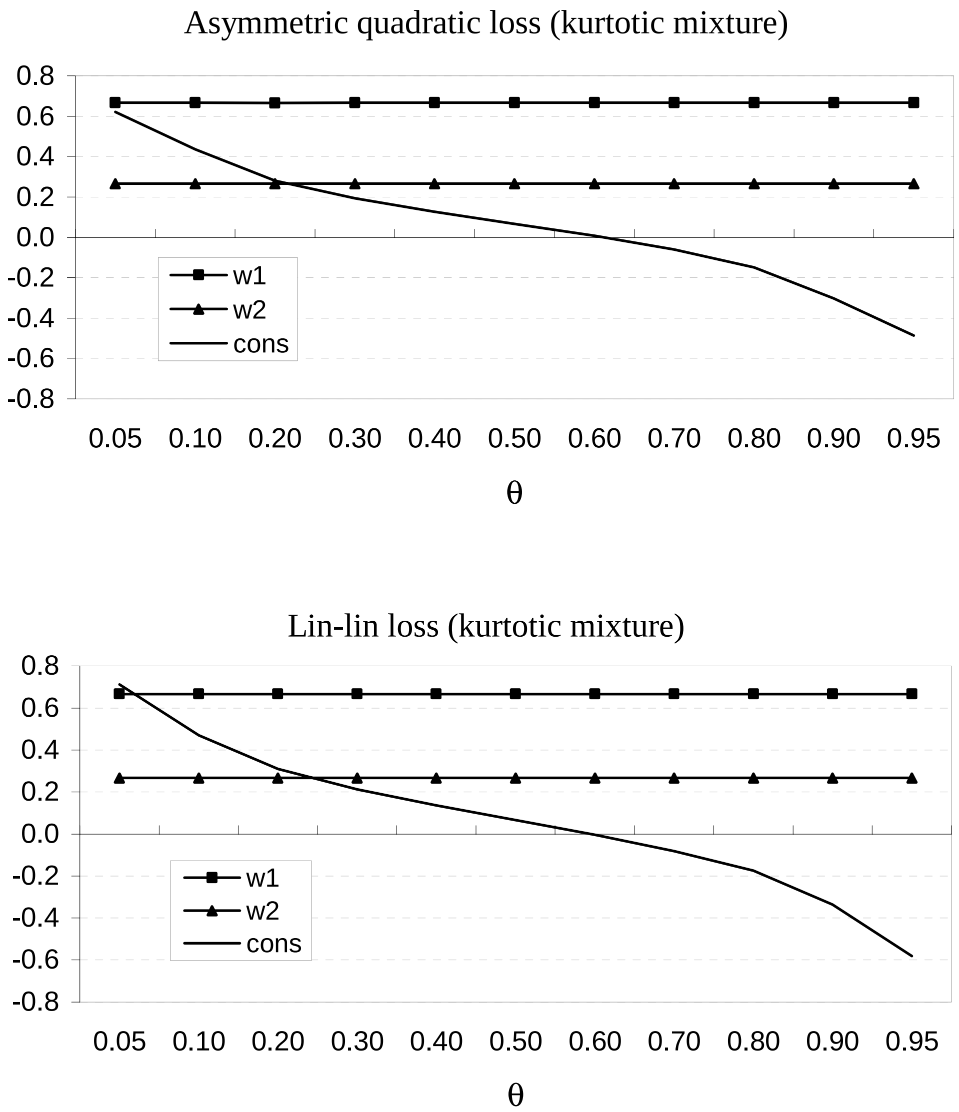


Figure 4: Time series of forecast errors

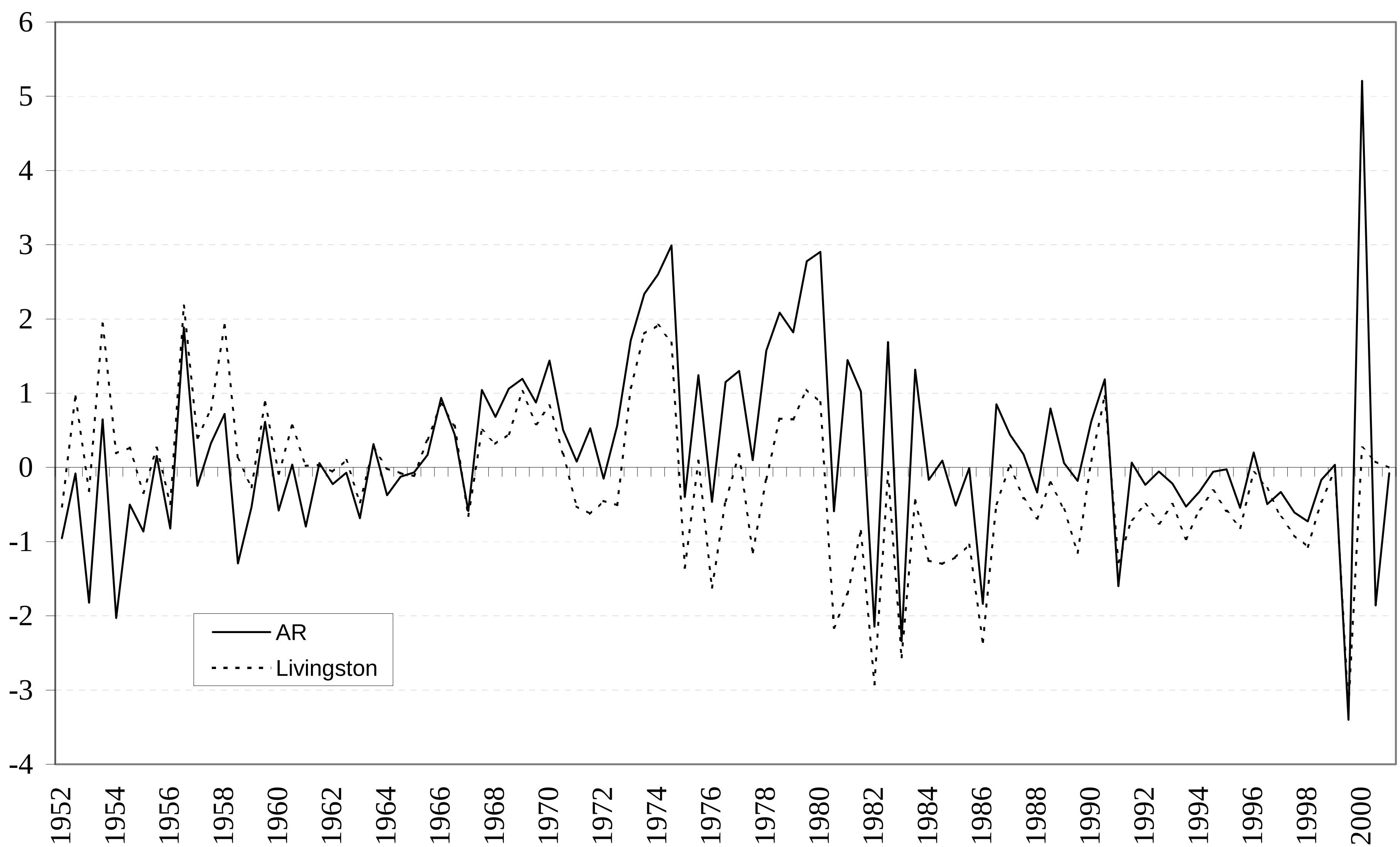
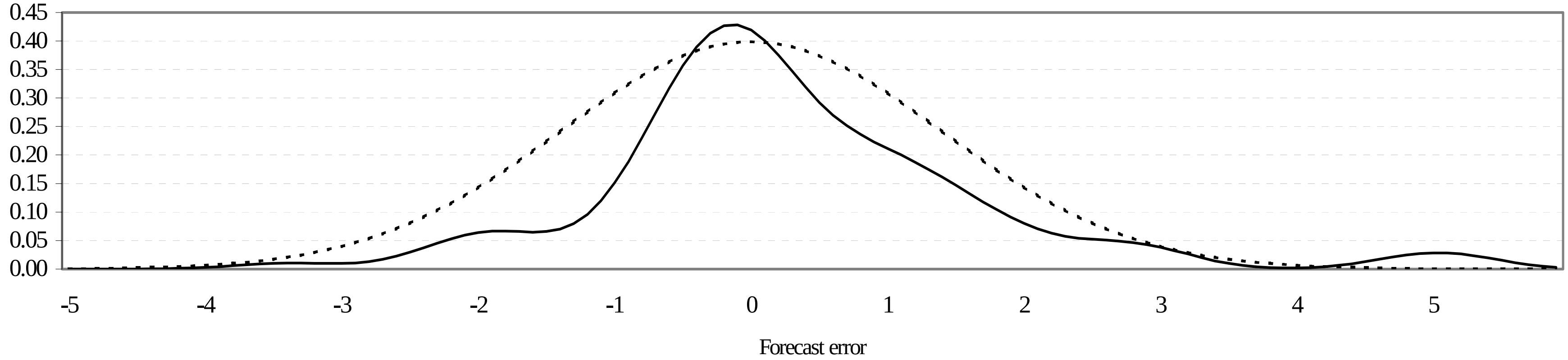


Figure 5: Forecast error densities from individual models

Density of forecast error from AR model



Density of forecast error from Livingston data

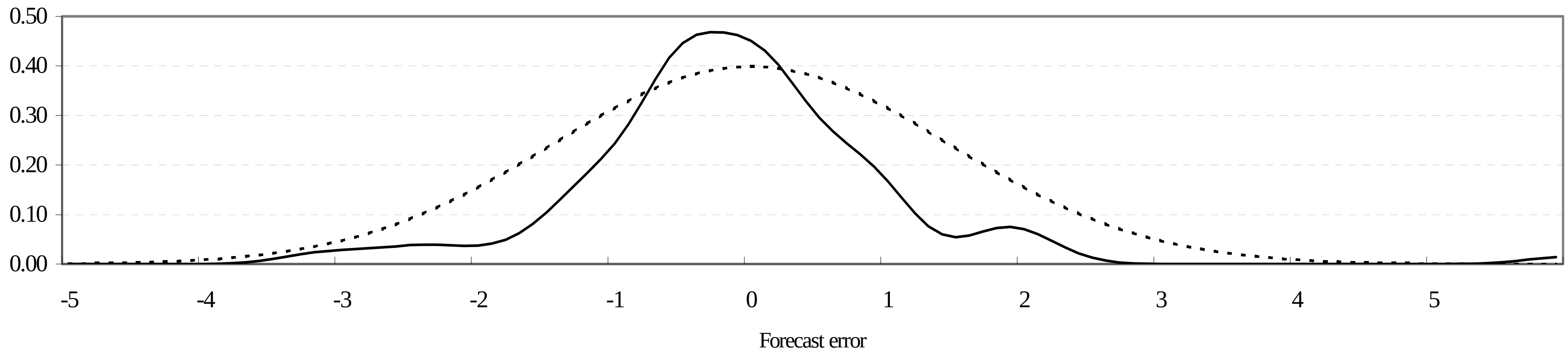
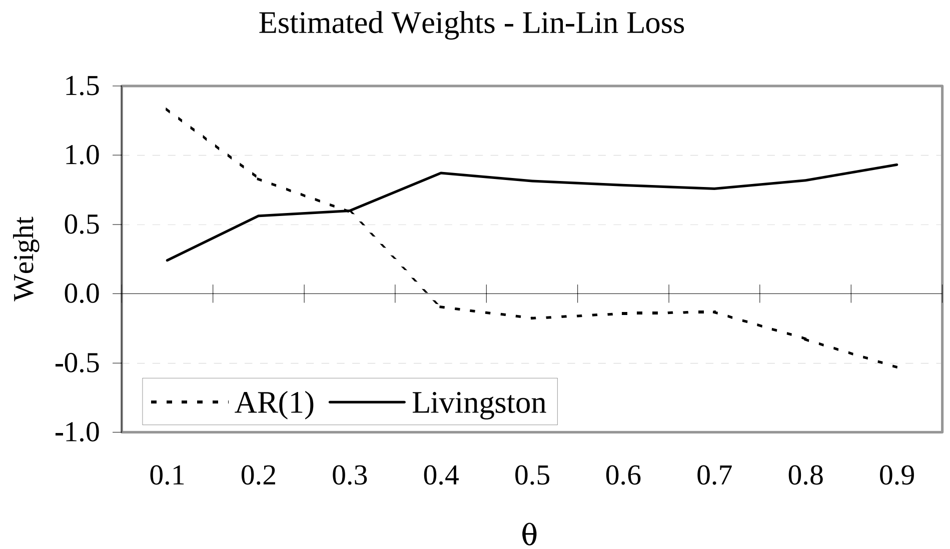


Figure 6: Forecast combination weights under different loss functions



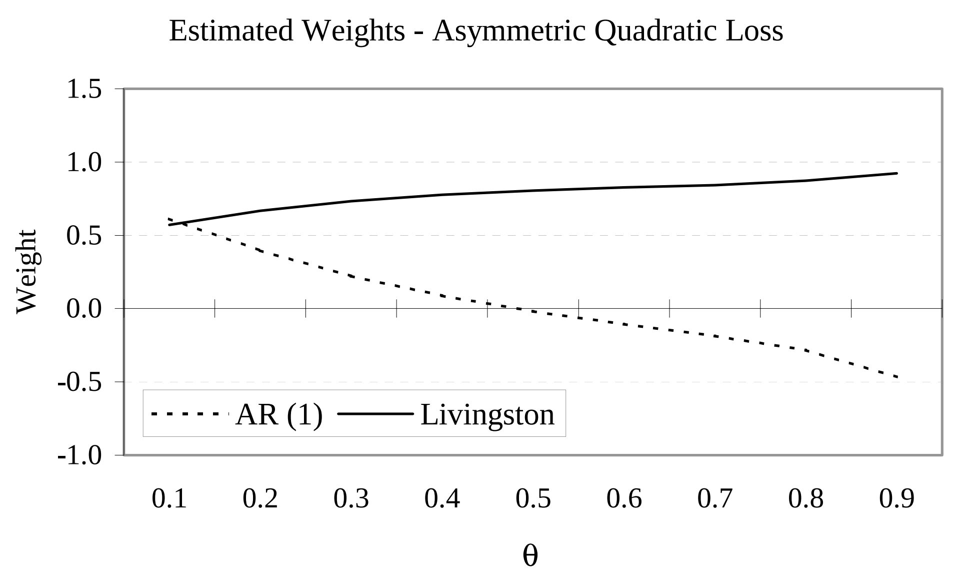


Figure 7: Forecast error densities from combined model

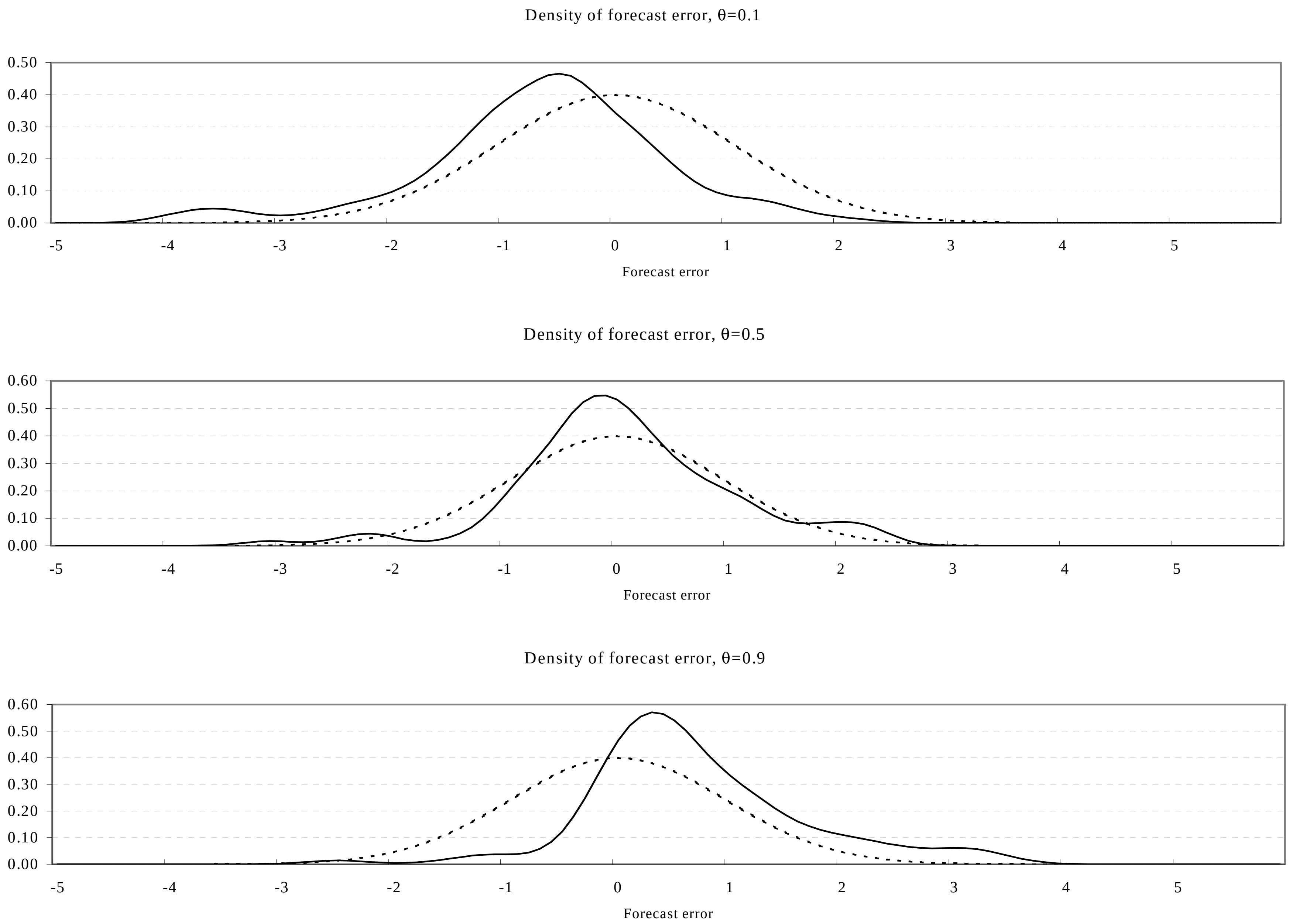


Figure 8: Disaggregated forecast error densities