Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19103-6297

pier@ssc.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 97-018

"Evaluating Density Forecasts"

by

Francis X. Diebold, Todd A. Gunther and Anthony S. Tay

Evaluating Density Forecasts

Francis X. Diebold*, Todd A. Gunther*, and Anthony S. Tay*#

* Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104

Department of Economics and Statistics
National University of Singapore
10 Kent Ridge Crescent
Singapore 119260.

First Draft: June 1996
This Print: November 11, 1996

Correspondence to:

F.X. Diebold
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104
fdiebold@mail.sas.upenn.edu

Abstract: We propose several methods for evaluating and improving density forecasts. We focus primarily on methods that are applicable regardless of the particular user's loss function, but we also show how to use information about the loss function when and if it is known. Throughout, we take explicit account of the relationships between density forecasts, action choices, and the corresponding expected loss.

Acknowledgments: Helpful discussion was provided by Clive Granger, Atsushi Inoue and Hashem Pesaran. All remaining inadequacies are ours alone. We thank the National Science Foundation, the Sloan Foundation, the University of Pennsylvania Research Foundation, and the National University of Singapore for support.

1. Introduction

Prediction occupies a distinguished position in econometrics, as in all the sciences; hence, evaluating predictive ability is a fundamental concern. Reviews of the forecast evaluation literature, such as Diebold and Lopez (1996), reveal that much attention has been paid to evaluating point forecasts. In fact, the bulk of the literature focuses on point forecasts, while noticeably smaller sub-literatures treat interval forecasts (e.g., Christoffersen, 1995) and probability forecasts (e.g., Clemen, Murphy and Winkler, 1995).

Rather remarkably, almost no attention has been given to evaluating *density forecasts*. At least two factors explain this neglect. First, until recently density forecasts have been little used in economics and elsewhere, in part because analytic construction of density forecasts has historically required restrictive and sometimes dubious assumptions, such as normality of innovations and neglect of parameter estimation uncertainty. Recent work using numerical and simulation techniques to construct density forecasts, however, has reduced our reliance on such assumptions.[1] In fact, improvements in computer technology have rendered the provision of density forecasts increasingly straightforward, and we predict that density forecasts will soon be commonplace.

Second, the problem of density forecast evaluation appears difficult. Although it is certainly possible to adapt techniques developed for the evaluation of point, interval and probability forecasts to the evaluation of density forecasts, such approaches lead to evaluation

---

[1] See, for example, the discussion of construction of density forecasts using resampling techniques surveyed in Berkowitz and Kilian (1996).

based only on limited aspects of the density forecast. The problem, then, is to define and implement a comprehensive measure of density forecast adequacy.

In this paper we do so; we propose several variants of a general method for density forecast evaluation. Our methods evaluate the density forecast in its entirety and explicitly account for the relationships between the density forecast, the action choice, and the resultant expected loss. In section 2, we present a detailed statement and discussion of the problem, and we provide the theoretical underpinnings of the methods that we introduce subsequently. Our treatment is related to, but nevertheless distinct from, the notable and complementary independent work of Granger and Pesaran (1996). In section 3 we present our methods of density forecast evaluation, and we also present methods for improving suboptimal density forecasts. We treat separately the case in which the user's loss function is unknown and the case in which it is known; the first case is likely to be most relevant in some situations, the second in others. We conclude in section 4.

2. Density Forecasts, Loss Functions and Action Choices

The Basic Framework

Let $f_t(y_t | \Omega_t)$ be the data-generating process governing a series y_t, where Ω_t denotes all variables that affect the density of y_t.[2] For notational convenience, we will often simply write $f_t(y_t)$, but the dependence on Ω_t should be understood. Suppose that a series of 1-step-ahead

---

[2] We indulge in the standard abuse of notation, which favors convenience over precision, and fail to distinguish between random variables and their realizations. The meaning will be clear from context.

density forecasts of y_t is available, made over m periods.[3] The forecasts might be available in closed form, or they might be simulated draws from the densities. In any case, we represent the series of forecasts by $\{p_t(y_t)\}_{t=1}^{m}$.

Let the series of realizations be denoted by $\{y_t\}_{t=1}^{m}$. We wish to evaluate the density forecasts by considering the historical performance of the forecaster. There is an intimate relationship between density forecasts, action choices, and loss functions that is relevant when evaluating forecasts. Each forecast user has a loss function $L(a_t, y_t)$, where a_t refers to an action choice. The action choice need not be a prediction of y_t. For example, a_t may refer to the amount of insurance coverage to purchase, with y_t representing the realized loss. The user chooses an action to minimize expected loss computed using the density which she believes to be the data-generating process. If she believes that $p_t(y_t)$, the prediction from the forecaster at time t, is the correct density, then she chooses an action a_t^* by solving[4]

$$a_t^*(p_t(\cdot)) = \underset{a \in A}{\operatorname{argmin}} \int L(a, y_t) p_t(y_t) dy_t .$$

The action choice defines the loss $\ell(y_t \mid a_t^*) = L(a_t^*, y_t)$ faced for every realization of the process $f_t(y_t)$. This loss is a random variable and possesses a distribution function, which we will call the loss distribution, and which depends only on the action choice.

The effect of the density forecast on the user's expected loss is easily seen. A density forecast translates into a loss distribution. Two different forecasts will in general lead to different

---

[3] The methods we describe can be modified in obvious fashion for h-step-ahead forecasts.

[4] We assume a unique minimizer. A sufficient condition is that A be compact and that L be convex in a.

action choices and hence different loss distributions. Different forecasts that lead to the same action choice are, to the user, equivalent. A "good" forecast will lead to an action choice which results in a comparatively low expected loss,

$$E[\ell(y_t \mid a_t^*)] = \int \ell(y_t \mid a_t^*)\, f_t(y_t) dy_t,$$

given the actual data-generating process $f_t(y_t)$.

<u>A Conceptual Framework for Evaluating Density Forecasts</u>

Suppose the user had the option of choosing between two forecasts in a given period, denoted by $p_j(y)$ and $p_k(y)$, where the subscript refers to the forecast and the time subscripts are omitted for convenience. The forecast user will prefer the forecast $p_j(y)$ if the mean of the loss distribution arising from following this forecast is less than the mean of the loss distribution obtained if $p_k(y)$ were followed instead, that is, if

$$\int \ell(y \mid a_j^*)\, f(y) dy \leq \int \ell(y \mid a_k^*)\, f(y) dy,$$

where a_j^* denotes the action that minimizes expected loss, given that the user bases the action choice on forecast j.

Ideally, we would like to find a way of assigning to each forecast a score $D(p_j)$, constructed from the history of density forecasts and realizations, which would measure the divergence of the realization from the forecast density, such that all users, *regardless of their loss function*, would prefer the forecast with the lower divergence. This would allow us to rank the forecasts. Unfortunately, the following proposition shows that no such measure exists for ranking two incorrect density forecasts.

Proposition 1. Let $f(y)$ be the density of y, p_j be a forecast of y, and $a \in A$ be a set of action choices. Let a_j^* be the optimal decision based on forecast p_j. Then no measure $D(\cdot)$ exists such that for arbitrary forecast densities p_j and p_k, both distinct from f, and for all possible loss functions $L(a, y)$,

$$D(p_j) \geq D(p_k) \Leftrightarrow \int L(a_j^*, y) f(y) dy \geq \int L(a_k^*, y) f(y) dy.$$

Proof. In order to establish the result, it is sufficient to find a pair of loss functions L_1 and L_2, a density function f governing y, and a pair of forecasts, p_j and p_k, such that,

$$\int L_1(a_k^*, y) f(y) dy \leq \int L_1(a_j^*, y) f(y) dy,$$

while

$$\int L_2(a_k^*, y) f(y) dy \geq \int L_2(a_j^*, y) f(y) dy.$$

That is, user 1 does better on average under forecast k, while user 2 does better under forecast j. It's straightforward to construct such an example. Suppose the true density function is N(0,1), and suppose that user 1's loss function is $L_1(a, y) = (y - a)^2$ and user 2's loss function is $L_2(a, y) = (y^2 - a)^2$. The optimal action choices are then $\int y\, p(y)\, dy$ and $\int y^2\, p(y)\, dy$ respectively. That is, user 1 bases her action choice on the mean, with higher expected loss occurring with larger errors in the forecast mean, while user 2's actions and expected losses depend on the error in the forecast of the uncentered second moment. In this context, consider two forecasts: forecast j is N(0,2) and forecast k is N(1,1). User 1 ranks forecast j above forecast k, because it leads to an action choice that in turns leads to a loss distribution with lower expected loss, but user 2 would rank forecast k above forecast j. □

To repeat: there is no way to rank two incorrect density forecasts such that all users will prefer the higher ranked forecast. However, if a forecast coincides with the true data-generating process, then it will minimize expected loss for all forecast users, regardless of their loss function, as we show in the following proposition.[5]

Proposition 2. Suppose that the forecast, $p_j(y)$, is identical to the data-generating process, i.e., $p_j(y) = f(y)$, and hence a_j^* minimizes the actual expected loss. Then for all possible density forecasts $p_k(y)$,

$$\int L(a_j^*, y) f(y) dy \leq \int L(a_k^*, y) f(y) dy,$$

i.e., choosing the action according to the true density gives the least expected loss.

Proof. The result follows immediately from the assumption that a_f^* minimizes expected loss over all possible actions, including those which might be chosen under alternative densities of y, such as $p_k(y)$. □

While no forecaster can be reasonably expected to reproduce the true data-generating process exactly, the preceding propositions suggest a direction for evaluating density forecasts. Without even taking loss functions into consideration, we know that the correct density is weakly superior to all forecasts. If there is statistically significant evidence that the realizations $\{y_t\}_{t=1}^m$ do not come from the forecast densities $\{p(y_t)\}_{t=1}^m$, we know (subject to Type I error) that some

---

[5] Granger and Pesaran (1996) independently arrive at the same conclusion.

users, depending on their loss functions, could potentially be better served by a different density forecast. In the following section, we propose tests to that effect.

3. Evaluating Density Forecasts

The Link Between Density Forecasts and the True Distribution: The Probability Integral Transform

The methods that we propose are based on the relationship between the data-generating process, f, and the density forecast, p, as related through the probability integral transform, z, of the realization of the process taken with respect to the density forecasts. The following lemma describes its distribution, q.

Lemma 1. Let y be the variable of interest with $f(y)$ as its density, and let $p(y)$ represent a density forecast of y. Let the variable $z(y)$ be the probability integral transform of y with respect to $p(y)$. That is,

$$\begin{aligned} z(y) &= \int_{-\infty}^{y} p(u)\,du \\ &= P(y). \end{aligned}$$

Then $z(y)$ will have support on the unit interval with density function

$$\begin{aligned} q(z) &= \left|\frac{\partial P^{-1}(z)}{\partial z}\right| f(P^{-1}(z)) \\ &= \frac{f(P^{-1}(z))}{p(P^{-1}(z))}. \end{aligned}$$

Proof: Use the fact that $p(y) = \frac{\partial P(y)}{\partial y}$ and $y = P^{-1}(z)$. □

Note that if $p(\cdot) = f(\cdot)$, $q(z)$ is simply the Uniform (0,1) density. Armed with Lemma 1, it is a simple matter to characterize completely the z series when the density forecasts are correct; we do so with the following proposition.

Proposition 3. Suppose a series $\{y_t\}_{t=1}^m$ is generated from $\{f_t(y_t|\Omega_t)\}_{t=1}^m$. If a sequence of density forecasts $\{p_t(y_t)\}_{t=1}^m$ coincides with $\{f_t(y_t|\Omega_t)\}_{t=1}^m$, then

$$\{z_t\}_{t=1}^m = \left\{\int_{-\infty}^{y_t} p_t(u)\,du\right\}_{t=1}^m \overset{iid}{\sim} U(0,1).$$

That is, the sequence of probability integral transforms of $\{y_t\}_{t=1}^m$ with respect to $\{p_t(y_t)\}_{t=1}^m$ is iid U(0,1).

Proof: The joint density of $\{y_t\}_{t=1}^m$ can be decomposed as

$$f(y_m,\ldots,y_1|\Omega) = f_m(y_m|\Omega_m)\, f_{m-1}(y_{m-1}|\Omega_{m-1}) \ldots f_1(y_1|\Omega_1).$$

The joint density of $\{z_t\}_{t=1}^m$ can therefore be written as

$$q(z_m,\ldots,z_1|\Omega) = \frac{f_m(P_m^{-1}(z_m)|\Omega_m)}{p_m(P_m^{-1}(z_m))} \cdot \frac{f_{m-1}(P_{m-1}^{-1}(z_{m-1})|\Omega_{m-1})}{p_{m-1}(P_{m-1}^{-1}(z_{m-1}))} \cdots \frac{f_1(P_1^{-1}(z_1)|\Omega_1)}{p_1(P_1^{-1}(z_1))}.$$

From Lemma 1, under the assumed conditions, each of the ratios above is a U(0,1) density. Hence the joint density of $\{z_t\}_{t=1}^m$ is an m-variate U(0,1) distribution. Because the marginals of a multivariate U(0,1) distribution are also U(0,1), the joint distribution is the product of the marginals, which is the definition of independence. Hence $\{z_t\}_{t=1}^m$ is distributed iid U(0,1). □

<u>Evaluating Density Forecasts When the Loss Function is not Specified</u>

The theory developed thus far suggests that we use the series $\{z_t\}_{t=1}^m$, derived from the history of realizations and density forecasts, to evaluate forecasts. We simply check whether

$\{z_t\}_{t=1}^{m}$ is iid U(0,1). The idea is based upon the same principle as the Kolmogorov-Smirnov and Cramer-von Mises tests, which check whether a random sample $\{y_t\}_{t=1}^{m}$ is drawn from density $p(y)$ by taking the probability integral transform of the sample with respect to $p(y)$. In this paper, we effectively consider a sample $\{y_t\}_{t=1}^{m}$ that is a realization from a *sequence* of densities, but Proposition 3 reveals that if a forecaster manages to capture the sequence of densities that forms the true data-generating process, then the probability integral transforms are still iid U(0,1). In our context, however, we do not begin with an iid *assumption*; rather, it's something *check*.

Simple tests of iid U(0,1) behavior are readily available. Recall, for example, that if $z_t \sim U(0,1)$ then $-2 \log z_t \sim \chi^2_2$.[6] Hence, if z is iid U(0,1), then $S = \sum_{t=1}^{m} -2 \log z_t \sim \chi^2_{2m}$. We could use the S-statistic to perform a simple significance test. Alternatively, we could perform any of the various well-known tests for uniformity, such as a runs test or a Kolmogorov-Smirnov test, all of which are actually joint tests of uniformity and iid.

Such tests, however, are not likely to be of much value in practical applications, because they are not constructive; that is, when rejection occurs, the tests generally provide no guidance as to *why*. If, for example, the S statistic rejects the hypothesis of iid U(0,1) behavior, is it because of violation of unconditional uniformity, violation of iid, or both? Moreover, even if we know that rejection comes from violation of uniformity, we'd like to know more: What, precisely, is the nature of the violation of uniformity, and how important is it? Similarly, even if we know that rejection comes from a violation of iid behavior, what precisely is its nature? Is z heterogeneous but independent, or is z dependent? If z is dependent, is the dependence operative primarily through the conditional mean, or are higher-ordered conditional moments, such as the

[6] See Johnson and Kotz (1970).

variance, relevant? Is the dependance strong and important, or is iid an adequate approximation, even if strictly false?

The nonconstructive nature of tests of iid U(0,1) behavior, and the nonconstructive nature of related separate tests of iid and U(0,1), which can easily be constructed, make us eager to trade (superficial) statistical rigor for more revealing methods of exploratory data analysis. First, as regards evaluating unconditional uniformity, we suggest visual assessment using the obvious graphical tool, a density estimate. Simple histograms are attractive in the present context, because they allow straightforward imposition of the constraint that z has support on the unit interval, in contrast to more sophisticated procedures such as kernel density estimates with the standard kernel functions. The estimated density can be visually compared to a U(0,1), and if desired the graphical analysis can be supplemented with a formal test of uniformity, robust to dependence and heterogeneity. Alternatively, we can estimate the c.d.f. of z and compare it to that of a U(0,1), by examining the corresponding q-q plot. The usual estimates of the density of z, the c.d.f. of z, and the q-q plot remain consistent regardless of the possible presence of dependence and heterogeneity in z.

Second, as regards evaluating whether z is iid, we again suggest visual assessment using the obvious graphical tool, the correlogram. Because we're interested potentially sophisticated forms of dependence -- not just linear dependence -- we examine not only the correlogram of z, but also those of powers of z. In practice, examination of the correlograms of z, z^2, z^3 and z^4 should be adequate; it will reveal any dependence operative through the conditional mean, conditional variance, conditional skewness, or conditional kurtosis. If desired, in conjunction with the correlogram inspection, standard tests for white noise may be applied to z and its powers.

The presence of particular forms of dependence in z can be informative in guiding forecasters and users about how to improve density forecasts. For instance, serial correlation in the z series may indicate that the mean dynamics have been inadequately modeled by the forecaster. A caveat, however, is that there is in general no one-to-one correspondence between the type of dependence found in z and the dependence in y missed by the forecasts. For example, assume that the true data-generating process is GARCH(p,q). Even if a forecaster correctly specifies the conditional variance function and perfectly estimates its parameters, there will be dependence in z if the forecaster assumes the wrong conditional density.

Adjusting Density Forecasts

In the previous section, we approached forecast evaluation from an historical perspective, evaluating the ability of a forecaster based on past realizations. The intent, of course, is to gauge the likely future accuracy of the forecaster based on past performance, assuming that the relationship between the correct density and the forecaster's predictive density remains fixed. Given that we observe systematic errors in the historical forecasts, the user may wish to simply reject the forecast. It may also turn out that these errors are irrelevant to the user, a case we further examine when we explicitly account for the user's loss function. Nevertheless, it is possible to take these errors into consideration when using the current forecast, just as it is possible to do so in the point forecast case. In the point forecast case, for instance, one can regress the y's on the $\hat{y}$'s, the predicted values, and use that relationship to construct an adjusted point forecast, allowing as well for various sorts of dynamics in the regression.

In the context of density forecasts, a similar procedure can be constructed by rewriting the relationship in Lemma 1. Suppose that the user is in period m and possesses a density forecast of y_{m+1}. From Lemma 1, we have

$$\begin{aligned} f_{m+1}(y_{m+1}) &= p_{m+1}(y_{m+1})\, q_{m+1}(P(y_{m+1})) \\ &= p_{m+1}(y_{m+1})\, q_{m+1}(z_{m+1}) \quad . \end{aligned}$$

Thus if we know $q_{m+1}(z_{m+1})$, we would know the actual distribution $f_{m+1}(y_{m+1})$. Since $q_{m+1}(z_{m+1})$ is unknown, an estimate $\hat{q}_{m+1}(z_{m+1})$ can be formed using the historical series of $\{z_t\}_{t=1}^{m}$, and an estimate of the true distribution $\hat{f}_{m+1}(y_{m+1})$ can then be constructed. If the sample $\{z_t\}_{t=1}^{m}$ turns out to be iid, then standard density estimation techniques can be used to produce the estimate $\hat{q}_{m+1}(z_{m+1})$. Otherwise, the estimation of $q_{m+1}(z_{m+1})$ becomes a non-trivial matter, which we defer to future research.

Evaluating Density Forecasts Using a Specific Loss Function

If a series of density forecasts has been systematically in error, it may still be the case that for a particular user, depending on her loss function, the systematic errors may be irrelevant. To be precise, the forecast may be such that the action choice induced by the forecast, a_p^*, minimizes the user's actual expected loss.[7] In such cases, which we now consider, the user's loss function can be incorporated into the evaluation process.

Consider a density forecast series, $\{p_t(y_t)\}_{t=1}^{m}$, and the corresponding action series, $\{a_{p,t}^*\}_{t=1}^{m}$, of a particular user. The series of action choices results in a series of potential losses $\{\ell(y_t \mid a_{p,t}^*)\}_{t=1}^{m}$, where $\ell(y_t \mid a_{p,t}^*) = L(a_{p,t}^*, y_t)$ and $y_t \sim f_t(y_t)$. We would like to compare

---

[7] Since we have assumed a unique minimizing action choice, this implies that $a_p^* = a_f^*$.

each period's realized loss with that period's expected loss under that optimal action choice $E_{f,t}[\ell(\cdot|a_{f,t})]$. The expected difference will be positive unless $a^*_{p,t} = a^*_{f,t}$. If we further assume that the forecaster correctly assesses the expected loss in each period, i.e., $E_{f,t}[\ell(\cdot|a_{f,t})] = E_{p,t}[\ell(\cdot|a_{p,t})]$, we can compute the differential

$$d_t = \ell(y_t|a^*_{p,t}) - E_p[\ell(\cdot|a^*_{p,t})].$$

Under the joint null hypothesis that the series of density forecasts is optimal relative to the user's loss function and that the forecaster correctly specifies the expected loss in each period, $E[d_t]=0$, which can be tested in the same way that Diebold and Mariano (1995) test whether two point forecasts are equally accurate under the relevant loss function.

4. Concluding Remarks

We have provided a characterization of optimal density forecasts, and we have proposed methods for evaluating whether reported density forecasts are optimal. Our methods are based on the series of probability integral transforms. We argued that this series contains much, if not all, of the information relevant to forecast users, regardless of their loss function. We showed how to use the series of probability integral transforms to judge whether a potentially large set of users, though not all, would reject a forecast, without explicitly specifying each individual user's loss function. We also showed, in the same framework not requiring specification of the loss function, how to improve a suboptimal density forecast by using information on previously-issued density forecasts and subsequent realizations. Finally, when information on the relevant loss function is available, we also showed how to evaluate a density forecast with respect to that loss function.

Our methods of density forecast evaluation have interesting parallels with well-known methods of evaluating point and interval forecasts. First, it's well-known that optimal point forecasts have corresponding 1-step-ahead errors that are iid with zero mean. Second, Christoffersen (1996) shows that the hit series corresponding to a series of correctly calibrated $(1-\alpha)\%$ interval forecasts is iid Bernoulli$(1-\alpha)$.[8] Our methods, in parallel, are based on the fact that the z series corresponding to a series of correct density forecasts is iid U(0,1).

Our method for improving incorrect density forecasts based on their historical track record also parallels a well-known method for improving suboptimal point forecasts, the Mincer-Zarnowitz (1969) regression. In a Mincer-Zarnowitz regression, we regress realizations y on point forecasts an intercept and $\hat{y}$,

$$y = \beta_0 + \beta_1 \hat{y} + \varepsilon.$$

Optimality of the point forecast with respect to some information set corresponds to $(\beta_0, \beta_1) = (0, 1)$ and at most h-dependence in ε if $\hat{y}$ is h-step-ahead. If a forecast fails the Mincer-Zarnowitz test, the regression nevertheless provides an immediate way to improve the forecast -- use $\hat{\beta}_0 + \hat{\beta}_1\hat{y} + \hat{\varepsilon}$ rather than $\hat{y}$.

We stress that, although we and others sometimes call density forecasts of the sort we consider in this paper "predictive densities," our methods are not Bayesian -- we evaluate density forecasts in terms of their behavior in repeated samples. Our evaluation methods are, however, similar in spirit to those used by predictivist Bayesians (e.g., Dawid, 1984; Seillier-Moiseiwitsch

[8] The hit sequence at any time is defined to be 1 if the realization is in the forecasted interval, and zero otherwise.

and Dawid, 1993). The methods by which the density forecasts themselves are produced are of no concern to us; they could be produced by any method, including Bayesian methods.[9]

The methods developed here may find wide application when density forecasts become more standard in economics and finance, as is happening already. The booming area of financial risk management, for example, is explicitly dedicated to providing density forecasts of future returns and to tracking certain aspects of the density, such as value at risk.[10] The day may not be far off in which risk management firms compute such density forecasts routinely, using simulation techniques to minimize reliance on ad hoc assumptions, and store draws from the density on line. Forecast purchasers will be granted access to the simulations and will be able to explore them for themselves, in a fashion similar to Geweke's (1994) suggestion for Bayesian reporting.

---

[9] A strict Bayesian, of course, would have little interest in our evaluation methods; conditional on a particular sample path and specification of the prior and likelihood, the predictive density simply is what it is, and there's nothing to evaluate. More eclectic Bayesians however, frequently evaluate Bayesian methods in classical terms. See, for example, the papers collected in Zellner (1984).

[10] See, for example, the popular RiskMetics system of J.P. Morgan & Co. (1996).

References

Berkowitz, J. and L. Kilian, 1996. Recent Developments in Bootstrapping Time Series, Manuscript, Department of Economics, University of Pennsylvania.

Christoffersen, P.F., 1995. Evaluating Interval Forecasts, Manuscript, Department of Economics, University of Pennsylvania.

Clemen, R.T., A.H. Murphy and R.L. Winkler, 1995. Screening Probability Forecasts: Contrasts Between Choosing and Combining, *International Journal of Forecasting*, 11, 133-146.

Dawid, A.P., 1984. Statistical Theory: The Prequential Approach, *Journal of the Royal Statistical Society, Series A*, 147, 278-292.

Diebold, F.X. and J.A. Lopez, 1996. Forecast Evaluation and Combination, *Handbook of Statistics*, forthcoming, North-Holland, Amsterdam.

Diebold, F.X. and R.S. Mariano, 1995. Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

Geweke, J., 1994. Bayesian Comparison of Econometric Models, Working Paper 532, Research Department, Federal Reserve Bank of Minneapolis.

Granger, C.W.J. and M.H. Pesaran, 1996. A Decision Theoretic Approach to Forecast Evaluation, Manuscript, Departments of Economics, University of California, San Diego and Cambridge University.

Johnson, N.L. and S. Kotz, 1970. *Continuous Univariate Distributions*, Volume Two, Houghton-Mifflin, Boston.

J.P. Morgan & Co, 1996. RiskMetrics Technical Document, Fourth Edition, J.P. Morgan & Co., New York.

Mincer, J. and V. Zarnowitz, 1969. The evaluation of economic forecasts, in J. Mincer (ed.), *Economic Forecasts and Expectations*, National Bureau of Economic Research, New York.

Seillier-Moiseiwitsch, F. and A.P. Dawid, 1993. On Testing the Validity of Sequential Probability Forecasts, *Journal of the American Statistical Association*, 88, 355-359.

Zellner, A., 1984. *Basic Issues in Econometrics*, University of Chicago Press, Chicago.