Estimating Loss Functions of Experts

Philip Hans Franses[*]
Rianne Legerstee
Richard Paap

Econometric Institute
Erasmus University Rotterdam

Econometric Institute Report 2011-42

December 15, 2011

Abstract

We propose a new and simple methodology to estimate the loss function associated with experts' forecasts. Under the assumption of conditional normality of the data and the forecast distribution, the asymmetry parameter of the lin-lin and linex loss function can easily be estimated using a linear regression. This regression also provides an estimate for potential systematic bias in the forecasts of the expert. The residuals of the regression are the input for a test for the validity of the normality assumption.

We apply our approach to a large data set of SKU-level sales forecasts made by experts and we compare the outcomes with those for statistical model-based forecasts of the same sales data. We find substantial evidence for asymmetry in the loss functions of the experts, with underprediction penalized more than overprediction.

Keywords: model forecasts; expert forecasts; loss functions; asymmetry; econometric models

[*]Corresponding author: Erasmus University Rotterdam, Econometric Institute, H 11-34, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, e-mail: franses@ese.eur.nl

1 Introduction

Sales forecasts are often the outcome of a process in which an expert with domain-specific knowledge modifies a model-generated forecast. Typically, simple extrapolation models are used to create such model forecasts, and often they are generated by automated statistical software which gets fed by lagged sales and other possibly relevant variables.

There is a long tradition in the sales forecasting literature to examine the quality of these expert forecasts relative to model forecasts (if these are available). Key questions are whether the domain-specific knowledge translates into improved forecasts, or whether experts downplay the model forecasts too much, thereby quoting less accurate forecasts. Classical studies are Blattberg and Hoch (1990) and Mathews and Diamantopoulos (1986) where various case studies are examined.

Recently this literature has seen a revived interest with the advent of a range of large data sets that allow for more generalizing statements. For example, Fildes et al. (2009) study thousands of expert and model forecasts, and conclude that expert forecasts tend to be biased and that expert forecasts are not necessarily better than model forecasts. Franses and Legerstee (2010), using a database with over 30,000 forecasts and realizations, show that, on average, model forecasts and expert forecasts are about equally good, but when expert forecasts are worse they are much worse.

A common finding in these two recent studies is that expert forecasts tend to exceed model forecasts, or in other words, judgmental adjustment is often positive. A potential explanation for this finding is that the experts dislike underpredicting more than overpredicting, perhaps due to planning reasons. Hence, when creating forecasts their loss function may not be a mean squared error (MSE) loss function symmetric around zero, but some other, asymmetric loss function. If such an alternative loss function is used indeed, this may then also explain why expert forecasts seem less accurate than model forecasts, as typically forecasts are evaluated using criteria like the root mean squared prediction error (RMSPE).

The loss function of experts is usually not known in practice. Given available data, one may however try to estimate this loss function by evaluating theoretical properties of loss functions against actual data. Various forms of asymmetric loss functions have been proposed in the literature, like, for example, the lin-lin loss function, the quad-quad loss function and the linex function proposed by Varian (1975). These loss functions have been frequently analyzed, for example, by analyzing the optimal forecast under a specific asymmetric loss function, see Zellner (1986) and Christoffersen and Diebold (1996, 1997), among others.

In this paper we are interested in estimating the parameters of loss functions given the availability of expert forecasts. Clatworthy et al. (2011) investigate whether financial analysts' loss functions are asymmetric or not, but they do not estimate the loss function. A notable exception is Elliott et al. (2005). These authors propose a linear Instrumental Variable (IV) estimator for the shape parameter of a general class of loss functions which signals the degree of asymmetry in the loss function. The general class of loss functions nests four popular loss functions, and these are the absolute deviation loss function and its asymmetric counterpart the lin-lin loss function, and the squared loss function and its asymmetric counterpart the quad-quad loss function. They use their methodology to estimate the asymmetry in forecasts of budget deficits for the G7 countries made by the IMF and OECD.

To estimate the loss function of experts in the sales forecasting industry we propose a methodology that differs from the methodology proposed by Elliott et al. (2005) in a number of ways. By making a normality assumption on the conditional distribution of the variable to be forecasted, and by that on the forecast distribution, we demonstrate that the estimation of the asymmetry parameter is simplified substantially. Elliott et al. (2005) need instrumental variables for their estimation method, but in our proposed methodology only simple linear regressions (OLS) are used, using panel data on expert forecasts and on the variable to be forecasted. If the normality assumption is valid, OLS is more efficient than using instrumental variables and the methodology can easily be extended to multiple-step ahead forecasts. Our proposed method can be used to estimate the key parameters of the well-known and useful linex loss function.

The outline of our paper is as follows. In Section 2 we show that for two well-defined loss functions, the lin-lin loss function and the linex loss function, simple regressions can be used to estimate the asymmetry parameter of the functions, provided the availability of the relevant data. In Section 3 we illustrate this methodology for a large database covering forecasts from a range of experts. We also consider statistical model forecasts to establish to what extent symmetric loss functions prevail. The robustness of our crucial assumption on the forecast distribution is tested in three ways. One way, for example, is to compare our estimates with those obtained with the methodology of Elliott et al. (2005). Upon estimating our two loss functions we find overwhelming support for the conjecture that experts may feel that negative forecast errors (meaning the forecasts are below actual sales) require more weight in the loss function than positive forecast errors. Section 4 concludes this paper with a summary and suggestions for further research.

2 Loss Functions

Suppose that Y_{t+1} is the random variable to be forecasted with forecast density $f(y_{t+1};\theta,\mathcal{Y}_t,\mathcal{X}_t)$ that may depend on parameters θ and lagged values $\mathcal{Y}_t = \{y_{t+1-j}\}_{j=1}^{J}$ and other exogenous variables summarized in $\mathcal{X}_t$. To simplify notation we write $f(y_{t+1};\theta)$ instead of $f(y_{t+1};\theta,\mathcal{Y}_t,\mathcal{X}_t)$. In this paper we confine our analysis to one-step ahead forecasts.

Given the forecast distribution, a point forecast p_{t+1} for Y_{t+1} can be obtained by specifying a loss function. For example, the quadratic loss function is given by

$$\mathrm{QL}(Y_{t+1}, p_{t+1}) = (p_{t+1} - Y_{t+1})^2, \tag{1}$$

where we adopt the convention that a forecast error is the forecast minus the realization. The point forecast $\hat{p}_{t+1}$ results from minimizing expected quadratic loss $\mathrm{E}[\mathrm{QL}(Y_{t+1}, p_{t+1})]$ with respect to p_{t+1}, where E denotes the expectation operator. In case of quadratic loss, this results in $\hat{p}_{t+1} = \mathrm{E}[Y_{T+1}|\theta]$. Hence, the optimal forecast is unbiased.

From a supply chain management point of view it can be necessary to put a higher penalty on negative forecast errors than on positive forecast errors. For example, if one forecasts sales, the consequences of a prediction which is lower than the realized demand may be worse than a prediction which is higher than the demand. In other words, being out of stock is worse than having a little too much stock. To allow for different penalties one may then consider an asymmetric loss function.

2.1 Asymmetric absolute loss function

An example of an asymmetric function is the lin-lin loss function, further also called the asymmetric absolute loss (AAL) function, which is given by

$$\mathrm{AAL}(Y_{t+1}, p_{t+1}) = \begin{cases} \alpha_A |p_{t+1} - Y_{t+1}| & \text{if } p_{t+1} \leq Y_{t+1} \\ |p_{t+1} - Y_{t+1}| & \text{if } p_{t+1} > Y_{t+1} \end{cases} \tag{2}$$

One sets $\alpha_A > 1$ if one wants to put more penalty on a forecast which is smaller than the true realization, see also Ferguson (1967). The optimal point forecast is obtained by minimizing expected loss, that is,

$$\mathrm{E}[\mathrm{AAL}(Y_{t+1}, p_{t+1})] = \int \mathrm{AAL}(y_{t+1}, p_{t+1}) f(y_{t+1}; \theta) dy_{t+1}. \tag{3}$$

The expected loss function $\mathrm{E}[\mathrm{AAL}(Y_{t+1}, p_{t+1})]$ can be written as

$$\int_{-\infty}^{p_{t+1}} (p_{t+1} - y_{t+1}) f(y_{t+1}; \theta) dy_{t+1} + \int_{p_{t+1}}^{\infty} \alpha_A (y_{t+1} - p_{t+1}) f(y_{t+1}; \theta) dy_{t+1}. \tag{4}$$

The first-order partial derivative is given by

$$\begin{aligned} \frac{\partial \mathrm{E}[\mathrm{AAL}(Y_{t+1}, p_{t+1})]}{\partial p_{t+1}} &= \int_{-\infty}^{p_{t+1}} f(y_{t+1}; \theta) dy_{t+1} - \int_{p_{t+1}}^{\infty} \alpha_A f(y_{t+1}; \theta) dy_{t+1} \\ &= F(p_{t+1}; \theta) - \alpha_A (1 - F(p_{t+1}; \theta)), \end{aligned} \tag{5}$$

where we used the Leibniz integral rule and where $F(\cdot; \theta)$ is the forecast distribution function of Y_{t+1} (with $f(\cdot; \theta)$ as its derivative). The optimal point forecast is obtained when this derivative is set equal to zero and solved for p_{t+1}, which results in

$$F(\hat{p}_{t+1}; \theta) = \frac{\alpha_A}{1 + \alpha_A}. \tag{6}$$

The point estimate corresponds to the $\alpha_A/(1+\alpha_A)$th percentile of the forecast distribution. Under symmetric loss ($\alpha_A = 1$) we obtain the median of the forecast distribution. For $\alpha_A > 1$ we have a forecast which is larger than the median, and for $\alpha_A < 1$ we obtain a forecast which is smaller than the median.

Hence, apparent biased forecasts of an expert may be due to the fact that an asymmetric loss function is used. Our main claim in this paper is that if we were to observe several forecasts of experts together with realizations of the forecasts, it is possible under some testable assumptions to estimate the value of α_A, see also Subsection 2.3 below.

Suppose that we have data with T forecasts where for each point forecast created at time $t = 1, ..., T$ the conditional forecast distribution is normal with mean m_t and variance s_t^2. Furthermore, assume that all forecasts are constructed using the same asymmetric absolute loss function. Under these assumptions the forecasts are thus generated by

$$p_{t+1} = m_t + s_t \Phi^{-1}\left(\frac{\alpha_A}{1+\alpha_A}\right), \tag{7}$$

where Φ^{-1} is the inverse CDF of a standard normal distribution.

Further assume that the realizations y_{t+1} result from a normal distribution with mean μ_t and variance σ_t^2 for $t = 1, \ldots, T$ and hence $y_{t+1} = \mu_t + \sigma_t \eta_t$, where η_t is a realized draw from a standard normal distribution. If there is a systematic bias in the forecast distribution it holds that $m_t = \mu_t + b$ with $b \neq 0$. If we consider the difference between p_{t+1} and y_{t+1} we obtain

$$(p_{t+1} - y_{t+1}) = b + s_t \Phi^{-1}\left(\frac{\alpha_A}{1+\alpha_A}\right) - \sigma_t \eta_t. \tag{8}$$

If we can obtain a consistent estimate of s_t and σ_t, one can use the simple regression

$$\frac{(p_{t+1} - y_{t+1})}{\hat{\sigma}_t} = \frac{1}{\hat{\sigma}_t}\beta_0 + \frac{\hat{s}_t}{\hat{\sigma}_t}\beta_1 + \varepsilon_t \tag{9}$$

to provide the estimate for $\beta_0 = b$ and for $\beta_1 = \Phi^{-1}(\alpha_A/(1+\alpha_A))$. An estimate of α_A can easily be obtained by solving

$$\frac{\alpha_A}{1+\alpha_A} = \Phi(\beta_1) \Rightarrow \alpha_A = \frac{\Phi(\beta_1)}{1-\Phi(\beta_1)}.$$

In sum, in this scenario it is possible for a forecaster to have an asymmetric loss function and a systematic bias in its forecasting distribution. The expression in (9) shows that it is possible to calibrate the loss function and the bias.

2.2 Linex loss function

An alternative nonlinear asymmetric loss function is the linear-exponential function, also called the linex (LIN) loss function, see Varian (1975) and Zellner (1986). This function is given by

$$\text{LIN}(Y_{t+1}, p_{t+1}) = \exp(\alpha_L(p_{t+1} - Y_{t+1})) - \alpha_L(p_{t+1} - Y_{t+1}) - 1 \tag{10}$$

with $\alpha_L \neq 0$. A negative value of α_L implies that a p_{t+1} lower than Y_{t+1} is more costly than a p_{t+1} higher than Y_{t+1}. To be more precise, if $\alpha_L < 0$, the linex loss function shows an almost exponential increase in loss to the left of the origin $(p_{T+1} - Y_{t+1}) = 0$ and an almost linear increase in loss to the right of the origin. A positive value of α_L implies the opposite and a $\alpha_L \to 0$ implies symmetric loss. Zellner (1986) shows that the point forecast which minimizes expected loss is given by

$$\hat{p}_{t+1} = -\alpha_L^{-1} \log \text{E}[\exp(-\alpha_L Y_{t+1})]. \tag{11}$$

Hence, if we assume that the forecast distribution of Y_{t+1} is normal with mean m_t and variance s_t^2, then the point forecast is given by

$$p_{t+1} = m_t - \frac{1}{2}\alpha_L s_t^2. \tag{12}$$

Again it is possible to estimate α_L in case we observe several forecasts of experts together with realized forecasts. Under the same conditions as above and using the same arguments, taking the difference between p_{t+1} and y_{t+1}

and dividing by $\hat{\sigma}_t$ results in the simple regression

$$\frac{(p_{t+1} - y_{t+1})}{\hat{\sigma}_t} = \frac{1}{\hat{\sigma}_t}\beta_0 + \frac{\hat{s}_t^2}{2\hat{\sigma}_t}\hat{s}_t^2\beta_1 + \varepsilon_t. \tag{13}$$

OLS provides the estimate for the systematic bias $b = \beta_0$ and asymmetry parameter $\alpha_L = -\beta_1$.

2.3 Parameter estimation

To run the regressions (9) and (13) we need estimates of s_t^2 and σ_t^2. If we have the availability of unbiased model forecasts ($mf_{t+1} = \mathrm{E}[y_{t+1}|\mathcal{Y}_t, \mathcal{X}_t]$) and the realizations y_{t+1}, the variance of the data can be estimated using

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_{t+1} - mf_{t+1})^2 \tag{14}$$

under the assumption that $\sigma_t^2 = \sigma^2$.

The variance of the forecast distribution of the expert s_t^2, however, cannot be estimated from the variance of the available expert forecasts as these forecasts may be biased and/or result from an asymmetric loss function. To estimate the variance we assume that s_t^2 is constant ($s_t^2 = s^2$ for $t = 1, \ldots, T$) and that the variance of the expert is equal to the variance of the forecast distribution of an econometric model which fits the data at hand and produces unbiased forecasts.

Because σ_t and s_t are constant over t we need panel data with expert forecasts and realizations in order to estimate the parameters in (9) and (13). In other words, if we have point forecasts for variables $i = 1, ..., N$ over periods $t = 1, ..., T$, denoted by $p_{i,t+1}$ and $mf_{i,t+1}$, we are able to estimate $\hat{s}_i^2$ and $\hat{\sigma}_i^2$ for each i. In case of the lin-lin loss function we can now estimate the bias and asymmetry parameter with the regression

$$\frac{(p_{i,t+1} - y_{i,t+1})}{\hat{\sigma}_i} = \frac{1}{\hat{\sigma}_i}\beta_0 + \frac{\hat{s}_i}{\hat{\sigma}_i}\beta_1 + \varepsilon_{i,t}, \tag{15}$$

where $b = \beta_0$ and $\alpha_A = \Phi(\beta_1)/(1-\Phi(\beta_1))$. In case of the linex function we can estimate the bias and asymmetry parameter with

$$\frac{(p_{i,t+1} - y_{i,t+1})}{\hat{\sigma}_i} = \frac{1}{\hat{\sigma}_i}\beta_0 + \frac{\hat{s}_i^2}{2\hat{\sigma}_i}\beta_1 + \varepsilon_{i,t}, \tag{16}$$

where $b = \beta_0$ and $\alpha_L = -\beta_1$.

2.4 Misspecification

Under our assumptions the error terms $\varepsilon_{i,t}$ for $i = 1, ..., N$ and $t = 1, ..., T$ should be normal with mean 0 and variance 1 in regressions (15) and (16). If this is not the case, (some of) the assumptions, such as the assumption of a normal forecast distribution, may not be valid or the loss function may not be adequate. It is therefore important to test if the estimated residuals are standard normally distributed.

If tests show that the error terms are not standard normally distributed or if there are other reasons to doubt whether the forecast density is normal, it is also possible to assume that the forecasts are lognormally distributed in case of the lin-lin loss function (AAL). Under this distribution, the forecasts are generated by

$$\log(p_{t+1}) = m_t + s_t\Phi^{-1}\left(\frac{\alpha_A}{1+\alpha_A}\right), \tag{17}$$

where m_t is the mean and s_t^2 the variance of $\log(p_{t+1})$ and Φ^{-1} is the inverse CDF of a standard normal distribution. Assume now that the realizations y_{t+1} result from a lognormal distribution with parameters μ_t and σ_t^2 for $t = 1, \ldots, T$ and hence $\log(y_{t+1}) = \mu_t + \sigma_t\eta_t$ where ε_t is a realized draw from a standard normal distribution. We can now write

$$(\log(p_{t+1}) - \log(y_{t+1})) = b + s_t\Phi^{-1}\left(\frac{\alpha_A}{1+\alpha_A}\right) - \sigma_t\eta_t, \tag{18}$$

where b is again the systematic bias in the forecast distribution, thus $m_t = \mu_t + b$. Using the estimates of s_t and σ_t and using the relevant panel data the regression

$$\frac{(\log(p_{i,t+1}) - \log(y_{i,t+1}))}{\hat{\sigma_i}} = \frac{1}{\hat{\sigma_i}}\beta_0 + \frac{\hat{s}_i}{\hat{\sigma_i}}\beta_1 + \varepsilon_{i,t} \tag{19}$$

provides $\beta_0 = b$ and $\beta_1 = \Phi^{-1}(\alpha_A/(1+\alpha_A))$ and hence $\alpha_A = \Phi(\beta_1)/(1 - \Phi(\beta_1))$. Again, if the assumptions are correct, including the assumption of lognormality of the forecast distribution, and the loss function is AAL, the error terms $\varepsilon_{i,t}$ for $i = 1, ..., N$ and $t = 1, ..., T$ should be normal with mean 0 and variance 1.

Another way to check if the assumptions are correct is to compare the results for the AAL loss function with the results as found with the method of Elliott et al. (2005). They use as a general loss function

$$L(Y_{t+1}, p_{t+1}) = [\alpha_E + (1 - 2\alpha_E) \cdot I(Y_{t+1} - p_{t+1} < 0)]|Y_{t+1} - p_{t+1}|^q, \tag{20}$$

where $I[\cdot]$ is an indicator function which takes a value of 1 if the statement between brackets is true and is 0 otherwise, where $\alpha_E \in (0, 1)$ and where they impose $q = 1$ or $q = 2$. By setting $q = 1$, the AAL loss function is obtained as defined above in (2), but with weight α_E for cases where $p_{t+1} \leq Y_{t+1}$ and with weight $1 - \alpha_E$ for cases where $p_{t+1} > Y_{t+1}$. Stated differently, $\alpha_E/(1 - \alpha_E) = \alpha_A$. Elliott et al. (2005) do not make assumptions on the distribution of the forecasts. Therefore, if the normality assumption is valid, their methodology should result in an $\hat{\alpha}_E$ for which $\hat{\alpha}_E/(1-\hat{\alpha}_E) \approx \hat{\alpha}_A$, where $\hat{\alpha}_A$ is obtained from (15). Differences between $\hat{\alpha}_E/(1-\hat{\alpha}_E)$ and $\hat{\alpha}_A$ might be a result of the chosen instrumental variables for the estimation of α_E or the use of $\hat{s}$ and $\hat{\sigma}$ instead of s and σ for the estimation of α_A or both.

In the next section we will illustrate the techniques and robustness checks described in this section for a range of forecasts made by many experts.

3 Illustration

We apply our methodology to an extensive panel data set. The data set covers SKU-level sales data and is described in detail in the next subsection. In Subsections 3.2 and 3.3 the results of our analysis are discussed.

3.1 Data set

For our case study we use monthly sales data of a large pharmaceutical company. The company has its headquarters in The Netherlands, and has local offices in various countries worldwide. The company uses an automated statistical package to create forecasts using lagged sales data as the only input. The experts know that these data are the only input. Each month model selection and parameter estimation are updated, whereby the package uses techniques such as Box-Jenkins and Holt-Winters. These model forecasts are then sent to the managers/experts in the local offices, after which they quote their own forecasts.

The forecasts are available for the months November 2004 through November 2006. They are created for various horizons, but we only use the 1-step-ahead forecasts in the analysis presented in this paper. In each country, forecasts are created by a different expert and hence we have forecasts for 35 countries and thus 35 distinct individuals. For confidentiality reasons we denote the countries with roman numbers I to XXXV. Forecasts are created for 1038 different products. In the notation of the previous section this means that i ranges from 1 to 1038. Per product we have a minimum of 15 and a maximum of 25 observations for which the model forecast, the expert forecast and realized sales are available to us. Thus, T depends on i and $15 \leq T_i \leq 25$ for $i = 1, ..., 1038$. All together, we have 24897 observations.

We denote the model forecasts as constructed by the statistical program of the company as MF, and the final forecasts from the experts are denoted as EF. The model that we use to estimate σ_i and s_i is for each i an AR(1) model for which the parameters are estimated over all available observations for i. For $mf_{i,t+1}$ $\forall i$ and $\forall t$ we consider the in-sample forecasts generated by these AR(1) models. Note that $MF_{i,t+1}$ and $mf_{i,t+1}$ are different forecasts, the first is the statistical model forecast as used by the company and the second is the forecast from the AR(1) model used to estimate σ_i and s_i.

The parameters in (15), (16) and (19) are estimated for each expert separately by multiplying the two variables in the regressions by dummy variables for the managers. We also estimate α_E per expert. Observations per expert range from 96 to 2132 with an approximate average of 710 observations.

3.2 Estimated asymmetry

We begin by analyzing the results as obtained under the assumption that the AAL function is used by the experts. Column 2 in Table 1 presents the estimated asymmetry parameter α_A per expert. We see that 26 of the 35 experts have an $\hat{\alpha}_A$ that is significantly different from 1 at a significance level of 10%. For 21 of these managers the difference is even significant at the 1% significance level. For all those 26 managers the $\hat{\alpha}_A$ exceeds 1, meaning that sales forecasts that are too low are penalized more than forecasts that are too high. On average, over 35 experts, $\hat{\alpha}_A$ has a value of 1.40, which indicates that too low forecasts are weighted 40% heavier than too high forecasts. To get some more insight into this value for $\hat{\alpha}_A$, see Figure 1.

[INSERT TABLE 1 AND FIGURE 1 ABOUT HERE]

The estimated systematic bias b for each expert can be found in Column 3 of Table 1. There are 11 experts with a significant systematic bias at the 1% significance level and another 2 experts with a significant systematic bias at the 5% significance level. Most of these are positive biases and most are linked to a significantly positive asymmetry parameter.

If we only take the 1% significance level into consideration, we can conclude that 15 experts have an asymmetric loss function, but no systematic bias. Another 6 experts have an asymmetric loss function and also a systematic bias. Only 5 experts have a systematic bias and no asymmetric loss function, and finally, only 9 experts seem to have a symmetric loss function and no systematic bias.

When we apply the test regression to the model forecasts MF, we obtain the results as reported in Columns 4 and 5 of Table 1. As we might expect from model forecasts based on techniques such as Box-Jenkins and Holt-Winters, we find much less evidence of asymmetry in the loss function and of systematic bias. For only 8 countries the $\hat{\alpha}_A$ are significantly different from 1 at the 1% significance level and in another 1 at the 5% significance level. The average of the 35 $\hat{\alpha}_A$ values is 1.03, which is very close to 1. Some evidence of systematic bias is found in 12 countries, but at the 1% significance level only 4 of these cases remain. In sum, the model-based forecasts in general seem unbiased and have been created using a symmetric loss function.

Now we turn to the results when we assume that the linex loss function is used by the experts. See Table 2 for the estimated asymmetry parameters and systematic biases again for both EF and MF. In the second column of this table we find $\hat{\alpha}_L$ for each expert. For 18 experts we find an $\hat{\alpha}_L$ significantly different from 0 (thus asymmetry) at a significance level of 10%. For 12 of these is the difference also significant at 1%. So this is almost half of the cases where we found asymmetry for the AAL function. All except 1 (which is only significant at the 10% level) have a negative asymmetry parameter, indicating that again negative forecast errors weigh more heavy than positive forecast errors. All except 2 (which are both again only significant at the 10% level) were also found to have an asymmetric loss function under AAL. On average, $\hat{\alpha}_L$ has a value of -0.0002. See Figure 2 for the shape of LIN with an α_L equal to this average estimate.

[INSERT TABLE 2 AND FIGURE 2 ABOUT HERE]

However, we do find more often a significant systematic bias under the linex loss function than under the lin-lin loss function, see Column 3 of Table 2. 22 experts have a $\hat{b}$ significantly different from 0 at the 10% significance level and for 16 of them is this difference also significant at the 1% level. In some instances the linear asymmetry as found under AAL seems to be replaced by a (more profound) systematic bias, see for example the experts denoted with IV, XX and XXX. In general, the bias is positive again.

In sum, we find that at the 1% significance level there are far more experts with a symmetric loss function (23) than with an asymmetric loss function (12) if we assume the linex loss function. 12 of the experts with a symmetric loss function also do not have a systematic bias, although 16 experts have a systematic bias. Results are also a bit more ambiguous, because there are more countries for which significant asymmetry and/or bias is found with the 5%- or 10%-significance level and not with the 1% significance level, as compared to the AAL situation.

Finally, we also compare these linex results for EF with the linex results for MF, see Columns 4 and 5 of Table 2. Again we do not find much evidence for asymmetry and systematic bias in the model forecasts. $\hat{\alpha}_L$ is on average -4.13e-06, so much closer to 0 than the average $\hat{\alpha}_L$ of -0.0002 found for EF. For only 10 countries is the asymmetry parameter significantly different from 0 at the 10% level and in only 3 countries at the 1% level. The number of significant systematic biases is 16 at the 10% level and 6 at the 1% level. So again these results confirm that statistical model forecasts are unbiased and derived from a symmetric loss function.

3.3 Specification checks

So far, we have analyzed the results given the assumptions underlying the analysis. To test these assumptions we now follow the strategy as outlined in Section 2.4.

The first step is to check if the error terms of the regression models (15) for AAL and (16) for LIN are standard normally distributed. To that end, we use the Kolmogorov-Smirnov test, see D'Agostino and Stephens (1986). The test is performed on the error terms of each country separately, so we have 35 test results. In the second and third column of Table 3 we see how often these 35 tests reject the null hypothesis of standard normally distributed error terms at the 1% significance level. For the asymmetric absolute loss function we see fairly low figures. For EF we see that in a little bit over one-third of the tests the null hypothesis is rejected and for MF this is a little bit over one-fifth. Note that the number of observations on which the normality test is performed is quite large (see Subsection 3.1), which means that the test has high power against tiny deviations of standard normality. For countries with many observations we may therefore choose an even lower significance level which implies that the number of rejections may even be lower.

[INSERT TABLE 3 ABOUT HERE]

For the linex loss function we find much higher numbers of rejection, namely 23 (66%) for EF and 12 (34%) for MF. As the numbers for AAL are much lower, this might indicate that we should not reject the assumption of a normal forecast distribution at this point, but that the assumption of a linex loss function is perhaps not an appropriate assumption. The AAL function seems to be the loss function that is more likely to be used by the managers creating the forecasts in this data set.

As we deal with sales forecasts in this application, which are always positive, it might be resonable to assume that the forecasts are lognormally distributed instead of normally. Therefore, we also estimate (19), again with separate coefficients for each country, and again we test if the error terms are standard normally distributed. We find overwhelming evidence that the forecast distribution is not lognormal, see Column 4 of Table 3. Both for EF and MF the null hypothesis of standard normal error terms is rejected for all 35 countries. This again indicates that assuming a normal forecast distribution seems acceptable for our data.

Our final specification check involves a comparison of our AAL results with those upon using the method of Elliott et al. (2005). Table 4 and Figures 3 and 4 give the results. Note that Columns 2 and 4 of Table 4 are the same as Columns 2 and 4 in Table 1, but are repeated for ease of comparison. Columns 3 and 5 present the results as obtained using the method of Elliott et al. (2005), where we used as instrumental variables a constant and one-month lagged sales. Remember that we expect $\hat{\alpha}_A$ and $\hat{\alpha_E}/(1-\hat{\alpha_E})$ to be approximately the same if the assumptions for our method are correct.

[INSERT TABLE 4 ABOUT HERE]

First note, from Table 4, that whenever $\hat{\alpha_A}$ is significantly larger than 1 at each significance level, $\hat{\alpha_E}/(1-\hat{\alpha_E})$ is never significantly smaller than 1 at each significance level. Furthermore, whenever $\hat{\alpha_A}$ is significantly smaller than 1 at the 1%-, 5%- or 10%-significance level (happens only twice for MF), $\hat{\alpha_E}/(1-\hat{\alpha_E})$ is never significantly larger than 1 at the 1%-, 5%- or 10%-significance level. Both statements also hold true when $\hat{\alpha_E}/(1-\hat{\alpha_E})$ is evaluated against $\hat{\alpha_A}$. These results indicate that we never find fully conflicting results with the two alternative methods.

The largest difference in results appears when we sometimes find a significant asymmetry with one method and no significant asymmetry with the other method. If we focus on the 1% significance level, this happens 8 times for EF and 2 times for MF, but in most of these cases (7) the other method also shows asymmetry at the 5% or 10% level. Hence, we find that both methods may differ in terms of the amount of asymmetry, but not in the sign of the asymmetry and hardly in the existence of the asymmetry.

To get a more precise idea of the size of the differences in estimated asymmetry parameters, we can take a look at the histograms in Figures 3 and 4. Here the differences between $\hat{\alpha}_A$ and $\hat{\alpha_E}/(1-\hat{\alpha_E})$ are depicted, for EF in the first figure and for MF in the second. Multiplying the differences with 100% shows the differences in percentages. Thus for example, a value of 0.1 indicates that the difference in weight between too low forecasts and too high forecasts is 10% higher according to $\hat{\alpha}_A$ than according to $\hat{\alpha}_E$.

[INSERT FIGURES 3 and 4 ABOUT HERE]

Although we see some outliers in the graphs, the largest one being that $\hat{\alpha_E}$ is 97% larger than $1 - \hat{\alpha_E}$ than that $\hat{\alpha}_A$ is larger than 1, on average the difference is around to be equal to 6% (0.06 in the figure). Furthermore, in 23 of the 35 countries the difference is smaller than 25% in absolute sense and in 31 of the 35 countries the difference is smaller than 50%. For MF, see Figure 4, these differences are even smaller, with an average difference of around 2.5% and a maximum difference of around 51%, both in absolute terms.

The larger differences do not necessarily seem to be related with the rejection of the standard normality of the residuals of the regression. The correlation between the absolute difference and the p-value of the Kolmogorov-Smirnov test is -0.14 for EF and -0.06 for MF. If we look at the test results for some countries with large differences in estimation results, we sometimes find rejection of the null hypothesis and sometimes we do not.

To conclude, we do not find large differences in the results of both methods and we take this as a final indication that the assumptions underlying our analysis do not need to be rejected, at least, for our data set at hand.

4 Conclusions

There is much available research on asymmetric loss functions for forecasters, but most of it is focussed on the theoretical discussion of possible shapes of those loss functions and on resulting optimal forecasts. Very little is known about which loss function is actually exercised by experts when they create their forecasts and rarely it is quantified to what extent the loss functions are asymmetric. We are aware of one study only, and this is presented in

Elliott et al. (2005).

In the present paper we propose a new and simple methodology to deduce the asymmetry parameter of the asymmetric absolute loss function and of the linex loss function. The derivation is based on some simplifying assumptions which can be held against actual data in a number of ways. The derivations were shown to lead to simple linear regressions.

We applied our methodology to a large data set of SKU-level sales forecasts where model forecasts are received by experts, after which they provided their final forecasts. We documented substantial evidence that the experts use an asymmetric loss function, where we diagnosed that most likely it is the asymmetric absolute loss function. Forecasts that are too low have a weight in the loss function that is on average 40% higher than forecasts that are too high.

The methodology proposed in this paper results in similar results as found with the methodology of Elliott et al. (2005) and in general we find no obvious indications that the assumptions underlying our analysis should be rejected. To what extent this is true for other data sets remains to be analyzed.

Further research on loss functions of experts could focus on multi-step-ahead forecasts. As forecasts errors might be correlated in such situations, the methodology might be a bit more complicated than the one presented here. Finally, forecast updates, that is, sequential forecasts for the same event, are also interesting to analyze.

A Tables

Table 1: The estimated asymmetry parameter α_A of AAL and estimated systematic bias b following from regression (15), for each expert. Columns 2 and 3 show results for expert forecasts and Columns 4 and 5 for statistical model forecasts. The asterisks in the second and fourth column indicate if the $\hat{\alpha}_A$'s are significantly different from 1 and the asterisks in the third and fifth column indicate if the $\hat{b}$'s are significantly different from 0, where one is for the 10%, two are for the 5% and three are for the 1% significance level.

Country/	**EF**		**MF**	
expert	$\hat{\alpha}_A$	$\hat{b}$	$\hat{\alpha}_A$	$\hat{b}$
I	1.134*	10.513***	0.964	3.596
II	1.659***	-4.163**	1.081	-4.502**
III	1.617***	-5.918***	1.212***	-3.046
IV	1.310***	2.651	1.036	18.650***
V	1.295***	20.427***	1.019	-4.434
VI	1.215***	3.006	1.072	-8.601
VII	1.784***	4.274**	0.990	1.874
VIII	1.772**	112.332***	0.857	79.284***
IX	1.339***	-1.835	1.222	-13.143**
X	1.089	-0.263	1.037	-1.934
XI	1.489***	-1.431	1.026	0.710
XII	1.432***	-8.009***	1.018	-5.830***
XIII	1.856***	0.641	1.284***	-1.700
XIV	1.144*	0.522	1.250***	2.049
XV	1.146	160.758***	1.092	-53.567**
XVI	1.674***	-0.699	1.058	-12.795**
XVII	1.343***	-2.810	1.207***	2.301
XVIII	2.511***	-8.023	1.219	-2.008
XIX	1.095	0.423	1.094	67.514
XX	1.396***	24.837***	0.845	-4.500
XXI	1.170	29.717***	0.839	24.509**
XXII	0.964	-3.958	0.904	-3.199
XXIII	1.540***	10.496***	0.841**	3.136*
XXIV	1.161	26.693***	0.964	18.843**
XXV	1.018	0.143	0.968	0.020
XXVI	1.337***	-1.446	0.892	-1.068
XXVII	1.088	-20.161	0.782	-1.016
XXVIII	0.846	-6.465	0.366***	-190.657
XXIX	1.810***	-4.480	1.521***	-3.126
XXX	1.925***	2.411	0.854	5.099**
XXXI	1.267***	-10.693***	1.328***	-9.900***
XXXII	1.369***	-2.620	1.031	5.679
XXXIII	1.425*	120.716	0.968	-140.564
XXXIV	1.202*	-20.129	0.948	-27.534
XXXV	1.454***	-5.086	1.288***	-3.092

Table 2: The estimated asymmetry parameter α_L of LIN and estimated systematic bias b following from regression (16), for each expert. Columns 2 and 3 show results for expert forecasts and Columns 4 and 5 for statistical model forecasts. The asterisks in the second and fourth column indicate if the $\hat{\alpha}_L$'s are significantly different from 0 and the asterisks in the third and fifth column indicate if the $\hat{b}$'s are significantly different from 0, where one is for the 10%, two are for the 5% and three are for the 1% significance level.

Country/	**EF**		**MF**	
expert	$\hat{\alpha}_L$	$\hat{b}$	$\hat{\alpha}_L$	$\hat{b}$
I	4.4e-05	15.009***	7.1e-05*	2.585
II	-3.7e-04***	2.528	-1.2e-04***	-3.548*
III	-2.1e-04***	3.775**	-8.2e-05***	0.852
IV	1.1e-05	11.290***	-1.7e-05	19.688***
V	-9.2e-06	25.171***	5.4e-06	-4.066
VI	-6.4e-05***	13.438***	-2.1e-05	-4.848
VII	-5.8e-04***	13.489***	-2.2e-04	1.522
VIII	-4.2e-04**	136.130***	-2.3e-05	72.009***
IX	-2.3e-04	10.143**	-2.0e-04	-5.096
X	8.7e-05	1.611	2.1e-04	-0.888
XI	-6.2e-04***	2.742**	-7.8e-05	0.954
XII	-6.4e-04***	-4.652***	-1.3e-04	-5.689***
XIII	-6.3e-04***	8.567***	-4.3e-04**	1.414
XIV	-6.2e-06	1.560	-3.1e-05*	3.755***
XV	-8.5e-04*	160.967***	-1.8e-04	-48.658***
XVI	-1.5e-03***	9.614**	-1.1e-04	-11.557**
XVII	-8.4e-05***	2.979	-3.9e-05**	6.029**
XVIII	-3.0e-04***	15.927***	1.4e-04	3.585
XIX	-4.4e-05	18.512	-5.5e-06	105.082
XX	-7.6e-05	33.160***	3.0e-04**	-8.123**
XXI	-4.0e-05	36.277***	-1.4e-06	16.901**
XXII	-7.2e-05	-4.624	4.6e-04	-4.377
XXIII	-8.9e-05	15.560***	8.4e-05	1.126
XXIV	-1.9e-05	35.699***	1.5e-05	16.767**
XXV	2.9e-05	1.036	2.6e-05	-1.126
XXVI	1.1e-05	-0.221	4.2e-05	-1.543
XXVII	1.8e-05	-10.861	3.3e-04	-21.134
XXVIII	1.9e-05*	-77.468	5.3e-05***	-652.308***
XXIX	-9.7e-06**	6.806	2.1e-06	4.913
XXX	1.2e-05	7.557***	4.7e-06	3.859**
XXXI	-4.1e-05*	-6.420**	-6.3e-05**	-4.799*
XXXII	-1.3e-04**	20.429**	2.7e-05	8.526
XXXIII	-2.5e-06	221.010***	-7.2e-07	-150.069**
XXXIV	-4.5e-05***	-5.803	-2.7e-05	-33.433*
XXXV	-3.0e-04***	3.617	-1.4e-04*	2.981

Table 3: This table shows the number of times out of 35 that the hypothesis that the error terms of the regressions (15) (Column 2), (16) (Column 3) and (19) (Column 4) are standard normally distributed is rejected. We use the Kolmogorov-Smirnov test with a significance level of 1%.

	AAL	**LIN**	**AAL log**
EF	13	23	35
MF	8	12	35

Table 4: The estimated α_A of AAL and estimated $\alpha_E/(1-\alpha_E)$ of general loss function (20) with $q = 1$ using the estimation method of Elliott et al. (2005), for each country. Columns 2 and 3 show results for expert forecasts and Columns 4 and 5 for statistical model forecasts. The asterisks in the second and fourth column indicate if the $\hat{\alpha}_A$'s are significantly different from 1 and the asterisks in the third and fifth column indicate if the $\hat{\alpha}_E$'s are significantly different from 0.5, where one is for the 10%, two are for the 5% and three are for the 1% significance level.

Country/ expert	**EF**		**MF**	
	$\hat{\alpha}_A$	$\hat{\alpha}_E/(1-\hat{\alpha}_E)$	$\hat{\alpha}_A$	$\hat{\alpha}_E/(1-\hat{\alpha}_E)$
I	1.134*	1.321***	0.964	1.091
II	1.659***	1.387***	1.081	1.001
III	1.617***	1.579***	1.212***	1.170***
IV	1.310***	1.170***	1.036	1.148**
V	1.295***	1.468***	1.019	0.851**
VI	1.215***	1.291***	1.072	1.085
VII	1.784***	1.618***	0.990	1.036
VIII	1.772**	2.253***	0.857	1.365
IX	1.339***	1.505***	1.222	1.057
X	1.089	1.045	1.037	0.888
XI	1.489***	1.506***	1.026	1.152**
XII	1.432***	1.180**	1.018	0.918
XIII	1.856***	1.778***	1.284***	1.078
XIV	1.144*	1.171**	1.250***	1.211***
XV	1.146	2.118***	1.092	0.780*
XVI	1.674***	1.464***	1.058	0.924
XVII	1.343***	1.455***	1.207***	1.360***
XVIII	2.511***	1.655***	1.219	0.929*
XIX	1.095	1.090	1.094	1.423*
XX	1.396***	1.296**	0.845	0.821**
XXI	1.170	1.802***	0.839	1.268
XXII	0.964	0.970	0.904	0.990
XXIII	1.540***	1.846***	0.841**	1.043
XXIV	1.161	1.442***	0.964	1.127
XXV	1.018	1.019	0.968	1.019
XXVI	1.337***	1.415***	0.892	1.105
XXVII	1.088	1.027	0.782	0.798**
XXVIII	0.846	1.136	0.366***	0.598***
XXIX	1.810***	1.651***	1.521***	1.007
XXX	1.925***	1.532***	0.854	1.087
XXXI	1.267***	1.586***	1.328***	1.421***
XXXII	1.369***	1.476***	1.031	1.124
XXXIII	1.425*	2.228***	0.968	0.916
XXXIV	1.202*	1.084	0.948	0.864
XXXV	1.454***	1.397***	1.288***	1.290***

B Graphs

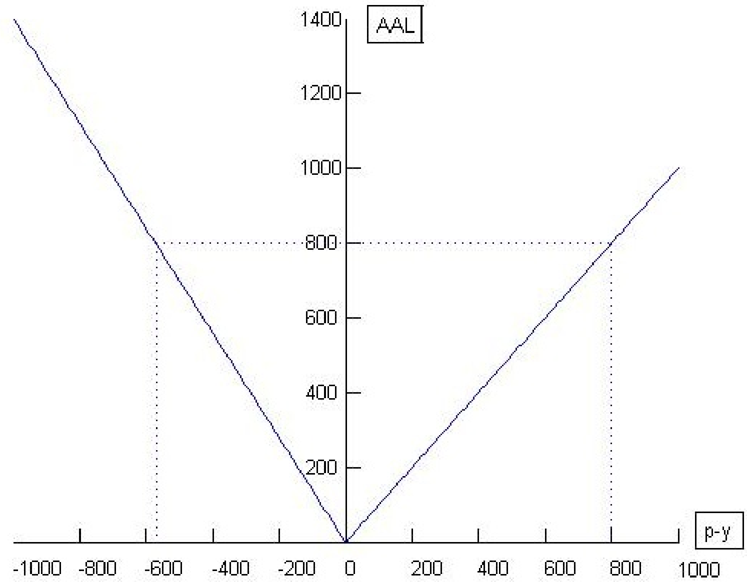

Figure 1: This figure shows the value of an AAL with $\alpha_A = 1.4$ for various values of the forecast error $p - y$.

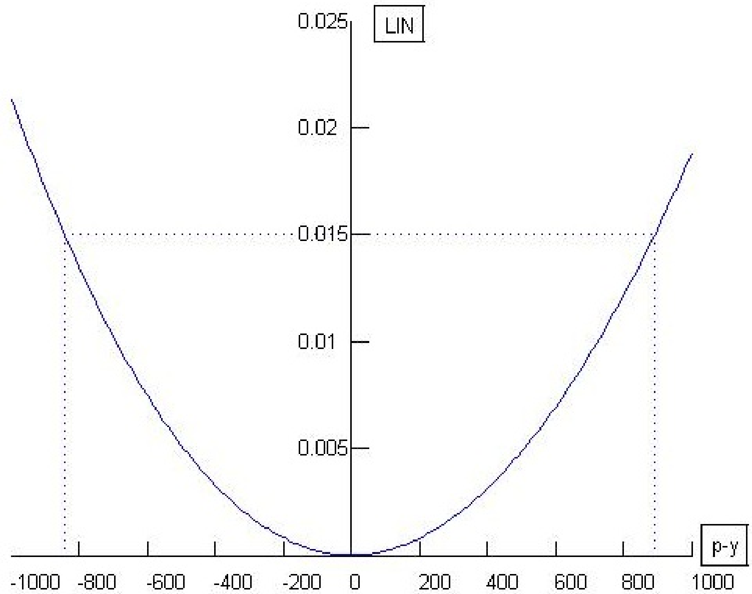

Figure 2: This figure shows the value of a LIN with $\alpha_L = -0.0002$ for various values of the forecast error $p - y$.

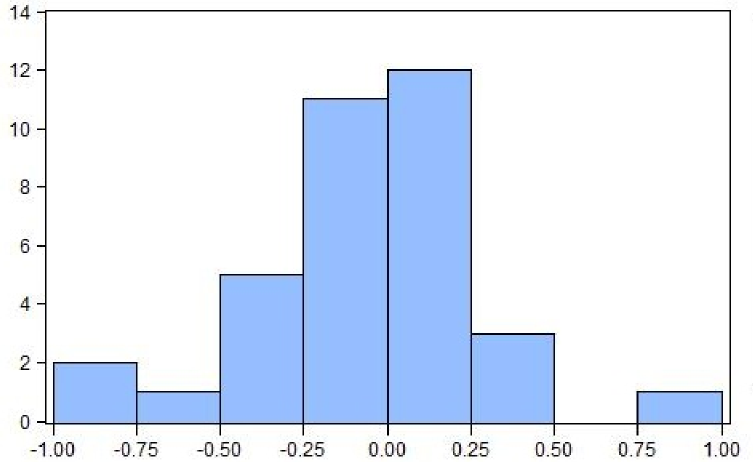

Figure 3: Histogram of differences between $\hat{\alpha}_A$ and $\hat{\alpha_E}/(1-\hat{\alpha_E})$ for 35 experts estimated over EF.

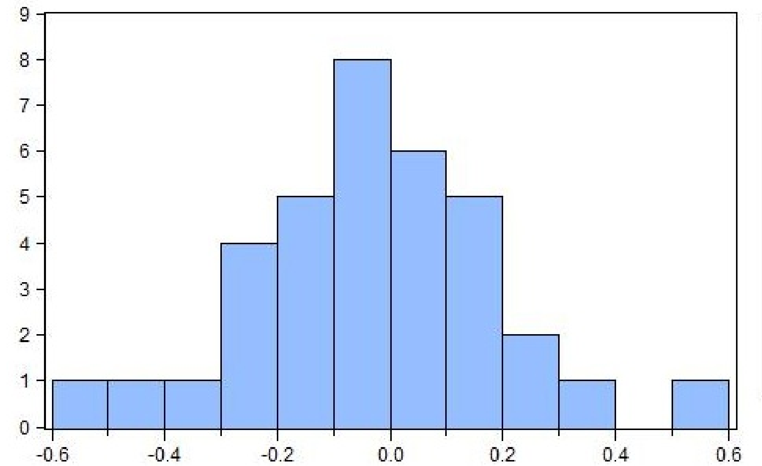

Figure 4: Histogram of differences between $\hat{\alpha}_A$ and $\hat{\alpha_E}/(1-\hat{\alpha_E})$ for 35 countries estimated over MF.

References

Blattberg, R. and Hoch, S. (1990). Database models and managerial intuition: 50% model + 50% manager. *Management Science*, 36(8):887–899.

Christoffersen, P. and Diebold, F. (1996). Further results on forecasting and model selection under asymmetric loss. *Journal of Applied Econometrics*, 11:561–571.

Christoffersen, P. and Diebold, F. (1997). Optimal prediction under asymmetric loss. *Econometric Theory*, 13:808–817.

Clatworthy, M., Peel, D., and Pope, P. (forthcoming 2011). Are analysts' loss functions asymmetric? *Journal of Forecasting.*

D'Agostino, R. and Stephens, M. (1986). *Goodness-of-fit techniques.* Marcel Dekker, Inc., New York.

Elliott, G., Komunjer, I., and Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *The Review of Economic Studies*, 72:1107–1125.

Ferguson, T. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, New York.

Fildes, R., Goodwin, P., Lawrence, M., and Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25:3–23.

Franses, P. and Legerstee, R. (2010). Do experts' adjustments on model-based SKU-level forecasts improve forecast quality? *Journal of Forecasting*, 29(3):331–340.

Mathews, B. and Diamantopoulos, A. (1986). Managerial intervention in forecasting: An empirical investigation of forecast manipulation. *International Journal of Research in Marketing*, 3:3–10.

Varian, H. (1975). A Bayesian approach to real estate assessment. In Fienberg, S. and Zellner, A., editors, *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, pages 195–208. North-Holland, Amsterdam.

Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81:446–451.