OPTIMAL PREDICTION WITH A GENERAL LOSS FUNCTION

B.D. MCCULLOUGH

Department of Decision Sciences
Drexel University, Philadelphia, PA 19104

Analytic estimators for asymmetric loss functions are known only in very special cases, and then only under a normality assumption. In practical situations, the true loss function must be crudely approximated. This paper proposes a bootstrap method for prediction under asymmetric and nonlinear loss functions that is freed from crude approximations to true loss functions and the normality assumption. The estimator proposed herein combines two computationally intensive methods: bootstrapping and nonlinear programming. Simulations demonstrate the viability of the proposed method. An application to interest rate data is presented.

INTRODUCTION

Prediction is an important forecasting tool, but as Whittle (1983, p. 106) has noted, “[P]rediction is rarely an end in itself. In most cases the predicted value, once obtained, is used to initiate or modify a course of action . . . In this larger context the problem of prediction appears only as incidental, and the central problem is that of regulation, i.e., of using past values to determine present action in such a way that the future course of the process is as near as possible to the desired one.” “Regulation” is, of course, “prediction under loss”. Forecasters have long been hampered in the pursuit of the larger goal of regulation due to the paucity of loss functions for which a suitable predictor is known. The predictor for a generic loss function cannot, in general, be found because the (conditional) distribution of the future value usually is unknown, so a predictor cannot be derived (Granger and Newbold, 1986 p. 121). Hence, forecasters are all but compelled to rely on quadratic loss with an assumption of normal errors. Two asymmetric loss functions for which closed-form predictors can be had are the asymmetric linear loss function (Granger, 1969) and the linear-exponential
('linex') loss function due to Varian (1975) and Zellner (1986); Cain (1994) gives prediction adjustments for the asymmetric quadratic loss function. All three methods require extremely specific forms for the loss function and a Gaussian assumption. Hence, they are of limited applicability in practical situations.

Another disadvantage of these methods is that all three require that the cost function be expressed in prediction-error loss form: if \( x \) is the true value and \( x_\hat{t} \) is the prediction, so that \( e_\hat{t} = x - x_\hat{t} \), then the loss function must be of the form \( L(e) \). From Granger (1969) and Christofferson and Diebold (1997) we know that, for the prediction-error loss form, the optimal predictor is simply the conditional mean plus a constant bias term. The bias term depends on the higher moments and the loss function. If the innovations are not i.i.d. then the bias term is not necessarily constant. Thus, for non-i.i.d. errors the prediction-error loss form is inappropriate.

While asymmetric loss functions rarely are used for prediction (but see Cain and Jannsen, 1995, and Whiteman, 1996), they frequently are used to assess competing forecast models (Weiss and Andersen, 1984; Diebold and Mariano, 1995). Also, they have been used to conduct "rationality tests" (Batchelor and Peel, 1998). In the context of making predictions according to the loss function by which forecasts are evaluated, recently much progress has been made in the case of asymmetric loss. Weiss (1996) approximates the optimal predictor based on the mean and variance of the conditional distribution. Christofferson and Diebold (1996, 1997) offer two fully parametric methods that are essentially based on the Weiner-Kolmogorov approach. The first approximates the optimal predictor via series expansions, and the second determines the exact optimal predictor for a piecewise linear approximation to the true loss function.

This paper proposes a bootstrap solution to the problem of prediction with a general loss function. Léger and Romano (1990) present a general framework for establishing consistency and convergence results for estimators that minimize an empirical bootstrap estimate of risk over a class of estimators. Prediction under loss can be viewed as a specific problem within such a general framework. This bootstrap method is not dependent upon a specific assumption concerning the innovations, as are the methods of Christofferson and Diebold, and it uses more information about the conditional distribution than does the method of Weiss, which concentrates on the first two moments of the conditional distribution. Additionally, it is not restricted to the prediction-error loss form, and thus embraces a much wider class of loss functions than the analytic methods.
Recent advances in bootstrap methodology now enable forecasters to estimate the conditional distribution of a future observation, enabling this information to be combined with a loss function without the artifice of assuming (conditionally) Gaussian forecast errors. Bootstrap methods for estimating the distribution of a future value, conditional on the observed sample, have been developed for several cases. Peters and Freedman (1985), Stine (1985), and Breiman (1992) treat the k-variable regression model. Thombs and Scuchany (1990) and Kabaila (1993) bootstrap the future distribution of an autoregressive process. Beran (1995) considers bootstrap prediction for random coefficient models. Applications of some of these methods are found in Peters and Freedman (1985), Veall (1987), Bernard and Veall (1987), and McCullough (1994, 1996).

The ability to bootstrap the distribution of the future value lends itself well to prediction under loss functions. The bootstrap distribution of the future value is multiplied by the loss function to form the expected loss (risk) function, which product then is minimized using standard nonlinear programming techniques. Since both the bootstrap and nonlinear programming are computationally intensive, this method is yet another example of the substitution of computational power for analytic complexity. Section two briefly reviews prediction under loss. To avoid needless complications arising from consideration of more sophisticated models, section three describes the simplest possible bootstrap of a future value. Section four combines bootstrap prediction with asymmetric loss functions. Section five presents four simulation experiments that compare predictions and associated costs of the bootstrap predictor (BP) with the known predictors for asymmetric linear (AL), linex (LX), and asymmetric quadratic (AQ) loss functions; the BP approximates all three. In the fourth experiment, the BP closely approximates the numerical solution of an analytically intractable nonlinear loss function that cannot be treated by the other methods: square loss for negative prediction error and square-root loss for positive prediction error. Section six considers an application to predicting interest rates when the loss function is not in prediction-error form. Section seven offers conclusions.

**PREDICTION UNDER A SPECIFIED LOSS FUNCTION**

For simplicity consider the scalar unconditional case, the extension to regression analysis and conditional prediction being obvious. Suppose that the random variable will take on the value \( x \) in the forecast period, and let its density be \( g(x) \). The forecast of \( x \) is \( x_f \) with associated forecast error \( e = x - x_f \). Let the loss

associated with a prediction error be given by \( L(x - x_j) \) for prediction-error loss form, and \( L(x, x_j) \) more generally. The predictor that minimizes expected loss is that value of \( x_j \) that minimizes risk

\[
\hat{x_j} = \min_{x_j} R, \quad R = \int_{-\infty}^{\infty} L(x, x_j) g(x) \, dx
\]  

(1)

Analytic solutions to (1) are likely to be unavailable. Numerical solutions under appropriate assumptions can be found, but whether these assumptions are valid in practice often is unknown. If \( L(\cdot) \) is symmetric about the origin and \( g(x) \) is symmetric about its mean, as in the case of quadratic loss with normal prediction error, then least squares is well-known to be optimal. When \( L(x - x_j) \) is asymmetric the situation is problematic, and a solution is known only for special cases, and then only under specific distributional assumptions.

Consider the AL loss function

\[
L(x - x_j) = \begin{cases} 
  a(x - x_j) & x - x_j > 0 \\
  b(x - x_j) & x - x_j < 0 \\
  0 & x - x_j = 0 
\end{cases}
\]

(2)

where \( a > 0, b < 0, a \neq b \). Granger (1969) has shown that the predictor that minimizes expected loss in the case that \( x \) is normally distributed is given by the AL loss predictor (ALP),

\[
x_j = \bar{x} + \sigma \Phi^{-1}(a/(a - b))
\]

(3)

where \( \bar{x} \) is the sample mean, \( \sigma \) is the population standard deviation (in practice replaced by the sample standard deviation), and \( \Phi(\cdot) \) is the cdf of the standard normal density.

The LX loss function is given by

\[
L(x - x_j) = b \{ \exp \left[ a(x - x_j) \right] - [a(x - x_j)] - 1 \}
\]

(4)

where \( a \neq 0 \) determines the shape of the loss function and \( b > 0 \) determines the scale. For small \( |a| \) the loss function is approximately symmetric. When \( a = 1 \) the loss function is decidedly asymmetric, with overestimation more costly than underestimation. For \( a < 0 \) and not small, the loss function rises exponentially for \( x_j - x < 0 \) and almost linearly for \( x_j - x > 0 \). Zellner (1986) shows that when \( x \) is normally distributed, the LX loss predictor (LXP) is given by

\[
x_j = \bar{x} - a\sigma^2/2
\]

(5)

which has expected cost equal to
\[ b \exp \left[ a(x_f - Ex) \right] \exp \left[ a^2 \sigma^2/2 \right] - a(x_f - Ex) - 1 \] \hspace{1cm} (6)

Consider the asymmetric quadratic loss function

\[
L(x - x_f) = \begin{cases} 
    a(x - x_f)^2 & x - x_f > 0 \\
    b(x - x_f)^2 & x - x_f < 0 \\
    0 & x - x_f = 0
\end{cases}
\] \hspace{1cm} (7)

where \( a > 0, b > 0, a \neq b \). Cain (1994) has shown that the predictor that minimizes expected loss in the case that \( x \) is normally distributed is given by the AQ loss predictor

\[ x_f = \bar{x} + \sigma \delta^* \] \hspace{1cm} (8)

where \( \delta^* \) is a solution of

\[ \phi(\delta) + \delta \left[ \frac{-a}{(a - b)} + \Phi(\delta) \right] = 0, \] \hspace{1cm} (9)

\( \phi(\cdot) \) is the pdf of the standard normal, and which may be solved by numerical methods. The corresponding expected cost is

\[ \sigma^2 [a - (a - b) \Phi(\delta^*)] \] \hspace{1cm} (10)

It will be shown that the BP solves the same problem as each of these three methods, in the sense that the answer provided by BP approximates the analytic solutions to the problems which ALP, LXP, and AQP solve.

**THE BOOTSTRAP**

To keep ideas clear and make the first implementations of this method as direct as possible, we consider the simplest possible setting for bootstrapping a future value, the univariate case as discussed by Bai and Olslen (1988). Let \( X_1, X_2, \ldots, X_n, X_{n+1} \) be i.i.d. random variables with common distribution function \( F(\cdot) \), mean \( \mu = 0 \) and variance \( \sigma^2 \). Assume \( E(X_i^4) < \infty \) and \( F''(\cdot) \) exists and is bounded. The \( X_i, i = 1, \ldots, n \), can be considered a learning sample and \( X_{n+1} \) a test case. We wish to estimate the distribution of \( X_{n+1} \). Let \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i/n \); so the residuals are \( e_i = x_i - \bar{x}, i = 1, \ldots, n \). Let \{e\} denote the collection of residuals after rescaling by \( \sqrt{n/(n-1)} \). By a uniform random draw with replacement from \{e\} form \( j = 1, \ldots, J \) bootstrap resamples \{e_j^*\} each of dimension \( n \). Create \( J \) bootstrap replicates \( \{x_j^*\} = \bar{x} + \{e_j^*\} \). For each \( j = 1, \ldots, J \) let \( \bar{x}_j^* \) be the mean of the \( j \)th bootstrap replicate \( \{x_j^*\} \). Stopping here, \( \bar{x}_j^*, j = 1, \ldots, J \) can be used to
form the empirical bootstrap distribution of the sample mean, i.e., to estimate the sampling distribution of $\bar{X}$. However, of interest is the bootstrap distribution of the future value $X_{n+1}$, which is formed as $x_{j}^{**} = x_{j}^{-} + e_{j}^{**}$, where each $e_{j}^{**}$ is again a uniform draw with replacement from $\{e\}$. The $x_{j}^{**}$ can be sorted to form the empirical distribution of $X_{n+1}$, which can then be multiplied by the loss function, with the resulting product treated by optimization methods. As mentioned earlier, there exist more sophisticated bootstraps for generating predictions for $k$-variable regression, random coefficient models, and autoregressive processes. Any of one of them can be used to generate $x_{j}^{**}$ for use with the BP.

**The Loss Function**

The bootstrap distribution of the future value can be multiplied point-by-point with the loss function to form the expected cost function. Because $g(x^{**}) \xrightarrow{d} g(x)$ (Bai and Olsheen, 1988), if the loss function is known, i.e., is non-stochastic, then it immediately follows that the product of the loss function and the bootstrap distribution converges weakly to the product of the loss function and the true distribution, i.e., $L(x, x_{j})g(x^{**}) \xrightarrow{d} L(x, x_{j})g(x)$. Assumption A1: $|L(x, x_{j})g(x^{**})| < h$ for some integrable function $h$. Assuming A1, by dominated convergence it follows that

$$
\hat{\text{R}}_{n} = \int L(x, x_{j})g(x^{**}) \, dx^{**} \xrightarrow{d} \int L(x, x_{j})g(x) \, dx = R
$$

(11)

i.e., the bootstrap estimate of risk converges weakly to the true risk. Let $\hat{\lambda}_{j}^{opt}$ represent the bootstrap choice that minimizes the left-hand side of (11) and let $\hat{\lambda}_{j}$ represent the value that minimizes the right-hand side of (11). By Theorem 2.1 of Léger and Romano (1990), $\hat{\lambda}_{j}^{opt} \to \hat{\lambda}_{j}$. Thus, the bootstrap estimator is "optimal" in the sense that $\hat{\lambda}_{j}^{opt}$ tends to $\hat{\lambda}_{j}$ asymptotically. Theoretically, a restriction on the cost function stems from assumption A1. In practice, this will be a mild requirement. Typically, $L(x, x_{j}) \to \infty$ as $|x - x_{j}| \to \infty$. Thus, one implication of A1 is that the loss function cannot increase at a faster rate than $g(x) \to 0$. In this context, a necessary condition is that $L(x, x_{j}) = o(g(x))$ where $o(\cdot)$ is the familiar ‘little o’ notation (observe that $g(x) = o(1)$ since $g(x)$ is a probability density).

Let $x_{j}^{**}$, $j = 1, 2, \ldots, J$ represent the bootstrap estimates of $x$ in the forecast period, and let $x_{j}$ represent the desired forecast. For a loss function $L(x - x_{j})$ the BP is the solution to the following program:
\[ \hat{x}_{j}^{**} = \min_{x_{j}} \sum_{j=1}^{J} L(x_{j}^{**} - x_{j}) \]  

(12)

As a practical matter, solving the above nonlinear program may require additional restrictions on the form of \( L(\cdot) \), for example convexity, depending upon the solver used to find the minimum. Provided that such assumptions are satisfied, a solution to this program can be found using traditional programming methods. The next section provides four examples. The first three examples demonstrate that the BP solves the same problem as each of the existing methods. The fourth example uses a loss function that cannot be approximated by the existing methods, yet is handled by BP. In addition to demonstrating the validity of BP, the examples that follow make clear how Eq. (12) may be implemented in practice for various loss functions.

**Four Examples**

The following experiment will be used to assess BP by comparing it with ALP, LXP, and AQP. A sample of size \( n + 1 \) is generated for the random variable \( X \). The first \( n \) observations are used to generate a prediction, which is then compared to observation \( n + 1 \) and loss is calculated. This experiment is repeated 1000 times with \( n = 100 \). To obtain reasonable accuracy in the tails of the bootstrap distribution, \( J = 999 \) (Efron, 1982). The program was written in RATS version 4.2 for the 386 PC (Doan, 1995), and takes approximately 30 minutes to run on a 133MHz Pentium. For analytic and computational simplicity, \( x \sim N(0, 1) \).

**Example 1: Asymmetric Linear Loss**

For AL loss with \( a = 0.2 \) and \( b = -0.8 \), the analytic solution is easily found. Since \( a/(a-b) = 0.200, \Phi^{-1}(0.200) = -0.8410 \), and from (3) the exact solution is \( x_{j} = -0.8410 \) after substituting \( \mu = 0 \) for \( \bar{x} \). Substituting \( x_{j} = -0.841 \) into (2) and taking expected value yields an expected cost of 0.2800. In practice, the ALP will use the sample mean and standard deviation, \( \bar{x} \) and \( s \), respectively. The BP solves the following program

\[ \min_{x_{j}} \sum_{j=1}^{J} 0.20(x_{j}^{**} - x_{j})\delta(x_{j}^{**} > x_{j}) - 0.80(x_{j}^{**} - x_{j})\delta(x_{j}^{**} < x_{j}) \]  

(13)

where \( \delta(x - x_{j}) = 1 \) if \( x < x_{j} \) and equals zero otherwise. This objective function has discontinuous derivatives, so solution is by the simplex method.

Simulation results are presented in Table 1 and Figure 1 displays the empirical distributions of forecasts and costs. The means and variances of the empirical
distributions of forecasts and costs are approximately equal. The ALP approximates the analytic solution, as does the BP. In particular, under linear loss and normality, the BP solves the same problem as the ALP. The means are roughly equivalent, though the variance of the ALP forecast is less than that of BP. This is not surprising, since the former is exact and contaminated only by sampling error, while the latter is consistent and contaminated not only by sampling error but by approximation error too.

### Table 1

**COMPARISON OF BP TO (1) ALP, (2) LXP, (3) AQP, AND (4) NUMERICAL SOLUTION OF SQUARE ROOT/SQUARE PROBLEM. RESULTS OF 1000 EXPERIMENTS**

(marginal significance level in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>$s\sqrt{1000}$</th>
<th>Hypothesis</th>
<th>Mean</th>
<th>$s\sqrt{1000}$</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ALP</td>
<td>-0.8401</td>
<td>0.0036</td>
<td>$t = 0.25$ (0.803)</td>
<td>0.2848</td>
<td>0.0075</td>
<td>$t = 0.64$ (0.522)</td>
</tr>
<tr>
<td></td>
<td>-0.8492</td>
<td>0.0046</td>
<td>$t = 1.78$ (0.075)</td>
<td>0.2860</td>
<td>0.0076</td>
<td>$t = 0.79$ (0.430)</td>
</tr>
<tr>
<td>2. LXP</td>
<td>0.4990</td>
<td>0.0039</td>
<td>$t = 0.25$ (0.803)</td>
<td>0.5307</td>
<td>0.0298</td>
<td>$t = 1.03$ (0.303)</td>
</tr>
<tr>
<td></td>
<td>0.4952</td>
<td>0.0043</td>
<td>$t = 1.12$ (0.263)</td>
<td>0.5332</td>
<td>0.0299</td>
<td>$t = 1.11$ (0.267)</td>
</tr>
<tr>
<td>3. AQP</td>
<td>-0.5491</td>
<td>0.0033</td>
<td>$t = 0.03$ (0.976)</td>
<td>0.3841</td>
<td>0.0186</td>
<td>$t = 0.48$ (0.625)</td>
</tr>
<tr>
<td></td>
<td>-0.5544</td>
<td>0.0036</td>
<td>$t = 0.41$ (0.159)</td>
<td>0.3867</td>
<td>0.0188</td>
<td>$t = 0.62$ (0.533)</td>
</tr>
<tr>
<td>4. BP</td>
<td>-0.3666</td>
<td>0.0052</td>
<td>$t = -1.082$ (0.279)</td>
<td>0.83133</td>
<td>0.02245</td>
<td>$t = -0.787$ (0.431)</td>
</tr>
</tbody>
</table>

![Graph](imageURL)

**Fig. 1.** Empirical distributions of forecasts (left) and costs (right) for ALP (solid lines) and BP (dashed lines)

*Jr. Comb., Inf. & Syst. Sci.*
EXAMPLE 2: LINEX LOSS

For LX loss with \( a = -1.0 \) and \( b = 1.0 \), from (5) the exact solution is \( x_f = 0.5 \) which inserted in (6) yields an expected cost of 0.5. In this setting, the BP is the solution to the following program

\[
\min \sum_{j=1}^{J} \exp \{ x_j^{**} - x_j \} - (x_j^{**} - x_j) - 1 
\]  

(14)

Simulation results are presented in Table 1, and Figure 2 displays the empirical distributions of forecasts and costs. The empirical distributions of forecasts have approximately equal means and variances, as do the cost distributions. Both the LXP and the BP approximate the exact solution: under linex loss and normality, LXP and BP solve the same problem.

EXAMPLE 3: ASYMMETRIC-QUADRATIC LOSS

For AQ loss with \( a = 0.2 \) and \( b = 0.8 \), the analytic solution is easily found. From (9) and (8), \( \delta^a = x_f = -0.5492 \) and from (10) expected cost is 0.375. The BP solves the following program:

\[
\min \sum_{j=1}^{J} 0.20(x_j^{**} - x_j)^2 \delta(x_j^{**} > x_j) - 0.80(x_j^{**} - x_j)^2 \delta(x_j^{**} < x_j) 
\]  

(15)

where \( \delta(x < x_j) = 1 \) if \( x < x_j \) and equals zero otherwise. This objective function has discontinuous derivatives and so solution is by the simplex method. Simulation results are presented in Table 1, and Figure 3 displays the empirical distributions.
Fig. 3. Empirical distributions of forecasts (left) and costs (right) for AQP (solid lines) and BP (dashed lines)

of forecasts and costs. The empirical distributions of forecasts and costs have approximately equal means and variances.

**EXAMPLE 4: SQUARE/SQUARE-ROOT LOSS**

Finally, consider a nonlinear loss function not amenable to approximation by AL, LX, or AQ. Let the loss function be

$$L(x - x_f) = \begin{cases} 
(x - x_f)^{1/2}, & x - x_f > 0, \\
(x - x_f)^2, & x - x_f < 0, \\
0, & x - x_f = 0 
\end{cases} \quad (16)$$

whence expected loss is given by

$$F(x_f) = \frac{1}{\sqrt{2\pi}} \int_{x_f}^{\infty} (x - x_f)^{1/2} \exp \left[-x^2/2\right] \, dx + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_f} (x - x_f)^2 \exp \left[-x^2/2\right] \, dx \quad (17)$$

Differentiating the above with respect to $x_f$ yields

$$F'(x_f) = -\frac{1}{2\sqrt{2\pi}} \int_{x_f}^{\infty} (x - x_f)^{-1/2} \exp \left[-x^2/2\right] \, dx + \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{x_f} (x - x_f) \exp \left[-x^2/2\right] \, dx \quad (18)$$

Solving (18) numerically using the secant method yields $x_f = -0.361$. Inserting this value into (17) yields expected cost of 0.849. We repeat the usual experiment. The BP predictor solves the following program:

$$\min_{x_f} \sum_{j=1}^{J} (x_f^{**} - x_f)^2 \delta(x_f^{**} > x_f) + (x_f^{**} - x_f)^{1/2} \delta(x_f^{**} < x_f) \quad (19)$$

*Jr. Comb., Inf. & Syst. Sci.*
Fig. 4. Loss function for interest rate prediction when the true interest rate is 0.10

Again, solution is by the simplex method. Results are presented in Table 1. BP approximates the numerical solution.

In each of the above examples, we have seen that BP approximates the exact analytic or numerical solution of the problem. These same experiments were repeated for a wide variety of non-Gaussian errors (results not reported), and in every case the BP proved markedly superior to the other methods. This is hardly surprising, since the BP can accommodate non-Gaussian errors, while the other methods cannot. In each case, though, the loss function was of the prediction-error form and the bootstrap of the future value was naive. The example in the following section has a nonlinear and asymmetric loss function that cannot be put in prediction-error form and that uses a bootstrap for autoregressive processes to generate the distribution of future values.

AN EXAMPLE WITH INTEREST RATES

As this example is purely illustrative, some simplifying assumptions are made. Consider a perpetual bond (consul) paying one unit per time period, so its price is \( 1/x \) where \( x \) is the interest rate. At the end of each period, before the beginning of the next period, trading takes place. After all trades are complete, the next period’s interest rate is revealed. Assume that markets are efficient so that if the next period’s interest rate is perfectly predicted, there are neither profits nor losses. The trader incurs a cost equal to the absolute value of the difference between the consul price implied by his forecast of the interest rate and the consul price implied by the
actual interest rate. Such a cost function is asymmetric and nonlinear, as can be seen from Figure 4. Note that it bears a strong resemblance to square/square-root loss, so we can expect the BP to handle this problem. A decrease in the interest rate of amount $\Delta x$ causes the price of the bond to rise by more than an increase in the interest rate of $\Delta x$ causes the price of the bond to fall.* Now let the interest rate in the future period be $x$ with prediction $x_f$. Underprediction, $x - x_f > 0$, causes a loss of $1/x_f - 1/x = (x - x_f)/x_f$, while overprediction, $x - x_f < 0$, causes a loss of $(x_f - x)/x_f$. These considerations give rise to a loss function of the form

$$L(x - x_f) = \begin{cases} 
(x - x_f)/x_f & x - x_f > 0 \\
-(x - x_f)/x_f & x - x_f < 0 \\
0 & x - x_f = 0 
\end{cases} \quad (20)$$

Due to its nonlinear nature it cannot be expressed in prediction-error loss form, and in fact depends not only on the magnitude of the error $x - x_f$ but upon the level of the process.

The BP solves the following program

$$\min_{x_f} \sum_{j=1}^{J} \frac{x_{f}^{**} - x_f}{x_{f}^{**} x_f} \delta(x_{f}^{**} > x_f) - \frac{x_{f}^{**} - x_f}{x_{f}^{**} x_f} \delta(x_{f}^{**} < x_f) \quad (21)$$

The interest rate series is the monthly secondary market rate for one-month CDs, series CDT1 on the St. Louis Federal Reserve database FRED, over the period 1983:1–1990:4 for a total of 88 observations. The student of economic history will note that during this period, interest rates were both high and volatile. The first four years are used to estimate a forecasting model, and then predictions are made for each of the next forty months, updating the model each period. The prediction for January 1987 is based on 40 observations, the prediction for February 1987 is based on 41 observations, etc.

An often recommended method of forecasting daily interest rates is Box-Jenkins ARIMA method. The usual identification procedures suggest an AR(1) model, since the partial autocorrelation cuts off abruptly after one lag and the autocorrelation declines gradually. The AR predictor is well-known to minimize quadratic loss. The Thomsbs-Scuchany (1990) method of bootstrapping autoregressive processes, as modified by McCullough (1994), is used to bootstrap the distribution of the interest rate in the future period. In the example at hand, BP does a better job of minimizing the cost of prediction error than the

*Suppose $x = 0.10$ so that the price is 10. If $x$ falls to 0.09 then the price rises to 11.11, while if $x$ rises to 0.11, then the price falls to 9.09.

*J. Econ., Inf. & Syst. Sci.*
Box-Jenkins method. The series of cost differences (cost based on Box-Jenkins prediction less cost based on BP prediction) has a mean of 0.04 and a standard deviation of the sample mean of 0.021, implying a t-statistic of 2.04 for two-sided test of the null hypothesis that the difference is zero. Thus, expected cost based on Box-Jenkins prediction exceeds expected cost based on BP. If the interest rate is 0.10, the consol price is $10; and while four cents on a $10 trade might not seem like much, it is about four dollars for a $1000 trade. In the Treasury bond market, the profit on a $1000 trade is less than ten dollars, so the proposed method might represent a substantial increase in the profit margin.

CONCLUSIONS

This paper sets forth a general method for prediction under asymmetric and nonlinear loss which combines two computationally intensive techniques: the bootstrap and nonlinear programming. Weak convergence of the bootstrap estimator is shown theoretically. Simulation results show that the bootstrap predictor approximates known results for asymmetric linear loss, linear-exponential loss, and asymmetric quadratic loss, whose solutions are known under a Gaussian assumption. In practice, if errors are known to be almost Gaussian and cost functions conform to asymmetric linear, linex, or asymmetric quadratic, then the appropriate analytic method might be preferred. In other cases, the bootstrap method might be preferred. As is typical of bootstrap applications the bootstrap predictor does not require a Gaussian assumption, so if errors are non-normal or if the cost function cannot be approximated by one of the asymmetric linear, linex, or asymmetric quadratic functions, then the bootstrap predictor might be preferred. The bootstrap predictor is consistent and subject to an approximation error that diminishes as sample size increases. More sophisticated bootstrap methods could improve the performance of the BP.

Given the dual estimation/programming requirements of this bootstrap predictor, many software packages will not be able to implement this method. Some packages do not easily perform bootstrapping routines. Other packages do not have routines to minimize an objective function. Those packages that have such minimization routines might not have a routine designed for this particular type of minimization problem. The Monte Carlo experiment originally was designed for a two variable regression, but the RATS solver too often failed to solve the associated programming problem. The same difficulty was encountered for setting $X \sim N(100, 10)$; scaling the problem to obtain a solution would have
reduced it to $X \sim N(0, 1)$. Hence, computational difficulties left only the simplest prediction problem to use in the Monte Carlo examples. Practical implementation of this procedure might well require specialized software and extensive computer programming, though RATS capably handled the empirical application presented in section six. Currently the method is limited to problems for which bootstrap prediction is feasible, such as i.i.d. and autoregressive cases. However, the current pace of bootstrap research suggests that these limitations will not long remain, and that bootstrap prediction for a much wider variety of models soon will exist. One can expect that, for most such methods of prediction, the method outlined in this paper will apply.

ACKNOWLEDGEMENTS

Thanks to D.A. Kendrick and A.L. Norman for reading previous versions of this paper, to S. Thore for reviewing the nonlinear programming methods, and for comments to participants at the 1997 International Conference on Combinatorics, Information Theory and Statistics.

REFERENCES


——— (1996), "Consistent forecast intervals when the future-period exogenous variables are stochastic", *Journal of Forecasting*, 15, 293–304.

S. Peters and D. Freedman (1985), "Using the bootstrap to evaluate forecasting equations", *Journal of Forecasting*, 4, 251–262.


