Christoffersen, P. and Diebold, F.X. (1997),
"Optimal Prediction Under Asymmetric Loss,"
Econometric Theory, 13, 808-817.

OPTIMAL PREDICTION UNDER ASYMMETRIC LOSS[1]

By Peter F. Christoffersen and Francis X. Diebold[2]

Keywords: Forecasting, loss function, asymmetric loss, nonlinear, heteroskedasticity

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. The potential necessity of allowing for asymmetric loss has long been acknowledged. 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). Practitioners routinely echo this sentiment (e.g., Stockman, 1987).

In this paper we treat the prediction problem under general loss structures, building on the classic work of Granger (1969). In Section 2, we characterize the optimal predictor for non-Gaussian processes under asymmetric loss. The results apply, for example, to important classes of conditionally heteroskedastic processes. In Section 3, we provide analytic solutions for the optimal predictor under two popular analytically-tractable asymmetric loss functions. In Section 4, we provide methods for approximating the optimal predictor under more general loss functions. We conclude in Section 5.

2. OPTIMAL PREDICTION FOR NON-GAUSSIAN PROCESSES

Granger (1969) studies Gaussian processes and shows that under asymmetric loss the optimal predictor is the conditional mean plus a constant bias term. Granger's fundamental result, however, has two key limitations. First, the Gaussian assumption implies a constant conditional prediction-error variance. This is unfortunate because conditional heteroskedasticity is widespread in economic and financial data. Second, the loss function must be of prediction-error form; that is, $L(y_{t+h},\hat{y}_{t+h}) = L(y_{t+h}-\hat{y}_{t+h}) = L(e_{t+h})$, where y_{t+h} is the h-step-ahead realization, $\hat{y}_{t+h}$ is the h-step-ahead forecast (made at time t), and e_{t+h} is the corresponding forecast error. More general functions of realizations and predictions are excluded.

Let us begin, then, by generalizing Granger's result to allow for conditional variance dynamics. We achieve this most simply by working in a conditionally-Gaussian, but not necessarily unconditionally-Gaussian, environment, with prediction-error loss. Subsequently we shall allow for both conditional non-normality and more general loss functions.

PROPOSITION 1: *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* $\hat{y}_{t+h} = \mu_{t+h|t}+\alpha_{t+h|t}$*, where* $\alpha_{t+h|t}$ *depends only on the loss function and the conditional prediction-error variance* $\sigma^2_{t+h|t} = var(y_{t+h}|\Omega_t) = var(e_{t+h}|\Omega_t)$.

PROOF: *See Appendix.*

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. Conditionally Gaussian GARCH processes, for example, fall under the jurisdiction of Proposition 1. 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.

The result of Proposition 1 that the "adjustment factor" depends only on the conditional variance depends crucially on conditional normality. We can dispense with conditional normality and still obtain a sharp result, however, which is a straightforward extension of Proposition 1.

PROPOSITION 2: *If* $y_{t+h}|\Omega_t$ *has conditional mean* $\mu_{t+h|t}$, *and a vector of (possibly time varying) conditional moments of order two and higher* $\lambda_{t+h|t}$, *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* $\hat{y}_{t+h} = \mu_{t+h|t}+\alpha_{t+h|t}$, *where* $\alpha_{t+h|t}$ *depends only on the loss function and* $\lambda_{t+h|t}$.

PROOF: *See Appendix.*

Note, however, that although Propostion 2 does not require a Gaussian process, it does require prediction-error loss. In Section 4 we will relax that assumption as well.

3. ANALYTIC SOLUTIONS UNDER LINEX AND LINLIN LOSS

Here we examine two asymmetric loss functions ("linex" and "linlin") for which it is possible to solve analytically for the optimal predictor. To maintain continuity of exposition, we work throughout this section with the conditionally Gaussian process $y_{t+h}|\Omega_t \sim N(\mu_{t+h|t},\sigma^2_{t+h|t})$.[3] For each loss function, we characterize the optimal predictor, $\hat{y}_{t+h} = \mu_{t+h|t}+\alpha_{t+h|t}$, and we compare its conditionally expected loss to that of two competitors, the conditional mean $\mu_{t+h|t}$, and the pseudo-optimal predictor $\hat{y}_{t+h} = \mu_{t+h|t}+\alpha_h$, where α_h depends only on the loss function and the unconditional prediction-error variance $\sigma^2_h = var(e_{t+h})$. 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 that it incorporates only a *constant* adjustment for asymmetry; thus, it is fully optimal only in the conditionally homoskedastic case $\sigma^2_{t+h|t} = \sigma^2_h,\ \forall\ t,\ h$.

3.1. Linex Loss

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

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

It is so-named because when $a>0$, loss is approximately linear to the left of the origin and approximately exponential to the right, and conversely when $a<0$. The optimal h-step-ahead predictor under linex loss solves

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

Differentiating and using the conditional moment-generating function for a conditionally Gaussian variate, we obtain $\hat{y}_{t+h} = \mu_{t+h|t}+\frac{a}{2}\,\sigma^2_{t+h|t}$. Similar calculations reveal that the pseudo-optimal predictor is $\hat{y}_{t+h} = \mu_{t+h|t}+\frac{a}{2}\sigma^2_h$, where $\sigma^2_h = var(e_{t+h})$ is the unconditional h-step-ahead prediction-error variance.

Proposition 1 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.[4] The reason is simple--when a is positive, for example, 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 it grows, so too does the optimal amount of bias, in order to avoid large positive prediction errors. Effectively, optimal prediction under asymmetric loss corresponds to conditional-mean prediction of a transformed series, where the transformation reflects both the loss function and the higher-order conditional moments of the original series.

For example, the optimal predictor of y_{t+h} under conditional normality and linex loss, $\hat{y}_{t+h} = \mu_{t+h|t} + \frac{a}{2}\sigma^2_{t+h|t}$, is the conditional mean of x_{t+h}, where $x_{t+h} = y_{t+h} + \frac{a}{2}\sigma^2_{t+h|t}$.[5]

Inserting the optimal, pseudo-optimal, and conditional mean predictors into the conditionally expected loss expression, we see that the conditionally-expected linex losses are $ba^2\sigma^2_{t+h|t}/2$, $b[\exp(a^2(\sigma^2_{t+h|t} - \sigma^2_h)/2) + a^2\sigma^2_h/2 - 1]$, and $b[\exp(a^2\sigma^2_{t+h|t}/2) - 1]$, respectively. By construction, the conditionally expected loss of the optimal predictor is less than or equal to that of any other predictor. Interestingly, however, it is not possible to rank the pseudo-optimal as superior to the conditional mean predictor. Tedious but straightforward algebra reveals that, for sufficiently small values of $\sigma^2_{t+h|t}$ (depending non-linearly on the values of a and σ^2_h), the conditionally expected loss of the conditional mean will be smaller than that of the pseudo-optimal predictor. In very low volatility times, the conditionally optimal amount of bias is very small, resulting in a lower conditionally expected loss for the conditional mean than for the pseudo-optimal predictor, the bias of which is optimal in "average" times, but too low in low-volatility times.

The situation is illustrated in Figure 1, in which we plot conditionally expected linex loss as a function of $\sigma^2_{t+h|t}$ for each of the three predictors. The conditionally expected loss of the optimal predictor is linear in $\sigma^2_{t+h|t}$ and is of course always lowest. The losses of the pseudo-optimal and the optimal predictors coincide when $\sigma^2_{t+h|t} = \sigma^2_h = 1$. As $\sigma^2_{t+h|t}$ falls below σ^2_h, the loss of the conditional mean intersects the loss of the pseudo-optimal predictor from above. As $\sigma^2_{t+h|t}$ gets close to zero, the optimal predictor incorporates progressively smaller corrections to the conditional mean, so the conditionally expected losses of the optimal and conditional mean predictors coincide.

3.2. Linlin Loss

The "linlin" loss function,

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

so-called because of its linearity on each side of the origin, was used by Granger (1969) and is the loss function underlying quantile regression. The optimal predictor solves

$$\min_{\hat{y}_{t+h}} \left\{ a \int_{\hat{y}_{t+h}}^{\infty} (y_{t+h}-\hat{y}_{t+h}) f(y_{t+h}|\Omega_t) dy_{t+h} - b \int_{-\infty}^{\hat{y}_{t+h}} (y_{t+h}-\hat{y}_{t+h}) f(y_{t+h}|\Omega_t) dy_{t+h} \right\}.$$

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 $F(\hat{y}_{t+h}|\Omega_t) = \frac{a}{a+b}$, where $F(y_{t+h}|\Omega_t)$ is the conditional c.d.f. of y_{t+h} and $f(y_{t+h}|\Omega_t)$ is the conditional density of y_{t+h}.

In the conditionally Gaussian case we have from Proposition 1 that

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

where $\Phi(z)$ is the *N(0,1)* c.d.f. It follows that the conditionally optimal amount of bias is $\alpha_{t+h|t} = \sigma_{t+h|t}\ \Phi^{-1}\left(\frac{a}{a+b}\right)$, so that $\hat{y}_{t+h} = \mu_{t+h|t} + \sigma_{t+h|t}\ \Phi^{-1}\left(\frac{a}{a+b}\right)$.[6] Similar calculations reveal that the pseudo-optimal predictor is $\hat{y}_{t+h} = \mu_{t+h|t} + \sigma_h\ \Phi^{-1}\left(\frac{a}{a+b}\right)$.

Now let us compute conditionally expected linlin loss for the optimal, pseudo-optimal and conditional mean predictors. Recall the formulae for the truncated expectation,

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

and substitute them into the expected loss expression to obtain

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

But under conditional normality,

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

where $\xi_{t+h|t} = \dfrac{\hat{y}_{t+h}-\mu_{t+h|t}}{\sigma_{t+h|t}}$ and $\phi(\cdot)$ is the *N(0,1)* p.d.f. Substituting into the conditionally expected loss expression, we obtain (after some algebraic manipulation)

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

For the optimal predictor, $\xi_{t+h|t} = \Phi^{-1}\left(\dfrac{a}{a+b}\right)$, yielding an expected loss of $(a+b)\sigma_{t+h|t}\phi\left(\Phi^{-1}\left(\dfrac{a}{a+b}\right)\right)$. For the pseudo-optimal predictor, $\xi_{t+h|t} = \Phi^{-1}\left(\dfrac{a}{a+b}\right)\dfrac{\sigma_h}{\sigma_{t+h|t}}$, yielding an expected loss of

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

For the conditional mean predictor, $\xi_{t+h|t} = 0$, yielding an expected loss of $E_t(L(y_{t+h}-\hat{y}_{t+h})) = (a+b)\sigma_{t+h|t}/\sqrt{2\pi}$. Qualitatively, the situation is identical to that shown in Figure 1 for the linex case.

4. APPROXIMATING THE OPTIMAL PREDICTOR

The analytic results above rely on simple loss functions. In general, however, it is not possible to solve analytically for the optimal predictor. Here we develop an approximately optimal predictor via series expansions. The approach is of interest because it frees us from two potentially restrictive assumptions -- conditional normality and prediction-error loss.

For the moment maintain the conditional normality assumption, and 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 series expansion around the unconditional (and time invariant) moments μ_h and σ_h^2,

$$\hat{y}_{t+h} \approx G(\mu_h,\sigma_h^2) + G^{\prime}(\mu_h,\sigma_h^2)\begin{pmatrix}\mu_{t+h|t}-\mu_h \\ \sigma_{t+h|t}^2-\sigma_h^2\end{pmatrix} + \frac{1}{2}\,(\mu_{t+h|t}-\mu_h,\ \sigma_{t+h|t}^2-\sigma_h^2)\ G^{\prime\prime}(\mu_h,\sigma_h^2)\begin{pmatrix}\mu_{t+h|t}-\mu_h \\ \sigma_{t+h|t}^2-\sigma_h^2\end{pmatrix}.$$

Rewrite this as

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

where $\beta = (\beta_0, \beta_1, ..., \beta_5)^{\prime}$ and $\beta_i = H_i(\mu_h, \sigma_h^2)$, $i = 0, 1, ..., 5$. Because the function $G(\cdot,\cdot)$ is generally unknown, so too are the $H(\cdot,\cdot)$ functions. But $\mu_{t+h|t}$ and $\sigma_{t+h|t}^2$ are known, and the minimization that defines $\hat{\beta}_N$ can be done over a very long simulated realization of length N, $\hat{\beta}_N = \underset{\beta\in B}{argmin} \sum_{t=1}^{N} L(y_{t+h}, y_{t+h}^{*}(\beta))$. Under regularity conditions given in the Appendix, the following proposition is immediate.

PROPOSITION 3: *As* $N\rightarrow\infty$, $y_{t+h}^{*}(\hat{\beta}_N)\rightarrow y_{t+h}^{*}(\beta_0)$, *where* $y_{t+h}^{*}(\beta_0)$ *is the best predictor within the* $y_{t+h}^{*}(\cdot)$ *family, with respect to the metric* $L(\cdot,\cdot)$.

PROOF: *See Appendix.*

A number of remarks are in order. First, the h-step-ahead conditional expectation and the corresponding conditional variance may be computed conveniently using the Kalman filter recursions. Second, if loss is in fact of prediction-error form, $L(e_{t+h})$, one may set $\beta_1 = 1$ and $\beta_3 = \beta_5 = 0$ *a priori*, due to Proposition 1. Third, it is clear that higher-order expansions in $\mu_{t+h|t}$ and $\sigma_{t+h|t}^2$ may be entertained and may lead to improvements. Fourth, conditional non-normality may be handled with expansions involving more than the first two conditional moments (e.g., involving conditional skewness and kurtosis). Fifth, and related, parametric economy can be achieved in conditionally non-Gaussian cases using the autoregressive conditional density framework of Hansen (1994). Hansen's framework exploits parametric conditional mean and variance functions but allows for higher-order conditional dynamics by letting the normalized variable $z_{t+h}(\theta) = \left(y_{t+h} - \mu_{t+h|t}(\theta)\right)/\sigma_{t+h|t}(\theta)$ follow a distribution with possibly time varying "shape" parameters, such as a t-distribution with time-varying degrees of freedom (and variance standardized to 1). Sixth, in both the conditionally Gaussian and conditionally non-Gaussian cases, one is of course not limited to series expansions; other nonparametric functional estimators may be used.

5. SUMMARY AND CONCLUDING REMARKS

This paper is part of a research program aimed at allowing for general loss structures in estimation, model selection, prediction, and forecast evaluation. Recently a number of authors have made progress toward that goal, including Weiss (1994) on estimation, Phillips (1994) on model selection, and Diebold and Mariano (1995) on forecast evaluation. Here we focused on prediction and 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 is that the conditionally optimal forecast is biased, and that the conditionally optimal amount of bias is time-varying in general and depends on higher-order conditional moments. Thus, even for models with linear conditional-mean structure, the optimal predictor is in general nonlinear, thereby providing a link with the broader nonlinear time series literature.

Interestingly, some important recent work in dynamic economic theory is very much linked to the idea of prediction under asymmetric loss discussed here. Building on 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 yields predictions different from the conditional mean.

University of Pennsylvania

APPENDIX

PROOF OF PROPOSITION 1: We seek the predictor that solves

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

(Here and throughout, $E_t(x)$ denotes $E(x|\Omega_t)$.) Without loss of generality we can write $\hat{y}_{t+h} = \mu_{t+h|t} + \alpha_{t+h|t}$ and $y_{t+h} = \mu_{t+h|t} + x_{t+h}$, so that

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

Because $f(x_{t+h}|\Omega_t)$ depends on $\sigma^2_{t+h|t}$ but not $\mu_{t+h|t}$, so too does the $\alpha_{t+h|t}$ that solves the minimization problem depend on $\sigma^2_{t+h|t}$ but not $\mu_{t+h|t}$. *Q.E.D.*

PROOF OF PROPOSITION 2: Precisely parallels that of Proposition 1. *Q.E.D.*

PROOF OF PROPOSITION 3: Following Amemiya (1985), we require three conditions:

(1) $\beta_0 \in B$, a compact subset of $\mathbb{R}^k$.

(2) $L_N(\beta) = \sum_{t=1}^{N} L(y_{t+h}, y^*_{t+h}(\beta))$ is continuous in $\beta \in B$ for all $y=(y_{1+h},...,y_{N+h})$ and is a

measurable function of y for all $\beta \in B$.

(3) $N^{-1}L_N(\beta)$ converges to a nonstochastic continuous function $L(\beta)$ in probability uniformly in $\beta \in B$ as $N \to \infty$, and $L(\beta)$ attains a unique global minimum at β_0.

Under the conditions, $\hat{\beta}_N = \underset{\beta \in B}{argmin} L_N(\beta)$ converges in probability to β_0 by the argument of Amemiya (1985, p. 107). Thus, $y^*_{t+h}(\hat{\beta}_N)$ converges in probability to $y^*_{t+h}(\beta_0)$ by continuity of $y^*_{t+h}(\hat{\beta}_N)$. *Q.E.D.*

REFERENCES

Amemiya, T.: *Advanced Econometrics*. Cambridge, Mass.: Harvard University Press, 1985.

Christoffersen, P.F. and Diebold, F.X. (1994): "Optimal Prediction Under Asymmetric Loss," National Bureau of Economic Research Technical Working Paper No. 167, Cambridge, Mass.

Diebold, F.X. and R.S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-265.

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

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

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, University of Chicago.

Phillips, P.C.B. (1994): "Bayes Models and Macroeconomic Activity," Manuscript, Yale

University.

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

Varian, H.: "A Bayesian Approach to Real Estate Assessment," in *Studies in Bayesian Econometrics and Statistics in Honor of L.J. Savage*, ed. by S.E. Feinberg and A. Zellner. Amsterdam: North-Holland, 1974.

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

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

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

FOOTNOTES

1. This paper is a heavily-revised and shortened version of parts of Christoffersen and Diebold (1994), which may be consulted for additional results, discussion, and examples.

2. We benefitted from constructive comments from the Co-Editor and two referees, as well as from Clive Granger, Hashem Pesaran, Enrique Sentana, Bob Stine, Jim Stock, Ken Wallis, and numerous conference and seminar participants. Remaining inadequacies are ours alone. We thank the National Science Foundation, the Sloan Foundation, the University of Pennsylvania Research Foundation, and the Barbara and Edward Netter Fellowship for support.

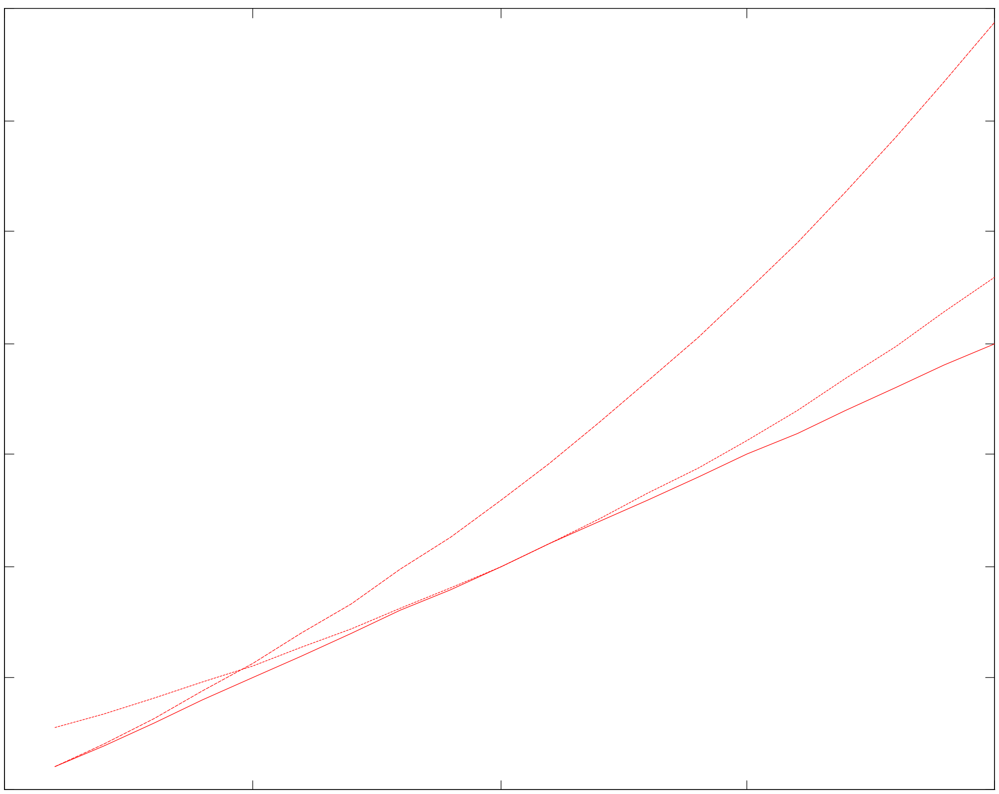
3. As will be made clear, however, although conditional normality is crucial to our derivation of the optimal predictor under linex loss, it may readily be discarded under linlin loss.

4. Note that as $a \to 0$ the conditionally optimal amount of bias approaches zero. Quadratic loss obtains as $a \to 0$, because if a is small one can replace the exponential part of the loss function by the first two terms of its Taylor series expansion, yielding the approximation $L(x) \propto x^2$.

5. Because y_{t+h} is conditionally normal with $E[y_{t+h}|\Omega_t] = \mu_{t+h|t}$, x_{t+h} is conditionally normal with $E[x_{t+h}|\Omega_t] = \mu_{t+h|t} + \frac{a}{2}\sigma^2_{t+h|t} = \hat{y}_{t+h}$.

6. Note that with linlin loss (in contrast to linex loss) it is very easy, even for non-Gaussian 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, $\hat{y}_{t+h} = F^{-1}\left(\frac{a}{a+b}|\Omega_t\right)$, so $\hat{y}_{t+h}$ is simply the *(a/(a+b))th* conditional quantile. When $a=b$, of course, $\hat{y}_{t+h}$ is the conditional median.

Figure 1
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*. The unconditional variance is fixed at 1.