TECHNICAL WORKING PAPER SERIES

OPTIMAL PREDICTION UNDER ASYMMETRIC LOSS

Peter F. Christoffersen
Francis X. Diebold

Technical Working Paper No. 167

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1994

Helpful discussion was provided by Hashem Pesaran, Ken Wallis, and seminar participants at the Tinbergen Institute and the April 1994 San Diego Conference on Multivariate Financial Econometrics. All errors are ours alone. We thank the National Science Foundation, the Sloan Foundation and the University of Pennsylvania Research Foundation for support. This paper is part of NBER's research program in Economic Fluctuations. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #167
October 1994

OPTIMAL PREDICTION UNDER ASYMMETRIC LOSS

ABSTRACT

Prediction problems involving asymmetric loss functions arise routinely in many fields, yet the theory of optimal prediction under asymmetric loss is not well developed. We study the optimal prediction problem under general loss structures and characterize the optimal predictor. We compute the optimal predictor analytically in two leading cases. Analytic solutions for the optimal predictor are not available in more complicated cases, so we develop numerical procedures for computing it. We illustrate the results by forecasting the GARCH(1,1) process which, although white noise, is non-trivially forecastable under asymmetric loss.

Peter F. Christoffersen
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297

Francis X. Diebold
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
and NBER

1. Introduction

A moment's reflection yields the insight that prediction problems involving asymmetric loss structures arise routinely, as a myriad of situation-specific factors may render "positive" errors more (or less) costly than "negative" errors, or conversely. The potential necessity of allowing for asymmetric loss has been acknowledged for some time. Granger and Newbold (1986), for example, note that although "an assumption of symmetry about the conditional mean ... is likely to be an easy one to accept, ... an assumption of symmetry for the cost function is much less acceptable" (p. 125). The applied literature has echoed this sentiment, in fields ranging from business-cycle forecasting and financial forecasting, to sales forecasting, electricity peak-load forecasting, and government revenue forecasting.[1]

Moreover, recent developments in economic theory suggest that asymmetric loss may arise quite generally. On the demand side, "loss aversion" is a fundamental component of the influential non-expected utility theory of Kahneman and Tversky (1979) and Tversky and Khaneman (1991). Loss aversion refers to situations in which present wealth serves as a focal point, with the utility of wealth rising only gradually above that point but falling sharply below that point.[2]

On the supply side, Stockman (1987) argues that decision making by firms, especially in financial markets, leads naturally to asymmetric loss functions. The intuition is that, in realistic market conditions, the ultimate variables that economic agents seek to influence (e.g., value of the firm) are likely to depend *nonlinearly* on profits obtained from speculation based on predictions (e.g., exchange rate predictions).

---

[1] See, for example, many of the articles in Makridakis and Wheelwright (1987).

[2] For an early development, see also Roy (1952).

This paper is part of a research program aimed at allowing for general loss structures in model selection, estimation, prediction, and forecast evaluation. Recently a number of authors have made progress toward that goal, including Weiss and Anderson (1984) and Weiss (1991, 1994) on model selection and estimation, and Diebold and Mariano (1994) on forecast evaluation. In this paper, we focus on prediction, beginning where Granger's (1969) classic paper ends. In section 2, we extend Granger's results to the case of conditionally Gaussian, but unconditionally non-Gaussian, processes. Among other things, this allows treatment of conditionally heteroskedastic processes. In section 3, we provide analytic solutions for the optimal predictor under two popular asymmetric loss functions. In section 4, we provide methods for computing the optimal predictor under more general loss functions for which analytic solutions are not possible, and also for processes that are not conditionally Gaussian. In section 5, we illustrate our results and methods by forecasting the GARCH(1,1) process. We conclude in Section 6.

2. Optimal Predictors for Unconditionally and Conditionally Gaussian Processes

Granger (1969) establishes optimality of the conditional mean predictor under symmetric loss, so long as the conditional distribution of the process being predicted is symmetric and one of two technical assumptions is met. Under asymmetric loss the conditional mean is no longer optimal, but a simple translation of the conditional mean *is* optimal for unconditionally Gaussian processes. More precisely,

Theorem 1 (Granger) If $\{y_t\} \sim N(\mu, \Sigma)$ is a Gaussian process and $L(e_{t+h})$ is any loss function defined on the h-step-ahead prediction error, $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$, then the optimal predictor is of the form $\hat{y}_{t+h} = \mu_{t+h|t} + \alpha$, where $\mu_{t+h|t} = E(y_{t+h}|\Omega_t)$, is the conditional mean,

$\Omega_t = \{y_t, y_{t-1}, \ldots\}$, and α is a constant that depends only on the loss function and the (constant) conditional prediction-error variance.

Granger's fundamental theorem has two key limitations. First, the process is assumed to be Gaussian, which implies a constant conditional prediction-error variance. This is unfortunate because conditional heteroskedasticity is widespread in economic and financial data.[3] Second, the loss function is defined directly on prediction errors. More general functions of predictions and realizations are ruled out.

Let us begin, then, by generalizing Granger's theorem to allow for conditional variance dynamics. We shall subsequently allow for more general loss functions and more general conditional distributions as well.

Theorem 2 If $y_{t+h}|\Omega_t \sim N(\mu_{t+h|t}, \sigma^2_{t+h|t})$ is a conditionally Gaussian process and $L(e_{t+h})$ is any loss function defined on the h-step-ahead prediction error e_{t+h}, then the optimal predictor is of the form $\hat{y}_{t+h} = \mu_{t+h|t} + \alpha_t$, where α_t depends only on the loss function and the conditional prediction-error variance $\sigma^2_{t+h|t} = \text{var}(y_{t+h}|\Omega_t) = \text{var}(e_{t+h}|\Omega_t)$.

Proof We seek the predictor that solves[4]

$$\min_{\hat{y}_{t+h}} E_t\{L(y_{t+h}-\hat{y}_{t+h})\} = \min_{\hat{y}_{t+h}} \int_{-\infty}^{\infty} L(y_{t+h}-\hat{y}_{t+h})\, f(y_{t+h}|\Omega_t)\, dy_{t+h}.$$

Without loss of generality we can write $\hat{y}_{t+h} = \mu_{t+h|t} + \alpha_t$ and $y_{t+h} = \mu_{t+h|t} + x_{t+h}$, so that

$$\underset{\hat{y}_{t+h}}{\text{argmin}}\, E_t\{L(y_{t+h}-\hat{y}_{t+h})\} = \underset{\alpha_t}{\text{argmin}} \int_{-\infty}^{\infty} L(x_{t+h}-\alpha_t)\, f(x_{t+h}|\Omega_t)\, dx_{t+h}.$$

Because $f(x_{t+h}|\Omega_t)$ depends on $\sigma^2_{t+h|t}$ but not $\mu_{t+h|t}$, so too does the α_t that solves the

[3] See Bollerslev, Engle and Nelson (1994) and Diebold and Lopez (1994).

[4] Here and throughout, $E_t(x)$ denotes $E(x|\Omega_t)$.

minimization problem depend on $\sigma^2_{t+h|t}$ but not $\mu_{t+h|t}$. ■

The upshot is that the optimal predictor under conditional normality is not necessarily just a constant added to the conditional mean, because the conditional prediction-error variance may be time-varying. The class of conditionally Gaussian GARCH processes, for example, falls under the jurisdiction of Theorem 2. Thus, under asymmetric loss, conditional variance dynamics are important not only for interval prediction, but also for *point* prediction. If loss is asymmetric but conditional heteroskedasticity is ignored, the resulting point predictions will be suboptimal and may have dramatically greater conditionally expected loss in consequence.

In closing this section, we note that its result (Theorem 2) depends crucially on conditional normality. It is apparent, moreover, that even if we maintain the conditional normality assumption but let loss be a general function of predictions and realizations, then the additive "correction" to the conditional mean predictor no longer obtains. That is, if $y_{t+h}|\Omega_t \sim N(\mu_{t+h|t}, \sigma^2_{t+h|t})$ and $L(y_{t+h}, \hat{y}_{t+h})$ is not of the form $L(y_{t+h}-\hat{y}_{t+h})$, then although the optimal predictor depends only on $\mu_{t+h|t}$ and $\sigma^2_{t+h|t}$, it appears impossible to characterize its form.

3. Analytic Solutions For Two Leading Loss Functions

Here we examine two asymmetric loss functions for which it is possible to solve analytically for the optimal predictor under conditional normality. We work with the process $y_{t+h}|\Omega_t \sim N(\mu_{t+h|t}, \sigma^2_{t+h|t})$. For each loss function, we characterize the optimal predictor, $\hat{y}_{t+h} = \mu_{t+h|t} + \alpha_t$, and we compare its conditionally expected loss to that of two competitors:

(1) the conditional mean, $\mu_{t+k|t}$.

(2) the pseudo-optimal predictor, $\hat{y}_{t+k} = \mu_{t+k|t} + \alpha$, where α depends only on the loss function and the unconditional prediction-error variance, $\sigma_k^2 = \text{var}(e_{t+k})$.

Note that the optimal predictor acknowledges loss asymmetry and the possibility of conditional heteroskedasticity through a possibly *time-varying* adjustment to the conditional mean. The conditional mean, in contrast, is always suboptimal as it incorporates *no* adjustment. The pseudo-optimal predictor is intermediate in the sense that it incorporates only a *constant* adjustment for asymmetry; thus, it is fully optimal only in the conditionally homoskedastic case for which $\sigma_k^2 = \sigma_{t+k|t}^2$. We include an examination of the conditionally expected loss of the pseudo-optimal predictor in order to explore the cost of ignoring conditional heteroskedasticity under asymmetric loss.

3a. Linex Loss

The "linex" loss function, introduced by Varian (1974) and further studied by Zellner (1986), is given by

$$L(x) = b\left[\exp(ax) - ax - 1\right], \quad a \in \mathbb{R}\backslash\{0\},\ b \in \mathbb{R}_+.$$

The linex loss function is so-named for its almost linear shape on one side of the origin, and almost exponential shape on the other. The parameter "a" plays an important role in the linex loss function; when a > 0 the loss function is approximately linear to the left of the origin and approximately exponential to the right, and conversely when a < 0. Moreover, quadratic loss is approximately nested within linex loss, because if a is small, one can accurately approximate the exponential part of the loss function by the first two terms in its Taylor series expansion, yielding

$$L(x) \approx b\left[\left(1 + ax + \frac{a^2x^2}{2}\right) - ax - 1\right] = \frac{ba^2}{2}x^2,$$

which is just a quadratic loss function, $L(x) = cx^2$. Thus, for small a, we expect the optimal predictor with respect to linex loss to be close to the optimal predictor with respect to quadratic loss. Various shapes of linex loss functions are illustrated in Figure 1.

Now let us compute the optimal h-step-ahead predictor under linex loss. The optimal predictor solves

$$\min_{\hat{y}_{t+h}} E_t \left\{b\left[\exp(a(y_{t+h}-\hat{y}_{t+h})) - a(y_{t+h}-\hat{y}_{t+h}) - 1\right]\right\},$$

which is equivalent to

$$b \min_{\hat{y}_{t+h}} \left\{E_t(\exp(ay_{t+h}))\exp(-a\hat{y}_{t+h}) - aE_t(y_{t+h}) + a\hat{y}_{t+h} - 1\right\}.$$

Before performing the minimization, we evaluate the expectation of the exponential by using the conditional moment-generating function for a conditionally Gaussian variate,

$$E_t\{\exp(ay_{t+h})\} = \exp\left(a\mu_{t+h|t} + \frac{a^2\sigma^2_{t+h|t}}{2}\right).$$

Substituting this expression into the conditionally expected linex loss function, and then differentiating, we obtain the first-order condition that defines the optimal predictor,

$$\hat{y}_{t+h} = \mu_{t+h|t} + \frac{a}{2}\sigma^2_{t+h|t}.$$

Similar calculations reveal that the pseudo-optimal predictor is

$$\hat{y}_{t+h} = \mu_{t+h|t} + \frac{a}{2}\sigma_h^2,$$

where $\sigma_h^2 = \mathrm{var}(e_{t+h})$ is the unconditional h-step-ahead prediction-error variance.

Theorem 2 shows that the optimal predictor under conditional normality is the conditional mean plus a function of the conditional prediction-error variance. Under linex loss, the function is a simple linear one, depending on the degree of asymmetry of the loss function, as captured in the parameter a.[5] The intuition is simple--when a is positive, positive prediction errors are more devastating than negative errors, so a negative conditionally expected error is desirable. The optimal amount of bias depends on the conditional prediction-error variance of the process. As the conditionally expected variation around the conditional expectation grows, so too does the optimal amount of bias, in order to avoid devastating large positive prediction errors.

Let us now compute the conditionally expected linex loss of the optimal, pseudo-optimal and conditional mean predictors. Consider first the optimal predictor, which when inserted in the conditionally expected loss expression yields

$$E_t\left\{ b\left[\exp(a(y_{t+h}-\mu_{t+h|t}-\frac{a\sigma_{t+h|t}^2}{2})) - a(y_{t+h}-\mu_{t+h|t}-\frac{a\sigma_{t+h|t}^2}{2}) - 1\right]\right\} = \frac{ba^2\sigma_{t+h|t}^2}{2}.$$

The conditionally expected loss of the pseudo-optimal predictor, on the other hand, is

[5] The correction is non-trivial for values of a not too close to zero. When a is close to zero the conditional mean is close to optimal, reflecting the fact that linex loss is then close to quadratic loss.

$$E_t\left\{L(y_{t+h} - \hat{y}_{t+h})\right\} = b\left[\exp(\frac{a^2}{2}(\sigma^2_{t+h|t}-\sigma^2_h)) + \frac{a^2}{2}\sigma^2_h - 1\right].$$

By construction, the conditionally expected loss of the pseudo-optimal predictor is at least as large as that of the optimal predictor.

Finally, when using the conditional mean as the predictor, the resulting conditionally expected linex loss is $b[\exp(\frac{a^2\sigma^2_{t+h|t}}{2}) - 1]$, which is strictly larger than the linex-optimal loss for all admissible values of a and b.

Notice that, perhaps contrary to expectations, it is not possible in general (i.e., for all values of $\sigma^2_{t+h|t}$) to rank the pseudo-optimal as superior to the conditional mean predictor in terms of conditionally expected loss. For the conditionally expected loss of the conditional mean to be higher than that of the pseudo-optimal predictor we need

$$\exp\left[\frac{a^2}{2}\sigma^2_{t+h|t}\right] - 1 \overset{?}{\geq} \exp\left[\frac{a^2}{2}(\sigma^2_{t+h|t}-\sigma^2_h)\right] + \frac{a^2}{2}\sigma^2_h - 1.$$

Subtracting $\frac{a^2}{2}\sigma^2_{t+h|t}$ from both sides yields

$$\exp\left[\frac{a^2}{2}\sigma^2_{t+h|t}\right] - \left[1 + \frac{a^2}{2}\sigma^2_{t+h|t}\right] \overset{?}{\geq} \exp\left[\frac{a^2}{2}(\sigma^2_{t+h|t}-\sigma^2_h)\right] - (1 + \frac{a^2}{2}(\sigma^2_{t+h|t}-\sigma^2_h)).$$

Because $\exp(x) - (1 + x)$ is increasing in x for $x > 0$, we infer that the conditionally expected loss of the conditional mean is indeed larger that the conditionally expected loss from the pseudo-optimal predictor when $\sigma^2_{t+h|t} > \sigma^2_h$. However, for a sufficiently small value of $\sigma^2_{t+h|t}$ (depending non-linearly on the value of a and σ^2_h) the conditionally expected loss of the conditional mean will be smaller than that of the pseudo-optimal predictor. The intuition is simply that in very low volatility times the conditionally optimal amount of bias is very small, resulting in a higher conditionally expected loss for the pseudo-optimal predictor (which injects too much bias) than for the conditional mean (which injects no bias).

These effects are illustrated in Figure 2, in which we plot conditionally expected loss as a function of $\sigma^2_{t+h|t}$, for each of the three predictors. As must be the case, the loss of the optimal predictor is always lowest. The losses of the pseudo-optimal and the optimal predictors coincide when $\sigma^2_{t+h|t} = \sigma^2_h = 1$. The loss of the conditional mean intersects the loss of the pseudo-optimal predictor from below at some value of $\sigma^2_{t+h|t}$ less than σ^2_h. As $\sigma^2_{t+h|t}$ gets close to zero the optimal predictor makes only a very small correction to the conditional mean and the losses of the two predictors get very close.

3b. Linlin Loss

Granger (1969) considers the simple asymmetric loss function,

$$L(y_{t+h}-\hat{y}_{t+h}) = \begin{cases} a|y_{t+h}-\hat{y}_{t+h}|, & \text{if } y_{t+h}-\hat{y}_{t+h} > 0 \\ b|y_{t+h}-\hat{y}_{t+h}|, & \text{if } y_{t+h}-\hat{y}_{t+h} \leq 0. \end{cases}$$

We call this the linlin loss function, because it is linearly increasing on both sides of the origin with the slopes a and b, respectively. The degree of asymmetry depends on the ratio of a to b. Various shapes of linlin loss functions are shown in Figure 3.

In the linlin case, the optimal predictor solves

$$\min E_t\,[L(y_{t+h}-\hat{y}_{t+h})] = a\int_{\hat{y}_{t+h}}^{\infty}(y_{t+h}-\hat{y}_{t+h})f(y_{t+h}|\Omega_t)dy_{t+h} - b\int_{-\infty}^{\hat{y}_{t+h}}(y_{t+h}-\hat{y}_{t+h})f(y_{t+h}|\Omega_t)dy_{t+h}.$$

The first-order condition is

$$-a(1-F(\hat{y}_{t+h}|\Omega_t)) + b\,F(\hat{y}_{t+h}|\Omega_t) = 0,$$

which is equivalent to

$$F(\hat{y}_{t+k}|\Omega_t) = \frac{a}{a+b},$$

where $F(y_{t+k}|\Omega_t)$ is the conditional c.d.f. of y_{t+k}.[6]

In the conditionally Gaussian case $y_{t+k}|\Omega_t \sim N(\mu_{t+k},\sigma^2_{t+k|t})$, we have

$$F(\hat{y}_{t+k}|\Omega_t) = Pr((y_{t+k} \leq (\mu_{t+k|t} + \alpha_t))|\Omega_t) = Pr\left(\left(\frac{y_{t+k}-\mu_{t+k|t}}{\sigma_{t+k|t}} \leq \frac{\alpha_t}{\sigma_{t+k|t}}\right)|\Omega_t\right) = \Phi\left(\frac{\alpha_t}{\sigma_{t+k|t}}\right) = \frac{a}{a+b},$$

where $\Phi(z)$ is the N(0,1) c.d.f., values for which are readily available in tables. From this we can obtain the α_t that when added to the conditional mean yields the optimal predictor,

$$\alpha_t = \sigma_{t+k|t}\,\Phi^{-1}\left(\frac{a}{a+b}\right).$$

Thus, the optimal predictor is

$$\hat{y}_{t+k} = \mu_{t+k|t} + \sigma_{t+k|t}\,\Phi^{-1}\left(\frac{a}{a+b}\right).$$

Similar calculations reveal that the pseudo-optimal predictor, which ignores conditional variance dynamics, is

$$\hat{y}_{t+k} = \mu_{t+k|t} + \sigma_k\,\Phi^{-1}\left(\frac{a}{a+b}\right).$$

Thus, with linlin loss, as with linex loss, it is easy to allow for conditional variance dynamics when constructing the optimal predictor.

Now let us compute conditionally expected linlin loss with the optimal, pseudo-

---

[6] Note that the optimal predictor depends only on the ratio a/b, because a/(a+b) = (a/b)/((a/b)+1).

optimal and conditional mean predictors. In general, the conditionally expected loss of the linlin predictor, $\hat{y}_{t+h}$, is

$$E_t\{L(y_{t+h}-\hat{y}_{t+h})\} = a\int_{\hat{y}_{t+h}}^{\infty}(y_{t+h}-\hat{y}_{t+h})f(y_{t+h}|\Omega_t)dy_{t+h} - b\int_{-\infty}^{\hat{y}_{t+h}}(y_{t+h}-\hat{y}_{t+h})f(y_{t+h}|\Omega_t)dy_{t+h}$$

Recall the formulae for the truncated expectation,

$$E_t\{y_{t+h}|(y_{t+h}>\hat{y}_{t+h})\} = \frac{\int_{\hat{y}_{t+h}}^{\infty} y_{t+h}f(y_t|\Omega_t)dy_{t+h}}{1-F(\hat{y}_{t+h})}$$

$$E_t\{y_{t+h}|(y_{t+h}<\hat{y}_{t+h})\} = \frac{\int_{-\infty}^{\hat{y}_{t+h}} y_{t+h}f(y_{t+h}|\Omega_t)dy_{t+h}}{F(\hat{y}_{t+h})},$$

and substitute into the expected loss expression to obtain

$$E_t\{L(y_{t+h}-\hat{y}_{t+h})\} = a(1-F(\hat{y}_{t+h}|\Omega_t))[E_t\{y_{t+h}|(y_{t+h}>\hat{y}_{t+h})\} - \hat{y}_{t+h}] - bF(\hat{y}_{t+h}|\Omega_t)[E_t\{y_{t+h}|(y_{t+h}<\hat{y}_{t+h})\} - \hat{y}_{t+h}].$$

Invoke conditional normality and use the properties of the truncated normal distribution to obtain

$$E_t\{y_{t+h}|(y_{t+h}>\hat{y}_{t+h})\} = \mu_{t+h|t} + \sigma_{t+h|t}\frac{\phi(\xi_t)}{1-\Phi(\xi_t)}$$

$$E_t\{y_{t+h}|(y_{t+h}<\hat{y}_{t+h})\} = \mu_{t+h|t} - \sigma_{t+h|t}\frac{\phi(\xi_t)}{\Phi(\xi_t)},$$

where $\xi_t = \frac{\hat{y}_{t+h}-\mu_{t+h|t}}{\sigma_{t+h|t}}$ and $\phi(\cdot)$ is the N(0,1) c.d.f. Substitute into the conditionally expected loss expression to obtain (after some algebraic manipulation)

$$E_t\{L(y_{t+h}-\hat{y}_{t+h})\} = (a+b)\sigma_{t+h|t}\phi(\xi_t) - a(\hat{y}_{t+h|t}-\mu_{t+h|t}) + (a+b)\Phi(\xi_t)(\hat{y}_{t+h|t}-\mu_{t+h|t}).$$

For the optimal predictor,

$$\xi_t = \Phi^{-1}\left(\frac{a}{a+b}\right),$$

yielding an expected loss of

$$E_t\{L(y_{t+h}-\hat{y}_{t+h})\} = (a+b)\sigma_{t+h|t}\phi\left(\Phi^{-1}\left(\frac{a}{a+b}\right)\right).$$

For the pseudo-optimal predictor,

$$\xi_t = \Phi^{-1}\left(\frac{a}{a+b}\right)\frac{\sigma_h}{\sigma_{t+h|t}},$$

yielding an expected loss of

$$E_t\{L(y_{t+h}-\hat{y}_{t+h})\} = (a+b)\sigma_{t+h|t}\phi\left(\Phi^{-1}\left(\frac{a}{a+b}\right)\frac{\sigma_h}{\sigma_{t+h|t}}\right)$$

$$- a\Phi^{-1}\left(\frac{a}{a+b}\right)\frac{\sigma_h}{\sigma_{t+h|t}} + (a+b)\Phi\left(\Phi^{-1}\left(\frac{a}{a+b}\right)\frac{\sigma_h}{\sigma_{t+h|t}}\right)\Phi^{-1}\left(\frac{a}{a+b}\right)\frac{\sigma_h}{\sigma_{t+h|t}}.$$

For the conditional mean predictor, $\xi_t = 0$, yielding an expected loss of

$$E_t\{L(y_{t+h}-\hat{y}_{t+h})\} = (a+b)\sigma_{t+h|t}\phi(0).$$

Figure 4 shows that, just as in the linex loss case of Figure 2, the linlin-optimal and pseudo-optimal predictors coincide when the conditional and unconditional prediction error variances coincide (at 1). Also, for a sufficiently low conditional prediction error variance, the loss of the pseudo-optimal predictor is actually higher than that of the conditional mean, which in turn approaches zero as does the loss of the optimal predictor.

In closing this section, we note that in the linlin case (in contrast to the linex case) it is very easy, even for general conditional distributions, to find the optimal predictor--just draw the conditional c.d.f. and read the value on the x-axis corresponding to a/(a+b). More formally,

$$\hat{y}_{t+k} = F^{-1}\left(\frac{a}{a+b} \mid \Omega_t\right),$$

so $\hat{y}_{t+k}$ is simply the (a/(a+b)th conditional quantile. When a=b, of course, $\hat{y}_{t+k}$ is the conditional median.

4. General Solutions to the Optimal Prediction Problem

We have characterized the optimal predictor under conditional normality, and we have computed it analytically in some leading special cases. In most cases, however, it is impossible to solve analytically for the optimal predictor. To see the difficulty associated with analytic solution, even for very simple loss functions and under conditional normality, consider the following natural generalization of quadratic loss ("quadquad" loss), in which loss is quadratic on each side of the origin, but positive errors cost more than negative errors (or conversely):

$$L(y_{t+k}-\hat{y}_{t+k}) = \begin{cases} a(y_{t+k}-\hat{y}_{t+k})^2, \text{ if } y_{t+k}-\hat{y}_{t+k} > 0 \\ b(y_{t+k}-\hat{y}_{t+k})^2, \text{ if } y_{t+k}-\hat{y}_{t+k} \leq 0 \end{cases}$$

Conditionally expected loss is

$$E_t\{L(y_{t+k}-\hat{y}_{t+k})\} = a\int_{\hat{y}_{t+k}}^{\infty}(y_{t+k}-\hat{y}_{t+k})^2 f(y_{t+k}|\Omega_t)\, dy_{t+k} + b\int_{-\infty}^{\hat{y}_{t+k}}(y_{t+k}-\hat{y}_{t+k})^2 f(y_{t+k}|\Omega_t)\, dy_{t+k}.$$

Differentiating with respect to the predictor, we obtain the first order condition

$$a\int_{\hat{y}_{t+k}}^{\infty}(y_{t+k}-\hat{y}_{t+k}) f(y_{t+k}|\Omega_t)\, dy_{t+k} + b\int_{-\infty}^{\hat{y}_{t+k}}(y_{t+k}-\hat{y}_{t+k}) f(y_{t+k}|\Omega_t)\, dy_{t+k} = 0.$$

It is clear that analytic solution of this first-order condition is impossible in general. Moreover, even in cases as highly-structured as conditional normality, analytic solution remains impossible. To see this, rewrite the first-order condition as

$$a(1-F(\hat{y}_{t+k}|\Omega_t))(E[y_{t+k}|(y_{t+k}>\hat{y}_{t+k})] - \hat{y}_{t+k}) + bF(\hat{y}_{t+k}|\Omega_t)(E[y_{t+k}|(y_{t+k}<\hat{y}_{t+k})] - \hat{y}_{t+k}) = 0.$$

Under conditional normality, expressions for the truncated expectations are given above. Inserting these representations for the truncated expectations, using $F(\hat{y}_{t+k}|\Omega_t) = \Phi(\xi_t)$ and cancelling terms yields[7]

$$(a-b)\phi(\xi_t)\sigma_{t+k|t} + (a-b)\Phi(\xi_t)(\hat{y}_{t+k}-\mu_{t+k|t}) - a(\hat{y}_{t+k}-\mu_{t+k|t}) = 0.$$

Thus, although conditional normality does yield some simplification, closed-form analytic solution remains impossible.

Existence and uniqueness of the optimal predictor are easily established under conditional normality, however. Denote the first-order condition that defines the optimal predictor by $g(\hat{y}_{t+k}) = 0$. Existence follows from $\lim_{\hat{y}_{t+k}\to\infty} g(\hat{y}_{t+k}) < 0$ and

[7] Notice that for a=b the conditional mean is of course optimal.

$\lim_{\hat{y}_{t+k} \to -\infty} g(\hat{y}_{t+k}) > 0$, together with continuity of the first-order condition. The two limits are easily verified; immediately, $\lim_{\hat{y}_{t+k} \to \infty} g(\hat{y}_{t+k}) = -\infty$ and $\lim_{\hat{y}_{t+k} \to -\infty} g(\hat{y}_{t+k}) = +\infty$. For uniqueness we need that $g'(\hat{y}_{t+k})$ be strictly negative everywhere. This too is easily verified; immediately,

$$g'(\hat{y}_{t+k}) = -a(1-\Phi(\xi_t)) - b\Phi(\xi_t),$$

which is strictly negative everywhere, because $a > 0$, $b > 0$ and $\Phi(\cdot)$ is a c.d.f.

This situation, in which existence and uniqueness of the optimal predictor are easily established but a closed-form does not exist, is typical. The good news, however, is that in such situations numerical algorithms (nonlinear equation solution algorithms in conjunction with numerical integration) may be used to compute the optimal predictor quickly and reliably, ***even in conditionally non-Gaussian cases***. One such loss function, and one that is particularly appealing in light of its simplicity and flexibility, is the piecewise-linear loss function.

4a. Piecewise-Linear Approximation of the Loss Function

Consider a loss function L(e) constructed by concatenating linear segments, such that the loss of zero is zero and it is nondecreasing on both sides of the origin. This may actually *be* the relevant loss function, or it may be used as an approximation to any prediction-error loss function.[8] Conveniently, the optimal predictor associated with this loss function falls into the class for which existence and uniqueness are easily established, and it is easily computed numerically.

Conditionally expected loss is

[8] Note that any desired level of approximation accuracy may be obtained by taking sufficiently many segments.

$$E_t\{L(y_{t+h}-\hat{y}_{t+h})\} = \sum_{i=1}^{I-1}\int_{\hat{y}_{t+h}+c_{i-1}}^{\hat{y}_{t+h}+c_i}(a_i(y_{t+h}-\hat{y}_{t+h})+b_i)f(y_{t+h}|\Omega_t)dy_{t+h} + \int_{\hat{y}_{t+h}+c_{I-1}}^{\infty}(a_I(y_{t+h}-\hat{y}_{t+h})+b_I)f(y_{t+h}|\Omega_t)dy_{t+h}$$

$$+ \sum_{j=1}^{J-1}\int_{\hat{y}_{t+h}+c^j}^{\hat{y}_{t+h}+c^{j-1}}(a^j(y_{t+h}-\hat{y}_{t+h})+b^j)f(y_{t+h}|\Omega_t)dy_{t+h} + \int_{-\infty}^{\hat{y}_{t+h}+c^{J-1}}(a^J(y_{t+h}-\hat{y}_{t+h})+b^J)f(y_{t+h}|\Omega_t)dy_{t+h}.$$

The first line denotes the pieces on the positive side of the origin and the second line the negative, i.e., $a_i \geq 0$, $\forall i$ and $a^j \leq 0$, $\forall j$. The c_i's and c^j's denote the breakpoints between segments, with $c^l < c^k < 0$ and $0 < c_k < c_l$, for all $l > k$. For zero loss at the origin we impose $b_1 = b^1 = c_0 = c^0 = 0$. To ensure that neighboring segments connect at the breakpoints we impose $b_i = b_{i-1} + (a_{i-1}-a_i)c_{i-1}$, $i = 2, 3, \ldots, I$, and similarly $b^j = b^{j-1} + (a^{j-1}-a^j)c^{j-1}$, $j = 2, 3, \ldots, J$.[9]

Differentiating with respect to the predictor, $\hat{y}_{t+h}$, and using Leibniz's rule, we obtain

$$\sum_{i=1}^{I-1}(a_ic_i+b_i)f(\hat{y}_{t+h}+c_i|\Omega_t) - \sum_{i=1}^{I-1}(a_ic_{i-1}+b_i)f(\hat{y}_{t+h}+c_{i-1}|\Omega_t) - (a_Ic_{I-1}+b_I)f(\hat{y}_{t+h}+c_{I-1}|\Omega_t)$$

$$- \sum_{i=1}^{I-1}a_i(F((\hat{y}_{t+h}+c_i)|\Omega_t)-F(\hat{y}_{t+h}+c_{i-1})) - a_I(1-F((\hat{y}_{t+h}+c_{I-1})|\Omega_t))$$

$$+ \sum_{j=1}^{J-1}(a^jc^{j-1}+b^j)f((\hat{y}_{t+h}+c^{j-1})|\Omega_t) + (a^Jc^{J-1}+b^J)f((\hat{y}_{t+h}+c^{J-1})|\Omega_t) - \sum_{j=1}^{J-1}(a^jc^j+b^j)f((\hat{y}_{t+h}+c^j)|\Omega_t)$$

$$- \sum_{j=1}^{J-1}a^j(F((\hat{y}_{t+h}+c^{j-1})|\Omega_t)-F((\hat{y}_{t+h}+c^j)|\Omega_t)) - a^JF((\hat{y}_{t+h}+c^{J-1})|\Omega_t) = 0,$$

which is the first-order condition that defines the optimal predictor. After some

---

[9] The familiar linlin loss function is of course a special case of the piecewise-linear loss function, corresponding to $a_i = a_1$, $i = 2, 3, \ldots, I$ and $a^j = a^1$, $j = 2, 3, \ldots, J$.

manipulation all p.d.f. terms cancel, leaving

$$-\sum_{i=1}^{I-1} a_i(F((\hat{y}_{t+k}+c_i)|\Omega_t)-F((\hat{y}_{t+k}+c_{i-1})|\Omega_t)) - a_I(1-F((\hat{y}_{t+k}+c_{I-1})|\Omega_t))$$

$$-\sum_{j=1}^{J-1} a^j(F((\hat{y}_{t+k}+c^{j-1})|\Omega_t)-F((\hat{y}_{t+k}+c^j)|\Omega_t)) - a^J F((\hat{y}_{t+k}+c^{J-1})|\Omega_t) = 0,$$

or equivalently (after a bit more manipulation),

$$\sum_{i=2}^{I}(a_i-a_{i-1})F((\hat{y}_{t+k}+c_{i-1})|\Omega_t) + \sum_{j=2}^{J}(a^{j-1}-a^j)F((\hat{y}_{t+k}+c^{j-1})|\Omega_t) + (a_1-a^1)F((\hat{y}_{t+k})|\Omega_t) - a_1 = 0.$$

This first-order condition cannot be solved analytically, but it is easy to solve numerically, given the conditional c.d.f. $F(y_{t+k}|\Omega_t)$. Sufficient conditions for existence and uniqueness of the solution are given in the following theorem.

Theorem 3 If:

(1) $a_i \geq a_{i-1}$, $i = 2, 3, \ldots, I$ and $a^{j-1} \geq a^j$, $j = 2, 3, \ldots, J$

(2) $f(y|\Omega) > 0$, $\forall y$

(3) $a_i > a_{i-1}$ for some i, or $a^{j-1} > a^j$, for some j,

then a solution to the first-order condition exists and is unique.

Proof Denote the first-order condition by $g(\hat{y}_{t+k}) = 0$. We shall show that $\lim_{\hat{y}_{t+k}\to\infty} g(\hat{y}_{t+k}) > 0$ and $\lim_{\hat{y}_{t+k}\to-\infty} g(\hat{y}_{t+k}) < 0$, so that the first-order condition has at least one root, by continuity of $g(\cdot)$. Immediately, $\lim_{\hat{y}_{t+k}\to\infty} g(\hat{y}_{t+k}) = -a^J$ and $\lim_{\hat{y}_{t+k}\to-\infty} g(\hat{y}_{t+k}) = -a_1$. These limits are strictly positive and negative, respectively, by condition (3) in conjunction with the fact that the a_i's are all non-negative and the a^j's are all non-positive. Now we establish uniqueness by showing that $g'(\hat{y}_{t+k}) > 0$, $\forall \hat{y}_{t+k}$. Immediately,

Notice that all terms are nonnegative from condition (1) in conjunction with the fact that the

$$g'(\hat{y}_{t+k}) = \sum_{i=2}^{I} (a_i - a_{i-1}) f((\hat{y}_{t+k} + c_{i-1}) | \Omega_t) + \sum_{j=2}^{J} (a^{j-1} - a^j) f((\hat{y}_{t+k} + c^{j-1}) | \Omega_t) + (a_1 - a^1) f(\hat{y}_{t+k} | \Omega_t).$$

a_i's are all non-negative and the a^j's are all non-positive, and because $f(\cdot)$ is a p.d.f Conditions (2) and (3) are sufficient to guarantee strict positivity, by guaranteeing that at least one term is strictly positive, but of course they are not necessary. ■

4b. Series Expansion Approximation of the Optimal Predictor

The method just discussed involves numerical solution of a piecewise-linear approximate loss function. Effectively, we obtain the exact solution to an approximate objective function. Now, in contrast, we shall obtain an approximate solution to the exact objective function.

The approach is of interest for at least two reasons. First, it enables us to dispense with the assumption that the loss function is of "cost-of-error" form, in favor of the general form $L(y_{t+k}, \hat{y}_{t+k})$. Second, and potentially more importantly, exact numerical solution under piecewise-linear loss (or any other loss function) can become very complicated if predictions at more than one horizon are desired, because the conditional c.d.f. will in general change with the horizon. Expressions for the requisite set of c.d.f.'s are typically very difficult to obtain. Thus, for example, even if the one-step-ahead conditional c.d.f. is Gaussian, the multi-step-ahead c.d.f.'s will generally *not* be Gaussian, and they are typically very difficult or impossible to characterize exactly. The method of this section, in contrast, requires computation of only a few low-ordered conditional moments, as opposed to the entire conditional distribution.

First consider the conditionally Gaussian case. Assume that the optimal predictor exists and is unique,

$$\hat{y}_{t+h} = G(\mu_{t+h|t}, \sigma^2_{t+h|t}),$$

where $G(\cdot, \cdot)$ is at least twice continuously differentiable. Then we can take a second order Taylor approximation around the unconditional (and time *in*variant) moments μ_h and σ^2_h,

$$\hat{y}_{t+h} \approx G(\mu_h, \sigma^2_h) + G'(\mu_h, \sigma^2_h)\begin{pmatrix}\mu_{t+h|t}-\mu_h \\ \sigma^2_{t+h|t}-\sigma^2_h\end{pmatrix} + (\mu_{t+h|t}-\mu_h, \sigma^2_{t+h|t}-\sigma^2_h)\, G''(\mu_h, \sigma^2_h)\begin{pmatrix}\mu_{t+h|t}-\mu_h \\ \sigma^2_{t+h|t}-\sigma^2_h\end{pmatrix}.$$

Rewrite this as

$$\hat{y}_{t+h} \approx \beta_0 + \beta_1\mu_{t+h|t} + \beta_2\sigma^2_{t+h|t} + \beta_3(\mu_{t+h|t})^2 + \beta_4(\sigma^2_{t+h|t})^2 + \beta_5(\mu_{t+h|t}\sigma^2_{t+h|t}) \equiv y^*_{t+h}(\beta),$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_5)'$ and $\beta_i = H_i(\mu_h, \sigma^2_h)$, $i = 0, 1, \ldots, 5$. Because the function $G(\cdot, \cdot)$ is generally unknown, so too are the $H(\cdot, \cdot)$ functions. But $\mu_{t+h|t}$ and $\sigma^2_{t+h|t}$ are known, and the minimization that defines $\hat{\beta}$ can be done over a very long simulated realization,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{t=1}^{N} L(y_{t+h}, y^*_{t+h}(\beta)).$$

As $N\to\infty$, $y_{t+h}(\hat{\beta})\to y^*_{t+h}(\beta_0)$, where $y^*_{t+h}(\beta_0)$ is the best predictor within the $y^*_{t+h}(\cdot)$ family, with respect to the metric $L(\cdot, \cdot)$.[10]

A number of remarks are in order.

(1) In the conditionally Gaussian case, the h-step-ahead conditional expectation and the corresponding conditional variance may be computed conveniently using the Kalman filter recursions.

[10] N is the size of the simulated sample.

(2) In the conditionally Gaussian case with "cost of error" loss, $L(e_{t+h})$, one may set β_1 = 1 and $\beta_3 = \beta_5 = 0$ *a priori*, due to Theorem 2.

(3) In the conditionally Gaussian case, it is nevertheless clear that higher-order expansions in $\mu_{t+h|t}$ and $\sigma^2_{t+h|t}$ may be entertained and may lead to improvements. We conjecture that if p denotes the order of the expansion, then as $N \rightarrow \infty$ and $p \rightarrow \infty$ with $p/N \rightarrow 0$, $y^*_{t+h}(\hat{\beta}) \rightarrow \hat{y}_{t+h}$.

(4) In conditionally non-Gaussian cases, expansions in more variables (e.g., involving conditional skewness and/or kurtosis) may be necessary and may be undertaken. We conjecture that if p denotes the order of the expansion and k denotes the number of variables, then as $N \rightarrow \infty$, $p \rightarrow \infty$, and $k \rightarrow \infty$, with $p/N \rightarrow 0$, $k/N \rightarrow 0$, and $k/p \rightarrow 0$, $y^*_{t+h}(\hat{\beta}) \rightarrow \hat{y}_{t+h}$.

(5) For any fixed N, in both the conditionally Gaussian and conditionally non-Gaussian cases, one might be able to obtain more accurate approximations to the optimal predictor than that obtained by Taylor series expansion by mixing Fourier terms with the Taylor terms, or by using methods such as neural networks.[11]

5. A GARCH(1,1) Example

We shall illustrate the results with the conditionally Gaussian GARCH(1,1) process under linlin loss. That is,

$$y_{t+1} = \varepsilon_{t+1}, \qquad \varepsilon_{t+1} | \varepsilon_t \sim N(0, \sigma^2_{t+1|t})$$

where

[11] See Kuan and White (1992).

$$\sigma^2_{t+1|t} = \omega + \alpha\varepsilon^2_t + \beta\sigma^2_{t|t-1}, \qquad \omega,\ \alpha,\ \beta > 0,\ \alpha+\beta < 1,$$

and

$$L(y_{t+h}-\hat{y}_{t+h}) = \begin{cases} a|y_{t+h}-\hat{y}_{t+h}|, & \text{if } y_{t+h}-\hat{y}_{t+h} > 0 \\ b|y_{t+h}-\hat{y}_{t+h}|, & \text{if } y_{t+h}-\hat{y}_{t+h} \leq 0. \end{cases}$$

Throughout, we normalize the unconditional variance to 1 by taking $\omega = (1-\alpha-\beta)$, and we set $\alpha = .2$ and $\beta = .75$.

We set the linlin loss parameters at a = .85, b = .15 (moderate asymmetry) or a = .95, b = .05 (high asymmetry). Throughout, we take

We compare the conditionally expected linlin losses associated with the optimal predictor, the pseudo-optimal predictor, and the conditional mean. This requires conditioning on an initial value of $\sigma^2_{t+1|t}$, and the results will, of course, vary with the value adopted. Here we set the initial conditional variance equal to the unconditional variance plus one standard deviation of the conditional variance,

$$\sigma^2_{t+1|t} = \sigma^2_t + \sqrt{var(\sigma^2_{t+1|t})}\,.$$

Calculation of $var(\sigma^2_{t+1|t})$, the variance of the conditional variance, is straightforward but somewhat tedious. We have

$$Var\left(\sigma^2_{t+1|t}\right) = E\left[\left(\sigma^2_{t+1|t}\right)^2\right] - \left(\sigma^2_t\right)^2.$$

But recall that $E_t\varepsilon^4_{t+1} = 3(\sigma^2_{t+1|t})^2$, so that $(\sigma^2_{t+1|t})^2 = (E_t\varepsilon^4_{t+1})/3$. Thus,

$$var\left(\sigma^2_{t+1|t}\right) = \frac{E\varepsilon^4_{t+1}}{3} - \left(\sigma^2_t\right)^2,$$

by the law of iterated expectations. But, as shown by Bollerslev (1986), the unconditional fourth moment is

$$E\varepsilon_{t+1}^4 = \frac{3\omega^2(1+\alpha+\beta)}{(1-\alpha-\beta)(1-\beta^2-2\alpha\beta-3\alpha^2)} = \frac{3(1-\alpha-\beta)(1+\alpha+\beta)}{1-\beta^2-2\alpha\beta-3\alpha^2},$$

because we set $\omega = (1-\alpha-\beta)$.

We will compare the conditionally expected losses of the optimal predictor, the pseudo-optimal predictor, and the conditional mean predictor, at various horizons and for various parameterizations of linlin loss. To do so, we need an expression for $\sigma_{t+h|t}$, which enters the earlier-derived expression for the optimal linlin predictor. Using results from Baillie and Bollerslev (1992), it is easy to show that for the GARCH(1,1) process,

$$\sigma_{t+h|t} = \sqrt{\sigma_t^2 + (\sigma_{t+1|t}^2-\sigma_t^2)(\alpha+\beta)^{h-1}} \ .$$

It is worth pointing out that the "optimal" predictor we use in this example is known to be truly optimal only for h = 1, because conditional normality holds only for h = 1. But, although the "optimal" predictor used in this example is in fact only an approximation to the optimal predictor when h > 1 (it is in fact an "improved" pseudo-optimal predictor), one expects it to perform better than the "constant adjustment" pseudo-optimal predictor, because it explicitly makes use of the time-varying conditional variance. Recognizing the abuse of language, we shall continue to refer to it as the "optimal predictor". Other approximations, such as the series expansion of section 4, are of course possible and may perform better, but they would introduce unnecessary complexity into the example.[12]

Because of the conditional non-normality when h > 1, we do not rely on the

[12] Because the optimal predictor in this case is the appropriate conditional fractile, one could also follow Baillie and Bollerslev (1992) and take a low-order inverse Edgeworth expansion to approximate the conditional fractile directly.

formulas derived in section 3b to compute the conditionally expected losses of the optimal, pseudo-optimal, and conditional-mean predictors. Instead, we compute them by simulation. At each of 20,000 replications, we draw a GARCH(1,1) realization of length 50, with the conditional variance initialized as discussed above, and we compute the loss of each of the three predictors at each of the 50 horizons. Finally, we average across replications.

In Figure 5, we show a typical GARCH(1,1) realization with $\alpha = .2$, $\beta = .75$, together with the real-time linlin-optimal, pseudo-optimal and conditional mean predictors, for linlin loss parameters a = .95 and b = .05. It is apparent that the optimal predictor injects more bias when conditional volatility is high, reflecting the fact that it accounts for both loss asymmetry and conditional heteroskedasticity. This conditionally optimal amount of bias is sometimes more and sometimes less than the constant bias associated with the pseudo-optimal predictor, which accounts for loss asymmetry but not conditional heteroskedasticity. Finally, of course, the conditional mean injects no bias, as it accounts neither for loss asymmetry nor conditional heteroskedasticity.

In Figure 6, we show the conditionally expected linlin loss of the pseudo-optimal predictor relative to that of the optimal predictor, across forecast horizons. All GARCH and linlin parameters are maintained at the earlier-discussed values of Figure 5. The conditionally expected loss from ignoring the conditional variance dynamics--that is, the conditionally expected loss from using the pseudo-optimal as opposed to the optimal predictor--is as high as 12% for short horizons. As with Granger (1981), although for very different reasons, the optimal predictor is successful at "forecasting white noise." Of course, as the prediction horizon increases, the cost of ignoring the conditional variance dynamics decreases, and the ratio of conditionally expected losses converges to 1.

In Figure 7, we show the conditionally expected linlin loss of the conditional mean relative to that of the optimal predictor. Although the cost of ignoring the conditional variance dynamics still decreases with horizon, the ratio of conditionally expected losses does not approach 1, because the conditional mean predictor ignores loss asymmetry in addition to conditional heteroskedasticity. The failure of the conditional mean to acknowledge the loss asymmetry affects predictive performance at *all* horizons.

Figures 8 and 9 parallel Figures 6 and 7, the difference being that we now set a = .95 and b = .05, so that loss is more highly asymmetric. The results are qualitatively identical, but quantitatively more pronounced.

6. Summary and Concluding Remarks

We have analyzed the optimal prediction problem under asymmetric loss. We computed the optimal predictor analytically in two leading tractable cases and showed how to compute it numerically in less tractable cases. A key theme that emerged was the dependence of the conditionally optimal amount of bias on the conditional variance, thereby providing a direct link from conditional heteroskedasticity to optimal *point* prediction, rather than simply to *interval* prediction. We illustrated this theme with an application to forecasting the GARCH(1,1) process under linlin loss.

Some important recent work in dynamic economic theory is very much linked to the idea of prediction under asymmetric loss discussed here. Building on Jacobson (1977) and Whittle (1990), Hansen, Sargent and Tallarini (1993) set up and motivate a general-equilibrium economy with "risk sensitive" preferences resulting in equilibria with certainty-equivalence properties. Thus, the prediction and decision problems may be done sequentially--but prediction is done with respect to a distorted probability measure that produces predictions that differ from the conditional mean.

As for extensions, we plan to use the predictors developed here to develop recursive prediction-based procedures for selecting forecasting models under the relevant loss function.[13] This of course also involves estimating models under the relevant loss function. The end result will be an integrated tool kit for model selection, estimation, prediction, and forecast evaluation under the relevant loss function. It will also be of interest to examine the ability of the parametric procedures developed here to forecast actual economic time series, and to compare the performance of our parametric predictors to White's (1992) nonparametric predictor,[14] as much of the forecasting literature suggests that simple, tightly parameterized--but nevertheless sophisticated--models tend to perform best in out-of-sample forecasting.[15]

[13] Important recent work along these lines, under a Kullback-Liebler distance metric, is reported in Phillips (1994).

[14] White develops his nonparametric prediction procedure under linlin loss, but it is readily extended to other loss functions.

[15] See Zellner (1992).

References

Baillie, R.T. and T. Bollerslev (1992), "Prediction in Dynamic Models with Time-Dependent Conditional Variances," *Journal of Econometrics*, 52, 91-113.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 3, 307-27.

Bollerslev, T., R.F. Engle and D.B. Nelson (1993), "ARCH Models," in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4. Amsterdam: North-Holland, forthcoming.

Diebold, F.X. and J. Lopez (1994), "ARCH Models," in K. Hoover (ed.), *Macroeconometrics: Developments, Tensions and Prospects*, forthcoming.

Diebold, F.X. and R.S. Mariano (1993), "Comparing Predictive Accuracy," Manuscript, Department of Economics, University of Pennsylvania.

Granger, C.W.J. (1969), "Prediction with a Generalized Cost of Error Function," *Operational Research Quarterly*, 20,199-207.

Granger, C.W.J. (1981), "Forecasting White Noise," in A. Zellner (ed.), *Proceedings of the Conference on Applied Time Series Analysis of Economic Data*. ASA-Census-NBER.

Granger, C.W.J. and Newbold, P. (1986), *Forecasting Economic Time Series* (Second edition). Orlando: Academic Press.

Hansen, B.E. (1994), "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705-730.

Hansen, L.P., T.J. Sargent and T.D. Tallarini (1993), "Pessimism, Neurosis, and Feelings About Risk in General Equilibrium," Manuscript, Department of Economics, University of Chicago.

Jacobson, D.H. (1977), *Extensions of Linear-Quadratic Optimal Control, Optimization and Matrix Theory*. New York: Academic Press.

Kahneman, D. and A. Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263-291.

Kuan, C.-M. and H. White (1992), "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews*,

Makridakis, S. and Wheelwright, S.C. (1987), *The Handbook of Forecasting: A Manager's Guide* (Second Edition). New York: John Wiley.

Phillips, P.C.B. (1994), "Bayes Models and Macroeconomic Activity," Manuscript, Department of Economics, Yale University.

Roy, A.D. (1952), "Safety First and the Holding of Assets," *Econometrica*, 20, 431-449.

Stockman, A.C. (1987), "Economic Theory and Exchange Rate Forecasts," *International Journal of Forecasting*, 3, 3-15.

Tversky, A. and Khaneman, D. (1991), "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, 106, 1039-1061.

Varian, H. (1974), "A Bayesian Approach to Real Estate Assessment," in S.E. Fienberg and A. Zellner (eds.), *Studies in Bayesian Econometrics and Statistics in Honor of L.J. Savage*, 195-208. Amsterdam: North-Holland.

Weiss, A.A. (1991), "Multi-step Estimation and Forecasting in Dynamic Models," *Journal of Econometrics*, 48, 135-49.

Weiss, A.A. (1994), "Estimating Time Series Models Using the Relevant Cost Function," Manuscript, Department of Economics, University of Southern California.

Weiss, A.A. and Andersen, A.P. (1984), "Estimating Forecasting Models Using the Relevant Forecast Evaluation Criterion," *Journal of the Royal Statistical Society A*, 137, 484-487.

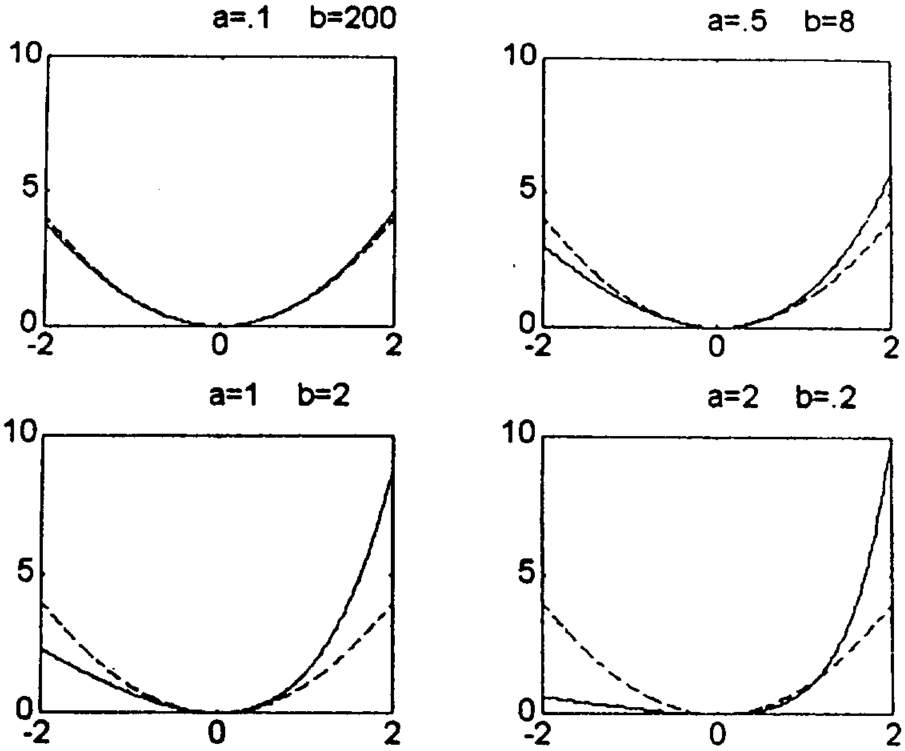
White, H. (1992) "Nonparametric Estimation of Conditional Quantiles Using Neural Networks," in C. Page and R. LePage (eds.), *Computing Science and Statistics, Proceedings of the 22nd Symposium on the Interface, Statistics of Many Parameters: Curves, Images, Spatial Models*. New York: Springer Verlag.

Whittle, P. (1990), *Risk-Sensitive Optimal Control*. New York: John Wiley.

Zellner, A. (1986), "Bayesian Estimation and Prediction Using Asymmetric Loss Functions," *Journal of the American Statistical Association*, 81, 446-451.

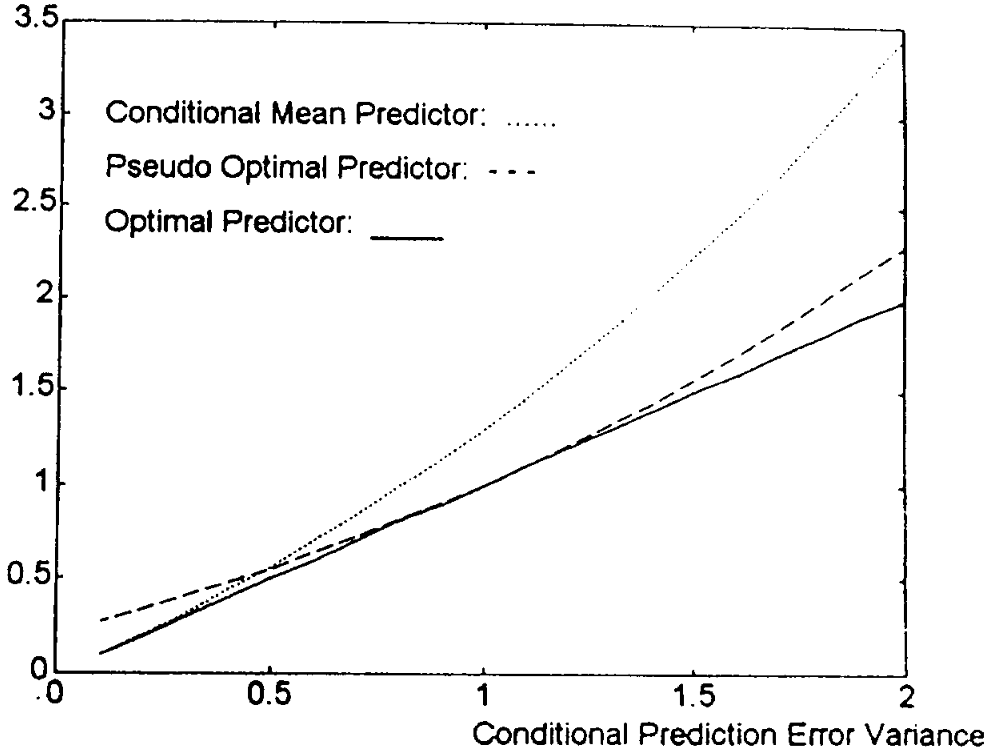
Zellner, A. (1992), "Statistics, Science and Public Policy," *Journal of the American Statistical Association*, 87, 1-6.

Figure 1
Quadratic Loss and Various Linex Loss Functions



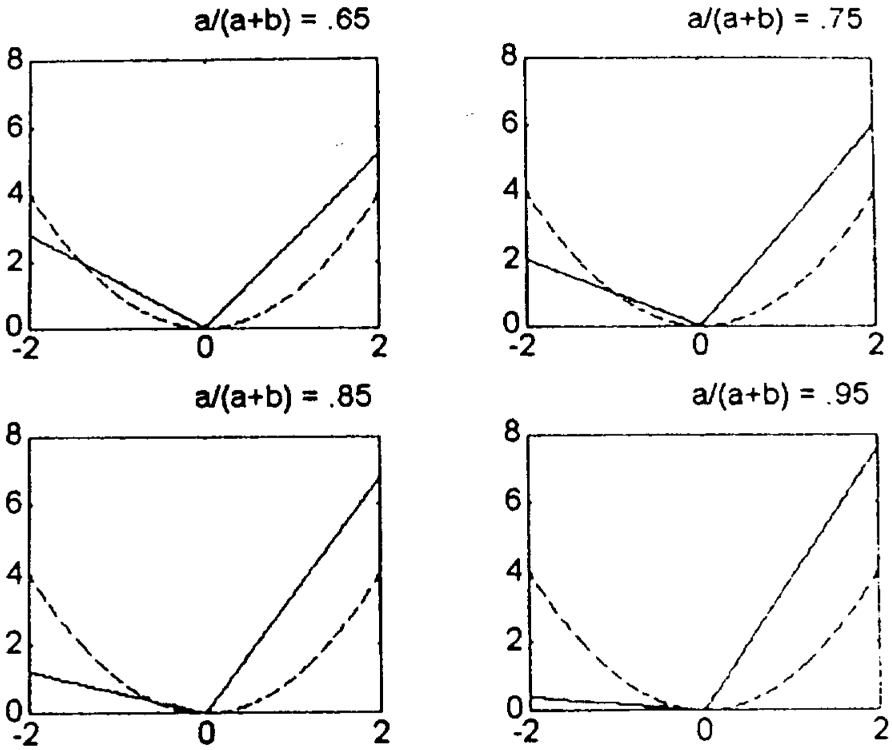
Notes to Figure: Quadratic loss appears as a dashed line and linex loss appears as a solid line. Linex loss is parameterized by a and b, where $L(x) = b[\exp(ax) - ax - 1]$, $a = R\backslash\{0\}$ and $b > 0$.

Figure 2
Conditionally Expected Linex Loss of
Conditional Mean, Pseudo-Optimal, and Optimal Predictors



Notes to Figure: The Linex loss parameters are set to a=1 and b=2. As the current *conditional* variance is changed (i.e., as one moves along the horizontal axis of the graph) the process' *unconditional* variance remains fixed at 1.

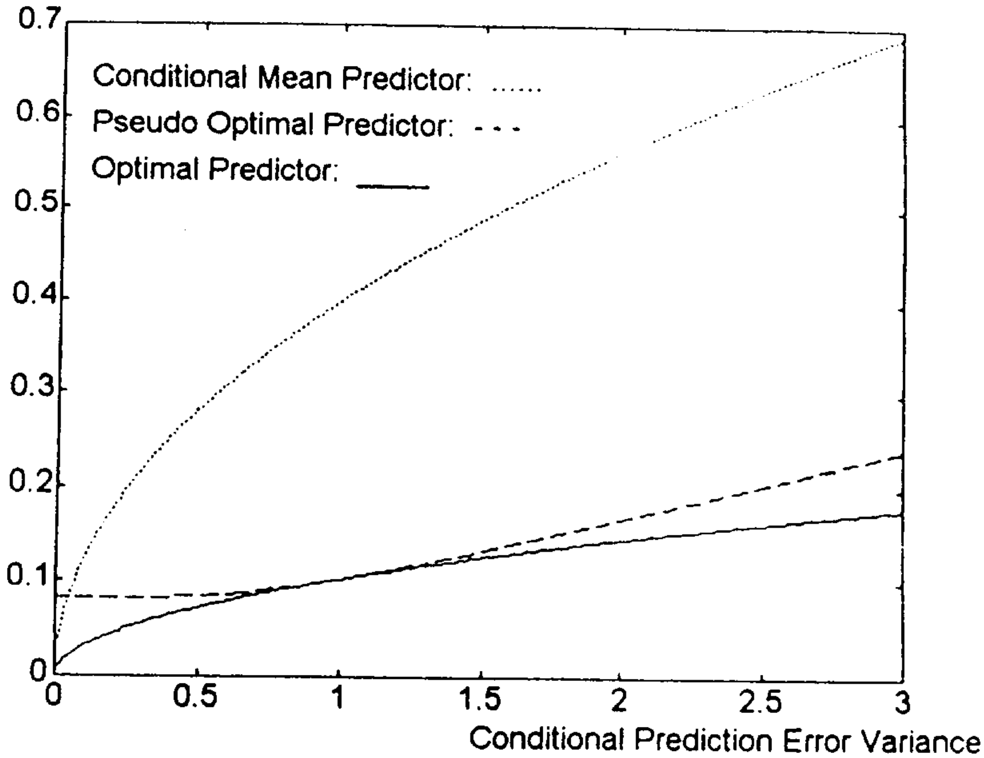
Figure 3
Quadratic Loss and Various Linlin Loss Functions



Notes to Figure: Quadratic loss appears as a dashed line and linlin loss appears as a solid line. Linlin loss is parameterized by a and b, where

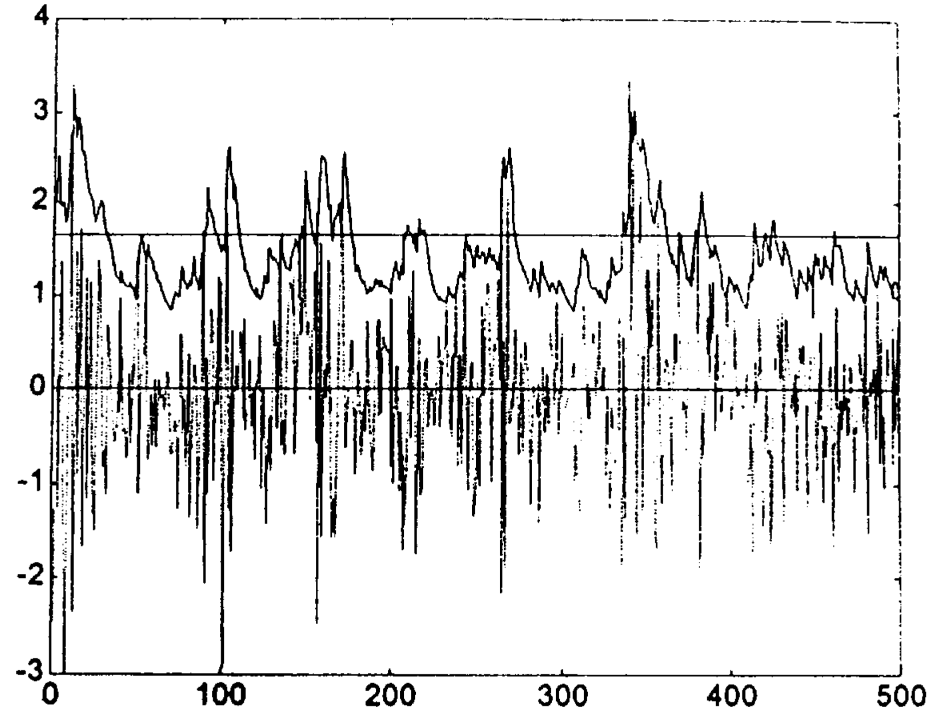
$$L(x) = \begin{cases} a|x|, & \text{if } x > 0 \\ b|x|, & \text{if } x \leq 0. \end{cases}$$

Figure 4
Conditionally Expected Linlin Loss of
Conditional Mean, Pseudo-Optimal, and Optimal Predictors



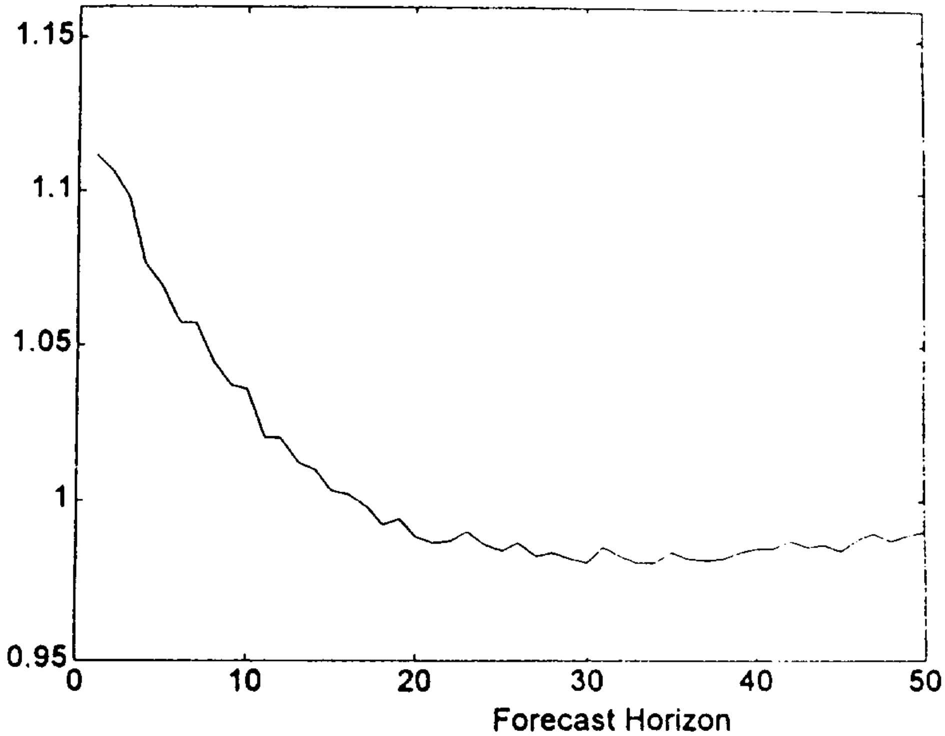
Notes to Figure: The Linlin loss parameters are set to a=.95 and b=.05. As the current *conditional* variance is changed (i.e., as one moves along the horizontal axis of the graph) the process' *unconditional* variance remains fixed at 1.

Figure 5
GARCH(1,1) Realization with
Linlin Optimal, Pseudo-Optimal, and Conditional Mean Predictors



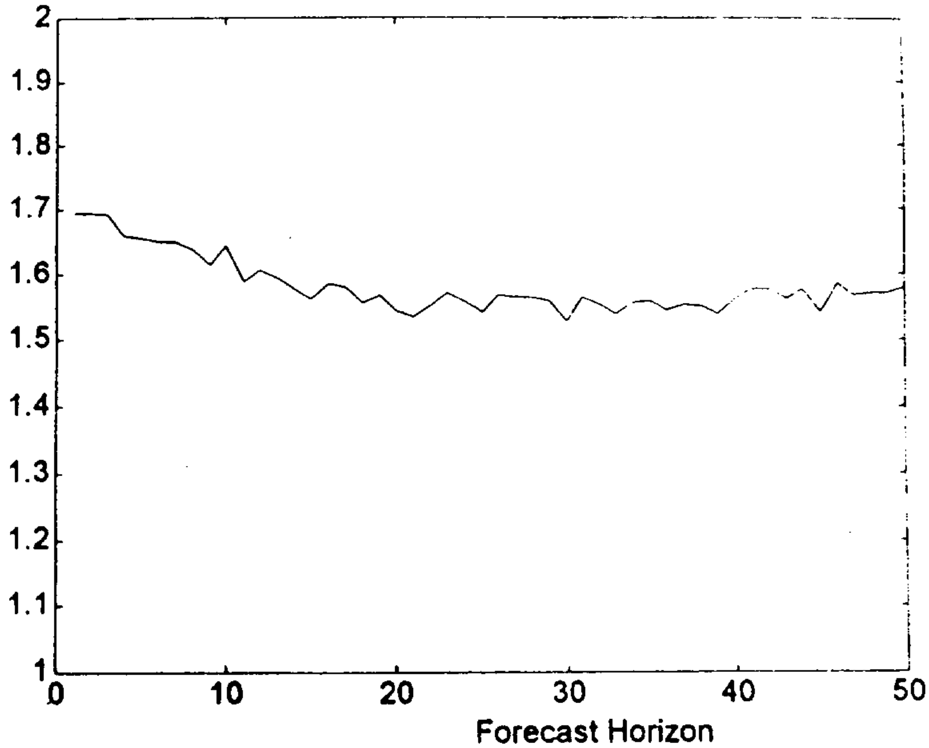
Notes to Figure: The linlin loss parameters are set to a = .95 and b = .05, so that a/(a+b) = .95. The GARCH(1,1) parameters are set to $\alpha = .2$ and $\beta = .75$. The dotted line is the GARCH(1,1) realization. The horizontal line at zero is the conditional mean predictor, the horizontal line at 1.65 is the pseudo-optimal predictor, and the time-varying solid line is the optimal predictor.

Figure 6
Ratio of Conditionally Expected Linlin Loss
of Pseudo-Optimal and Optimal Predictors



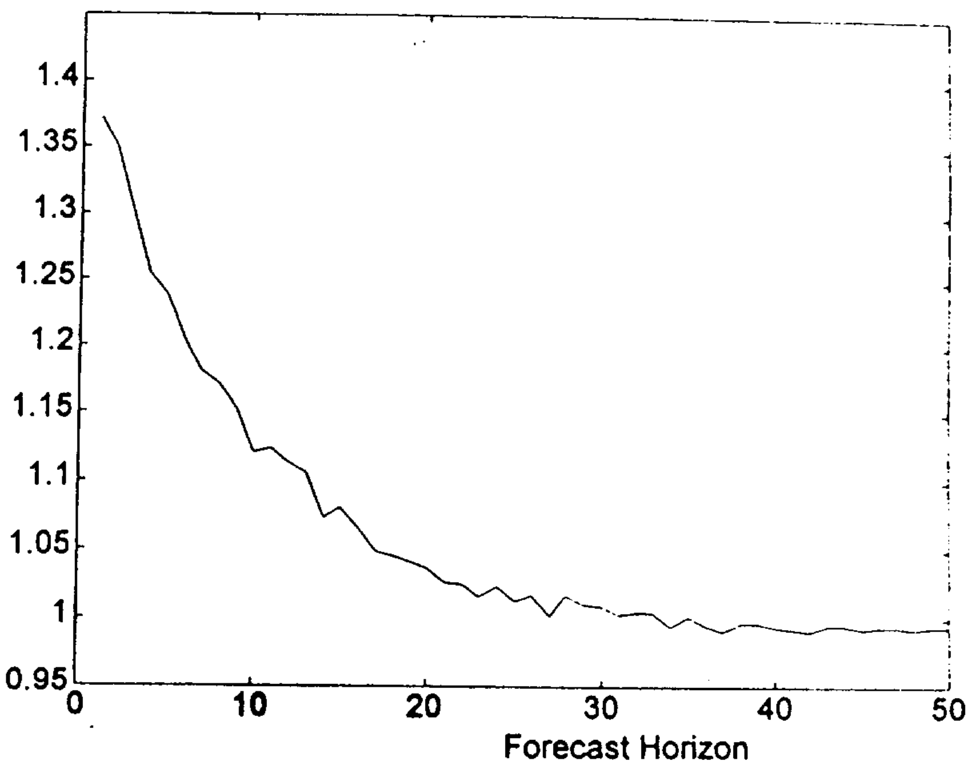
Notes to Figure: The linlin loss parameters are set to a = .85 and b = .15, so that a/(a+b) = .85. The GARCH(1,1) parameters are set to α = .2 and β = .75.

Figure 7
Ratio of Conditionally Expected Linlin Loss of Conditional Mean and Optimal Predictors



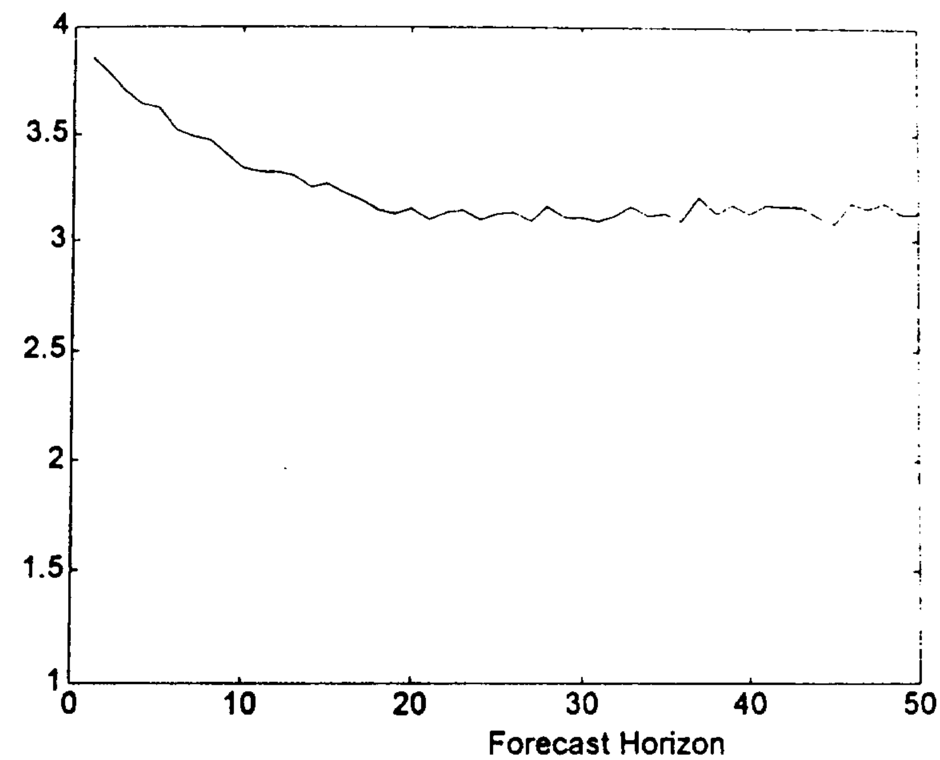
Notes to Figure: The linlin loss parameters are set to a = .85 and b = .15, so that a/(a+b) = .85. The GARCH(1,1) parameters are set to α = .2 and β = .75.

Figure 8
Ratio of Conditionally Expected Linlin Loss of Pseudo-Optimal and Optimal Predictors



Notes to Figure: The linlin loss parameters are set to a = .95 and b = .05, so that a/(a+b) = .95. The GARCH(1,1) parameters are set to α = .2 and β = .75.

Figure 9
Ratio of Conditionally Expected Linlin Loss of Conditional Mean and Optimal Predictors



Notes to Figure: The linlin loss parameters are set to a = .95 and b = .05, so that a/(a+b) = .95. The GARCH(1,1) parameters are set to α = .2 and β = .75.