DISCUSSION PAPER SERIES

No. 4037

PROPERTIES OF OPTIMAL FORECASTS

Andrew J Patton and Allan G Timmermann

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

ISSN 0265-8003

PROPERTIES OF OPTIMAL FORECASTS

Andrew J Patton, London School of Economics (LSE)
Allan G Timmermann, University of California, San Diego and CEPR

Discussion Paper No. 4037
August 2003

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 4037

August 2003

ABSTRACT

Properties of Optimal Forecasts*

Evaluation of forecast optimality in economics and finance has almost exclusively been conducted under the assumption of mean squared error loss. Under this loss function optimal forecasts should be unbiased and forecast errors serially uncorrelated at the single period horizon with increasing variance as the forecast horizon grows. Using analytical results we show in this Paper that all the standard properties of optimal forecasts can be invalid under asymmetric loss and non-linear data-generating processes and thus may be very misleading as a benchmark for an optimal forecast. Our theoretical results suggest that many of the conclusions in the empirical literature concerning sub-optimality of forecasts could be premature. We extend the properties that an optimal forecast should have to a more general setting than previously considered in the literature. We also present new results on forecast error properties that may be tested when the forecaster's loss function is unknown but restrictions can be imposed on the data-generating process, and introduce a change of measure, following which the optimum forecast errors for general loss functions have the same properties as optimum errors under MSE loss.

JEL Classification: C22, C52 and C53
Keywords: efficient markets, forecast evaluation, loss function and rationality

Andrew J Patton
London School of Economics
Houghton Street
London
WC2A 2AE
Email: a.patton@lse.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=159418

Allan G Timmermann
Department of Economics
University of California, San Diego
9500 Gilman Drive
La Jolla CA 92093-0508
USA
Tel: (1 858) 534 4860
Fax: (1 858) 534 7040
Email: atimmerm@ucsd.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=116464

*The authors would like to thank Jeremy Berkowitz, Sean Campbell, Carlos Capistran, Peter Christoffersen, Frank Diebold, Graham Elliott, Raffaella Giacomini, Clive Granger, Peter Hansen, Oliver Linton, Mark Machina, Francisco Penaranda, Kevin Sheppard, Hal White and seminar participants at Brown, CORE, LSE, McGill, UC Riverside, Stanford, Warwick Business School and the 2002 EC^2 conference for useful comments. All remaining deficiencies are the responsibility of the authors. Patton gratefully acknowledges financial support from the IAM Programme in Hedge Fund Research at LSE.

Submitted 29 July 2003

1 Introduction

Knowledge of the properties possessed by an optimal forecast is crucial in many areas of economics and finance and is used, inter alia, in tests of the efficient market hypothesis in foreign exchange, bond and stock markets, and tests of the rationality of decision makers in a variety of macroeconomic applications. Almost without exception empirical work has relied on testing properties that optimal forecasts have under mean squared error (MSE) loss.[1] These properties include unbiasedness of the forecast, lack of serial correlation in one-step-ahead forecast errors, serial correlation of order $h-1$ at the h-period horizon and non-decreasing forecast error variance as the forecast horizon grows. Although such properties seem sensible, they are in fact established under a set of very restrictive assumptions on the decision maker's loss function.

Increasingly the assumption of symmetric loss has been questioned in the literature. Christoffersen and Diebold (1997), Diebold (2001), Granger and Newbold (1986), Granger and Pesaran (2000), Pesaran and Skouras (2001), Skouras (2001) and West, Edison and Cho (1993) call for a more decision theoretic approach to forecasting that considers the losses derived from over- and underpredictions. There is often no reason why losses should be symmetric around a zero forecast error (the perfect prediction). For instance, financial analysts' forecasts have been found to be strongly biased[2] and it is easy to understand why, as underprediction of corporate earnings is likely to lead to a worsened relationship between the analyst and the firm in question, while overpredictions are unlikely to generate similar pressures.

In this paper we demonstrate that none of the properties traditionally associated with tests of optimal forecasts carry over to a more general setting with asymmetric loss and possible nonlinear dynamics in the data generating process. While bias of the optimal forecast has been established by Granger (1969, 1999) and characterized analytically for certain classes of loss functions and forecast error distributions by Christoffersen and Diebold (1997), failure of the remaining optimality

[1] See, e.g., Brown and Maital (1981), Cargill and Meyer (1980), De Bondt and Bange (1992), Dokko and Edelstein (1989), Figlewski and Wachtel (1981), Keane and Runkle (1990, 1998), Lakonishok (1980), Mishkin (1981), Muth (1961), Pesando (1975) and Schroeter and Smith (1986) and Zarnowitz (1985).

[2] See De Bondt and Thaler (1990), Abarbanell and Bernard (1992) and Michaely and Womack (1999) for example.

properties has not previously been shown.[3,4]

As a precursor to our general results, we first derive closed-form results in the context of a commonly used asymmetric loss function (linear-exponential, or "linex") and a widely used nonlinear data generating process, namely the regime switching model suggested by Hamilton (1989). We find that not only can the optimal forecast be biased, but the forecast errors can be serially correlated of arbitrarily high order and both the unconditional and conditional forecast error variance may be *decreasing* functions of the forecast horizon.

We next extend the properties that an optimal forecast should have to a more general setting than that previously considered in the literature. Our results suggest that conclusions in the empirical literature concerning suboptimality of forecasts may have been premature. We prove that the *expected loss*, rather than the forecast error variance, is a non-decreasing function of the forecast horizon and that a "generalized forecast error" has mean zero, limited serial correlation and is a martingale difference sequence at the single-period horizon.

We also introduce a transformation from the usual probability measure to an "MSE-loss probability measure", under which the optimal forecasts are unbiased and forecast errors are serially uncorrelated, in spite of the fact that these properties generally fail to hold under the physical measure. These results are analogous to the change of measure from the physical measure to the risk-neutral measure, under which assets may be priced as though investors are risk-neutral.

Finally, we establish some surprising new results that trade off restrictions on the loss function against restrictions on the data generating process. In situations where the conditional higher order moments of the forecast variable are constant, we show that although the optimal forecast may well be biased, the one-step optimal forecast errors are not serially correlated while the h-step forecast errors are at most MA(h-1). This holds irrespective of the shape of the loss function. This offers

[3] Under asymmetric loss functions such as lin-lin and linex and assuming a conditionally Gaussian process, Christoffersen and Diebold (1997) establish that the optimal forecast is biased and characterize the optimal bias analytically. Their study does not, however, consider the other properties of optimal forecast errors such as lack of serial correlation and non-decreasing variance.

[4] Hoque, *et al.* (1988), and Magnus and Pesaran (1987 and 1989) discuss violations of the standard properties of optimal forecasts caused by estimation error, rather than by a choice of loss function different from MSE. In this paper we consider the case of zero estimation error, to rule this out as a cause of apparent violations. Violations caused by forecasters behaving strategically are discussed in Ehrbeck and Waldmann (1996), Laster, *et al.* (1999), Prendergast and Stole (1996) and Scharfstein and Stein (1990), among others.

a new way to test optimality of forecast errors that is robust to the loss function, but requires restrictions on the underlying data generating process. This result will be useful in the common situation where the shape of the loss function is unknown, whereas the restrictions on the data generating process can be tested empirically. We also establish conditions on the data generating process such that the optimal forecast error variance is non-decreasing in the forecast horizon.

The outline of the paper is as follows. Section 2 first summarizes the properties of optimal linear predictions under squared error loss and next demonstrates how each of these properties can be violated under asymmetric loss in the context of a nonlinear data generating process. Section 3 establishes properties of optimal forecasts under general loss and verifies that these are satisfied for the model considered in Section 2. This section also contains the change of measure results. Section 4 derives testable properties of the forecast errors when restrictions are imposed on the data generating process while Section 5 concludes. An appendix contains the list of assumptions employed in various places, technical details and proofs.

2 Standard Optimality Properties and their Violation under Asymmetric Loss

Suppose that a decision maker is interested in forecasting some univariate time series, $Y = \{Y_t; t = 1, 2, ...\}$, h steps ahead given information at time t, $\mathcal{F}_t$. We assume that $Y \equiv \{Y_t : \Omega \to \mathbb{R},\ t = 1, 2, ...\}$ is a stochastic process on a complete probability space $(\Omega, \mathcal{F}, P)$, where $\Omega = \mathbb{R}^{m\infty} \equiv \times_{t=1}^{\infty}\mathbb{R}^m$, and $\mathcal{F} = \mathcal{B}^{m\infty} \equiv \mathcal{B}(\mathbb{R}^{m\infty})$, the Borel σ-field generated by $\mathbb{R}^{m\infty}$. Y_t is thus adapted to the information set available at time t, denoted $\mathcal{F}_t$. At a minimum $\mathcal{F}_t$ includes the filtration generated by $\{Y_{t-k}; k \geq 0\}$, but it may also be expanded to include other information.[5] We denote the conditional distribution of Y_{t+h} given $\mathcal{F}_t$ as $F_{t+h,t}$, i.e. $Y_{t+h}|\mathcal{F}_t \sim F_{t+h,t}$, and the conditional density as $f_{t+h,t}$. Point forecasts conditional on $\mathcal{F}_t$ are denoted by $\hat{Y}_{t+h,t}$.

Under well-known conditions decision makers can be assumed to maximize their expected utility. Realized utility depends on the value of some outcome variable(s) and the decision maker's actions.

[5] The assumption that Y_t is adapted to $\mathcal{F}_t$ rules out the direct application of the results in this paper to, for example, volatility forecast evaluation. In such a scenario the object of interest, conditional variance, is not measurable with respect to $\mathcal{F}_t$. Using imperfect proxies for the object of interest can cause difficulties, as pointed out by Hansen and Lunde (2003).

Forecasts matter to realized utility in as far as they determine actions. Assuming that the decision maker's optimal action is a strictly monotonic function of the forecast, Machina and Granger (2003) show that expected utility maximization is equivalent to minimizing expected loss as a function of the outcome, Y_{t+h}, and the prediction, $\hat{Y}_{t+h,t}$. Under a set of further restrictions on the decision maker's optimization problem the loss function simply depends on the h-step-ahead forecast error

$$e_{t+h,t} = Y_{t+h} - \hat{Y}_{t+h,t}. \tag{1}$$

Machina and Granger derive the exact relationship between utility and loss representations of the decision maker's optimization problem including the restrictions required for the loss function only to depend on the forecast error.

2.1 Optimality Properties under Mean Squared Error Loss

The vast majority of work on optimal forecasts assumes a squared error loss function:

$$L(e_{t+h,t}) = ae^2_{t+h,t},\ a > 0. \tag{2}$$

The focus is usually on the mean squared error (MSE) across observations. Under this loss function, optimal forecasts have the following standard properties:

Proposition 1 *Let the loss function be*

$$L\left(Y_{t+h} - \hat{Y}_{t+h,t}\right) = \left(Y_{t+h} - \hat{Y}_{t+h,t}\right)^2,$$

and assume that Y is covariance stationary with $|E_t\left[Y_{t+h}\right]| < \infty$ and $E_t\left[Y^2_{t+h}\right] < \infty$ for all t and h. Then

1. *The optimal forecast of Y_{t+h} is $E_t\left[Y_{t+h}\right]$ for all forecast horizons h;*

2. *The optimal forecast error is conditionally (and unconditionally) unbiased;*

3. *The optimal h-step forecast error exhibits zero serial correlation beyond the $(h-1)\,th$ lag;*

   and if we further assume that Y is covariance stationary, then we obtain:

4. *The unconditional variance of the optimal forecast error is non-decreasing as a function of the forecast horizon.*

The proof is listed in the appendix. $E_t[.]$ is shorthand notation for $E[.|\mathcal{F}_t]$, the conditional expectation given $\mathcal{F}_t$. Properties such as these have been extensively tested in empirical studies of optimality of predictions or rationality of forecasts. The proposition shows that the standard properties of optimal forecasts are generated by the assumption of mean squared error loss alone; in particular, assumptions on the data generating process (DGP) (beyond covariance stationarity and finite first and second moments) are not required.

2.2 Asymmetric loss and a nonlinear process

Using a specific example with reasonable assumptions about the forecaster's loss function and the DGP we next show that the above properties cease to be valid when the assumption of MSE loss is relaxed. Our example is an idealized case, where in addition to knowing the form of the DGP, the forecaster is assumed to also know the parameters of the DGP, removing estimation error from the problem. Forecasts in this example are thus perfectly optimal.

We establish our results in the context of the linear-exponential (linex) loss function, which allows for asymmetries:

$$L\left(e_{t+h,t};a\right)=\exp\left\{ae_{t+h,t}\right\}-ae_{t+h,t}-1,\ a\neq 0. \tag{3}$$

This loss function has been used extensively to demonstrate the effect of asymmetric loss, c.f. Varian (1974), Zellner (1986) and Christoffersen and Diebold (1997). An optimal forecast is defined by minimizing the conditional expected loss:

$$\begin{aligned}\hat{Y}^*_{t+h,t} &\equiv \underset{\hat{Y}_{t+h,t}}{\arg\min}\, E\left[L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)|\mathcal{F}_t\right]\\ &= \underset{\hat{Y}_{t+h,t}}{\arg\min}\int L\left(y,\hat{Y}_{t+h,t}\right)f_{t+h,t}\left(y\right)dy,\end{aligned}$$

where $Y_{t+h}|\mathcal{F}_t$ has density $f_{t+h,t}$. Under the assumption that we may interchange the expectation and differentiation operators, the first order condition for the optimal forecast, $\hat{Y}^*_{t+h,t}$, takes the form

$$E_t\left[\frac{\partial L\left(Y_{t+h}-\hat{Y}^*_{t+h,t};a\right)}{\partial\hat{Y}_{t+h,t}}\right]=a-aE_t\left[\exp\left\{a\left(Y_{t+h}-\hat{Y}^*_{t+h,t}\right)\right\}\right]=0.$$

We derive analytical expressions for the optimal forecast and the expected loss using a popular nonlinear data generating process, namely a regime switching model of the type proposed by

Hamilton (1989).[6] Thus suppose that $\{Y_t\}$ is generated by a simple Gaussian mixture model driven by some underlying state process, S_t :

$$\begin{aligned} Y_{t+1} &= \mu + \sigma_{s_{t+1}} v_{t+1} \\ v_{t+1} &\sim i.i.d.\ N(0,1) \\ s_{t+1} &= 1, ..., k < \infty \end{aligned} \tag{4}$$

We assume that the state indicator function, S_{t+1}, is independently distributed of all past, current and future values of v_{t+1}, and that S_t is observable at time t.[7] The state-specific means and variances can be collected in $k \times 1$ vectors, $\boldsymbol{\mu} = \mu \boldsymbol{\iota}$, $\boldsymbol{\sigma}^2 = \left[\sigma_1^2, ..., \sigma_k^2\right]'$, where $\boldsymbol{\iota}$ is a $k \times 1$ vector of ones. Conditional on a given realization of the future state variable, $S_{t+1} = s_{t+1}$, Y_{t+1} is Gaussian with mean μ and variance $\sigma^2_{s_{t+1}}$, but future states are unknown at time t so Y_{t+1} can be strongly non-Gaussian given current information, $\mathcal{F}_t$.

At each point in time the state variable, S_{t+1}, takes an integer value between 1 and k. Following Hamilton (1989), we assume that the states are generated by a first-order stationary and ergodic Markov chain with transition probability matrix

$$\mathbf{P}(s_{t+1}|s_t) = \mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & \vdots \\ \vdots & \vdots & \ldots & p_{k-1k} \\ p_{k1} & \cdots & p_{kk-1} & p_{kk} \end{bmatrix}, \tag{5}$$

where each row of $\mathbf{P}$ sums to one. The vector comprising the probability of being in state s_{t+h} at time $t+h$ given $\mathcal{F}_t$ is denoted by $\hat{\boldsymbol{\pi}}_{s_{t+h},t}$, i.e. $\hat{\boldsymbol{\pi}}_{s_{t+h},t} = \left(\Pr(S_{t+h} = 1|\mathcal{F}_t), ..., \Pr(S_{t+h} = k|\mathcal{F}_t)\right)'$, while $\bar{\boldsymbol{\pi}}$ is the vector of unconditional or ergodic state probabilities that solve the equation $\bar{\boldsymbol{\pi}}'\mathbf{P} = \bar{\boldsymbol{\pi}}'$. Note that $\hat{\boldsymbol{\pi}}_{s_t,t}$ will be a vector of ones and zeros, as the variable S_t is $\mathcal{F}_t$-measurable.

[6] Christoffersen and Diebold (1997) characterize analytically the optimal bias under linex loss and a conditionally Gaussian process with ARCH disturbances. They derive analytically the optimal time-varying bias as a function of the conditional variance. For our purposes, however, this process is less well-suited to show violation of all four properties forecast errors have in the standard setting since this requires characterizing the forecast error distribution at many different horizons, h. The problem is that while the one-step-ahead forecast error distribution is Gaussian for a GARCH(1,1) process, this typically does not hold at longer horizons, c.f. Drost and Nijmann (1993).

[7] In most applications of regime switching models the state is assumed to be unobservable. To eliminate estimation error from the forecaster's problem in this example we assume that the state is directly observable.

Consider the h-step-ahead forecasting problem. Using the conditional normality of v_{t+h}, the expected loss is[8]

$$
\begin{aligned}
E_t\left[L\left(e_{t+h,t};a\right)\right] &= E_t\left[\exp\left\{a\left(Y_{t+h}-\hat{Y}_{t+h,t}\right)\right\}\right]-aE_t\left[Y_{t+h}\right]+a\hat{Y}_{t+h,t}-1 \\
&= \sum_{s_{t+h}=1}^{k}\hat{\pi}_{s_{t+h},t}E_t\left[\exp\left\{a\left(Y_{t+h}-\hat{Y}_{t+h,t}\right)\right\}|S_{t+h}=s_{t+h}\right] \\
&\quad -a\sum_{s_{t+h}=1}^{k}\hat{\pi}_{s_{t+h},t}E_t\left[Y_{t+h}|S_{t+h}=s_{t+h}\right]+a\hat{Y}_{t+h,t}-1 \\
&= \hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\exp\left\{a\boldsymbol{\mu}-a\hat{Y}_{t+h,t}+\frac{a^2}{2}\boldsymbol{\sigma}^2\right\}-a\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\mu}+a\hat{Y}_{t+h,t}-1. \qquad (6)
\end{aligned}
$$

Differentiating with respect to $\hat{Y}_{t+h,t}$ and setting the resulting expression equal to zero gives the first order condition

$$1=\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\exp\left\{a\boldsymbol{\mu}-a\hat{Y}^*_{t+h,t}+\frac{a^2}{2}\boldsymbol{\sigma}^2\right\}.$$

Solving for $\hat{Y}^*_{t+h,t}$ we get an expression that is easier to interpret:

$$\hat{Y}^*_{t+h,t}=\mu+\frac{1}{a}\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right), \qquad (7)$$

where $\boldsymbol{\varphi}\equiv\exp\left\{\frac{a^2}{2}\boldsymbol{\sigma}^2\right\}$. The h-step forecast error associated with the optimal forecast, denoted $e^*_{t+h,t}$, is

$$e^*_{t+h,t}=\sigma_{s_{t+h}}v_{t+h}-\frac{1}{a}\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right).$$

This expression makes it easy for us to establish the violation of property 1 in our setup:

Proposition 2 *The unconditional and conditional bias in the optimal forecast error arising under linex loss (3) for the Markov switching process (4)-(5) is given by:*

$$
\begin{aligned}
E_t\left[e^*_{t+h,t}\right] &= -\frac{1}{a}\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right) \qquad (8)\\
E\left[e^*_{t+h,t}\right] &= -\frac{1}{a}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h\rightarrow-\frac{1}{a}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right)\ \text{as } h\rightarrow\infty
\end{aligned}
$$

where $\boldsymbol{\lambda}_h\equiv\log\left(\mathbf{P}^h\boldsymbol{\varphi}\right)$. *Thus, generically, the optimal forecast is conditionally and unconditionally biased at all forecast horizons, h, and the bias persists even as h goes to infinity.*

[8] All $\exp\{\cdot\}$ and $\log(\cdot)$ operators are applied element-by-element to vector and matrix arguments.

The proof of the proposition is given in the appendix. For purposes of exposition, we present some results for a specific form of the loss function ($a = 1$) and regime switching process:

$$\begin{aligned}
\boldsymbol{\mu} &= [0,0]' \\
\boldsymbol{\sigma} &= [0.5,2]' \\
P &= \begin{bmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{bmatrix} \text{ so} \\
\bar{\boldsymbol{\pi}} &= \left[\frac{2}{3},\frac{1}{3}\right]'.
\end{aligned}$$

The unconditional mean of Y_t is zero, and the unconditional variance is $\bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 = 1.5$. This parameterization is not dissimilar to the empirical results obtained when this model is estimated on financial data. For this particular parameterization the optimal bias in $e^*_{t+1,t}$ is -1.17, indicating that it is optimal to over-predict. Figure 1 shows the density of $e_{t+h,t}$ and also plots the linex loss function. The density function has been re-scaled so as to match the range of the loss function. This figure makes it clear why the optimal bias is negative: the linex loss function with $a = 1$ penalizes positive errors (under-predictions) more heavily than negative errors (over-predictions). The optimal forecast is in the tail of the unconditional distribution of Y_t: the probability mass to the right of the optimal forecast is only 10%. Under symmetric loss the optimal forecast is the mean, and so under symmetric distributions the amount of probability mass either side of the forecast would be 50%. In Figure 2 we plot the optimal forecast bias as a function of the forecast horizon (using the steady-state weights as initial probabilities). The bias for this case is an increasing (in absolute value) function of h and asymptotes to -1.17.

We next demonstrate the violation of property 2. Let $\odot$ be the Hadamard (element-by-element) product. The result is as follows:

Proposition 3 *The variance of the forecast error arising under linex loss (3) for the Markov switching process (4)-(5) associated with the optimum forecast is given by*

$$Var\left(e^*_{t+h,t}\right) = \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot\mathbf{I} - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\lambda}_h. \tag{9}$$

This variance need not be a decreasing function of the forecast horizon, h. In the limit as h goes to infinity, the forecast error variance converges to the steady-state variance, $\bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2$.

Rather than studying the variance of the forecast error, $Var\left(e^*_{t+h,t}\right)$, it is more common to consider the mean squared error of the forecast. These two measures are closely related but differ by the squared bias. The corresponding result for the MSE is as follows:

Corollary 1 *The mean-square forecast error arising under linex loss (3) for the Markov switching process (4)-(5) associated with the optimum forecast is given by*

$$MSE\left(e^*_{t+h,t}\right) = \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 + \frac{1}{a^2}\boldsymbol{\lambda}'_h\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot\mathbf{I}\right)\boldsymbol{\lambda}_h$$

The MSE need not be a decreasing function of the forecast horizon, h. In the limit as h goes to infinity, the MSE converges to $\bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 + \left(\frac{1}{a}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right)\right)^2$.

A surprising implication of Proposition 3 is that it is not always true that $Var\left(e^*_{t+h,t}\right)$ will converge to $\bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2$ from below, that is, $Var\left(e^*_{t+h,t}\right)$ need not be increasing in h. Depending on the form of $\mathbf{P}$ and $\boldsymbol{\sigma}^2$, it is possible that $Var\left(e^*_{t+h,t}\right)$ actually *decreases* towards the unconditional variance of Y_t. Corollary 1 shows that a similar result is true for the mean-square forecast error.

Using the numerical example described above the unconditional variance of the optimal forecast error as a function of the forecast horizon is shown in Figure 3. It is clearly possible that the forecast error at the distant future has a lower variance than at the near future.[9] The reason for this surprising result lies in the mismatch of the forecast objective function, L, and the variance of the forecast error, $Var(e_{t+h,t})$, and thus does not occur when using quadratic loss (see next section). A similar mismatch of the objective function and the performance metric has been discussed by Christoffersen and Jacobs (2002), Corradi and Swanson (2002) and Sentana (1998).

Using the expression for $Var\left(e^*_{t+h,t}\right)$ in Proposition 3, we can consider two interesting special cases. First, suppose that $\sigma_1 = \sigma_2 = \sigma$ so the variable of interest is *i.i.d.* normally distributed with constant mean and variance. In this case we have:

$$Var\left(e^*_{t+h,t}\right) = \bar{\boldsymbol{\pi}}'\boldsymbol{\iota}\sigma^2 + \frac{1}{a^2}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right)\boldsymbol{\iota}'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot\mathbf{I} - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\iota}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right) = \sigma^2.$$

And so the optimal forecast error variance is constant for all forecast horizons as we would expect.

The second special case arises when the transition matrix takes the form $P = \boldsymbol{\iota}\bar{\boldsymbol{\pi}}'$. That is, the probability of being in a particular state is independent of past information, so the density of the

[9] Using the same numerical example it can be shown that the MSE decreases when moving from $h = 1$ to $h = 2$ but increases with h for $h \geq 2$. We do not report this figure in the interests of parsimony.

variable of interest is a constant mixture of two normal densities and thus is $i.i.d$ but may exhibit arbitrarily high kurtosis. In this case we have $\boldsymbol{\lambda}_h = \boldsymbol{\iota} \log(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi})$ for all h, so:

$$\begin{aligned} Var\left(e^*_{t+h,t}\right) &= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 + \frac{1}{a^2}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right)\boldsymbol{\iota}'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot I - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\iota}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right) \\ &= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2. \end{aligned}$$

Thus the optimal forecast error variance is constant for all forecast horizons. This special case shows that it is not the fat tails of the mixture density that drives the curious result regarding decreasing forecast error variance in our example. Rather, it is the combination of asymmetric loss and persistence in the conditional variance.

Violation of properties 3 and 4: Now consider the autocorrelation function of the optimal forecast errors. In the standard linear, quadratic loss framework an optimal h-step forecast is an MA process of order no greater than $(h-1)$. This implies that all autocovariances beyond the $(h-1)^{th}$ lag are zero. In our setting this need not hold:

Proposition 4 *The h-step-ahead forecast error arising under linex loss (3) for the Markov switching process (4)-(5) is serially correlated with autocovariance*

$$Cov\left[e^*_{t+h,t}, e^*_{t+h-j,t-j}\right] = \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2\mathbf{1}_{\{j=0\}} + \frac{1}{a^2}\boldsymbol{\lambda}'_h\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot P^j - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\lambda}_h. \tag{10}$$

Although this converges to zero as j goes to infinity, it can be non-zero at lags larger than h.

Using the same parameterization as in the earlier example, the autocorrelation function for various forecast horizons is presented in Figure 4. Notice the strong autocorrelation even at lags much longer than the forecast horizon, h. Thus the optimal forecast error in our set-up need not follow an $MA(h-1)$ process and the one-step-ahead forecast error need not be serially uncorrelated (property 3).[10,11]

In stark contrast, under MSE loss, all the results hold for the regime switching process previously

[10] Batchelor and Peel (1998) also noted that the optimal forecast errors obtained under linex loss may be serially correlated.

[11] We can again consider the two special cases: iid Normal ($\sigma_1 = \sigma_2 = \sigma$), and iid mixture of normals ($\mathbf{P} = \boldsymbol{\iota}\bar{\boldsymbol{\pi}}'$). Following the same logic as for the analysis of forecast error variance, it can be shown that in both of these cases the autocorrelation function equals zero for all lags greater than zero. We discuss this result more generally in Section 4.

considered. The verification of the first two properties is simple. The third property is verified by:

$$\begin{aligned} Var\left(e^*_{t+h,t}\right) &= E\left[\sigma^2_{s_{t+h}}\nu^2_{t+h}\right] \\ &= \sum_{s_{t+h}=1}^{k} \bar{\boldsymbol{\pi}}_{(s_{t+h})}\sigma^2_{s_{t+h}} E\left[\nu^2_{t+h}|S_{t+h}=s_{t+h}\right] \\ &= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2, \end{aligned}$$

which is constant for all horizons. The autocovariance of the optimal forecast errors under the regime switching process is

$$\begin{aligned} Cov\left(e^*_{t+h,t}, e^*_{t+h-j,t-j}\right) &= E\left[\sigma_{s_{t+h-j}}\sigma_{s_{t+h}}\nu_{t+h-j}\nu_{t+h}\right] \\ &= \sum_{s_{t+h-j=1}}^{k}\sum_{s_{t+h}=1}^{k} \bar{\boldsymbol{\pi}}_{\left(s_{t+h-j}\right)}\pi_{s_{h+t}|t+h-j}\sigma_{s_{t+h-j}}\sigma_{s_{t+h}} \times \\ &\quad E\left[\nu_{t+h-j}\nu_{t+h}|S_{t+h-j}=s_{t+h-j}, S_{t+h}=s_{t+h}\right] \\ &= 0 \text{ for } j \neq 0. \end{aligned}$$

Thus the optimal forecast errors are conditionally and unconditionally unbiased, have constant unconditional variance as a function of the forecast horizon, and are serially uncorrelated at all lags. This is because under MSE loss the optimal forecast error in this example is simply the term $\sigma_{s_{t+h}}\nu_{t+h}$, which is (heteroskedastic) white noise.

3 Properties of Optimal Forecasts under General Conditions

While quadratic loss is commonly used in empirical work, in a more general setting the optimal forecast, $\hat{Y}^*_{t+h,t}$, is chosen to minimize the expected loss, where the loss function need not be a function solely of the forecast error:

$$L = L\left(Y_{t+h}, \hat{Y}_{t+h,t}\right).$$

Under general loss the first order condition for the optimal forecast becomes[12],

$$0 = E_t\left[\frac{\partial L\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)}{\partial \hat{Y}_{t+h,t}}\right] = \int \frac{\partial L\left(y, \hat{Y}^*_{t+h,t}\right)}{\partial \hat{Y}_{t+h,t}} f_{t+h,t}(y)\, dy. \quad (11)$$

[12] This result relies on the ability to interchange the expectation and differentiation operators. We discuss the conditions under which this is possible below.

This condition can be rewritten using what Granger (1999) refers to as the (optimal) generalized forecast error,[13] $\psi^*_{t+h,t}$,[14]

$$\psi^*_{t+h,t} \equiv \frac{\partial L\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)}{\partial \hat{Y}_{t+h,t}} \tag{12}$$

so that (11) simplifies to

$$E_t[\psi^*_{t+h,t}] = \int \psi^*_{t+h,t} f_{t+h,t}\left(y\right) dy = 0. \tag{13}$$

Under a broad set of conditions $\psi^*_{t+h,t}$ is therefore a martingale difference sequence with respect to the information set used to compute the forecast, Ω_t. The generalized forecast error is closely related to the "generalized residual" often used in the analysis of discrete, censored or grouped variables, see Gourieroux, *et al.* (1987) and Chesher and Irish (1987) for example. Both the generalized forecast error and the generalized residual are based on first-order (or 'score') conditions.

Often $\psi^*_{t+h,t}$ can be derived explicitly. For the regime switching process/linex loss example the generalized forecast error is

$$\psi^*_{t+h,t} = a - a\exp\left\{a\sigma_{s_{t+h}}\nu_{t+h} - \log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right)\right\}. \tag{14}$$

Under MSE loss, the h-step generalized forecast error is:

$$\psi^*_{t+h,t} = -2\left(Y_{t+h} - \hat{Y}^*_{t+h,t}\right) = -2e^*_{t+h,t}, \tag{15}$$

and so the generalized forecast error is simply the negative of twice the standard forecast error. It turns out that the close relation of the standard forecast error and the generalized forecast error under mean squared error loss is the reason for the standard forecast error having such nice properties in that case. As we showed in the previous section, the properties of the standard forecast error do not hold for asymmetric loss and nonlinear processes; they do, however, hold for the generalized forecast error. We now turn our attention to proving properties of the generalized forecast error analogous to those for the standard case.

[13] Granger (1999) only considers loss functions that have the forecast error as an argument, and so defines the generalised forecast error as $\psi^*_{t+h,t} \equiv \partial L\left(e_{t+h,t}\right)/\partial e_{t+h,t}$. Our definition is more general than this.

[14] While this term is appropriate under prediction-error loss, more generally $\psi^*_{t+h,t}$ can be viewed as the marginal loss associated with a particular prediction, $\hat{Y}_{t+h,t}$.

3.1 Unbiasedness of the generalized forecast error

It is easy to establish that, although the forecast error, $e^*_{t+h,t}$, need not be unbiased, the generalized forecast error, $\psi^*_{t+h,t}$, is unbiased:

Proposition 5 *Let assumptions L1, L2* and L3 hold. Then the generalized forecast error has conditional (and unconditional) mean zero.*

Assumptions are listed in the appendix. For the regime switching process (4)-(5) the conditional mean of the generalized forecast error is

$$\begin{aligned}
E_t\left[\psi^*_{t+h,t}\right] &= a - a\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right)^{-1} E_t\left[\exp\left\{a\sigma_{s_{t+h}}\nu_{t+h}\right\}\right] \\
&= a - a\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right)^{-1} \hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h \exp\left\{\frac{a^2}{2}\boldsymbol{\sigma}^2\right\} \\
&= 0,
\end{aligned}$$

and $E\left[\psi^*_{t+h,t}\right] = 0$ by the law of iterated expectations. Thus the generalized forecast error has conditional and unconditional mean zero for all forecast horizons.

3.2 Non-decreasing expected loss as a function of the forecast horizon

In the standard framework the optimal forecast is unbiased and the loss function is quadratic. This leads to the equality of the optimal forecast error variance and the expected loss from the optimal forecast:

$$E\left[L\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)\right] = E\left[e^{*2}_{t+h,t}\right] = Var\left(e^*_{t+h,t}\right). \tag{16}$$

In general this equality will not hold, and indeed the optimal forecast error variance is not necessarily of interest; rather, the quantity of interest is the expected loss from the forecast. For the regime switching process we showed that the variance of the optimal forecast error need not be non-decreasing with the forecast horizon, contrary to results in the standard framework. The reason for this is a mismatch of the forecaster's loss/objective function and variance. Under general loss functions, if we instead look at the unconditional expected loss as a function of the forecast horizon we obtain the following result:

Proposition 6 *Let assumptions D1 and L2 hold. Then the unconditional expected loss of an optimal forecast error is a non-decreasing function of the forecast horizon. The conditional expected loss, however, need not be a non-decreasing function of the forecast horizon.*

The unconditional expected loss as a function of the forecast horizon behaves as follows in the regime switching example.

Corollary 2 *Under linex loss (3) and the regime switching process (4)-(5) the unconditional expected loss is*

$$E\left[L\left(Y_{t+h},\hat{Y}^{*}_{t+h,t};a\right)\right]=\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h\rightarrow\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right)\ \text{ as } h\rightarrow\infty.$$

For the numerical example used above, Figure 5 shows the expected loss as a function of the forecast horizon. As expected it is a non-decreasing function of h.

3.3 Serial correlation in the generalized forecast error

A property of optimal h-step ahead forecast errors under MSE loss is that they are MA processes of order no greater than $h-1$. In a non-linear framework an MA process need not completely describe the dependence properties of the generalized forecast error. However the autocorrelation function of the generalized forecast error will match some $MA\,(h-1)$ process:

Proposition 7 *Let assumptions L1, L2* and L3 hold. Then the generalized forecast error from an optimal h-step forecast made at time t exhibits zero correlation with any function of any element of the time t information set, Ω_t. In particular, the generalized forecast error will exhibit zero serial correlation for lags greater than $(h-1)$.*

Typically the transformation from the forecast error to the generalized forecast error depends on a finite set of parameters. If these are known or subject to estimation, the result implies restrictions on the loss function that are easy to test. For an example of this, see Elliott *et al.* (2002).

For completeness, we derive the autocorrelation function for the optimal generalized forecast error for our regime switching example.

Corollary 3 *The generalized forecast error from an optimal h-step forecast made at time t under the regime switching process (4)-(5) and assuming linex loss (3) has the following autocovariance* *function:*

$$Cov\left[\psi^*_{t+h,t},\psi^*_{t+h-j,t-j}\right]=\begin{cases}-a^2+a^2\sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\left(\pi'_{s_t,t}P^h\varphi\right)^{-2}\left(\pi'_{s_t,t}P^h\varphi^4\right) & j=0\\ \\ \begin{array}{c}-a^2+a^2\sum_{s_{t-j}=1}^{k}\bar{\pi}_{(s_{t-j})}\left(\pi'_{s_{t-j},t-j}P^h\varphi\right)^{-1}\times\\ \sum_{s_t=1}^{k}\pi_{s_t,t-j}\left(\pi'_{s_t,t}P^h\varphi\right)^{-1}\cdot\left(\varphi'\odot\left(\pi'_{s_t,t}P^{h-j}\right)\right)P^j\varphi\end{array} & 0<j<h\\ \\ 0 & j\geq h\end{cases}$$

where $\varphi^4\equiv\exp\left\{2a^2\sigma^2_{s_{t+h}}\right\}$.

Using the numerical example above, Figure 6 presents the autocorrelation function for the optimal generalized forecast error. As implied by Proposition 7, the autocorrelations are non-zero for $j<h$ and drop sharply to zero when $j\geq h$.

3.4 Properties of the optimal forecast error under a change of measure

So far we have shown that by changing our object of analysis from the usual forecast error to the 'generalized' forecast error we can obtain the usual properties of unbiasedness and zero serial correlation. We next consider changing the probability measure used to compute the properties of the forecast error. This analysis is akin to the use of risk-neutral densities in asset pricing, c.f. Harrison and Kreps (1979). In asset pricing one may scale the objective (or physical) probabilities by the stochastic discount factor (or the discounted ratio of marginal utilities) to obtain a risk-neutral probability measure and then apply risk-neutral pricing methods. Here we will scale the objective probability measure by the ratio of the marginal loss, $\partial L/\partial\hat{y}$, to the forecast error, and then show that under the new probability measure the standard properties hold. We call the new measure the "MSE-loss probability measure". The resulting method thus suggests an alternative means of evaluating forecasts made using asymmetric loss functions.

3.4.1 Unbiasedness and zero serial correlation under a change of measure

The conditional distribution of the forecast error, $f_{e_{t+h,t}}$, given Ω_t and a forecast $\hat{Y}_{t+h,t}$, satisfies:

$$f_{e_{t+h.t}}\left(e;\hat{Y}_{t+h,t}\right)=f_{t+h,t}\left(\hat{Y}_{t+h,t}+e\right) \tag{17}$$

for all $\left(e,\hat{Y}_{t+h,t}\right)\in R^2$ where $f_{t+h,t}$ is the conditional distribution of Y_{t+h} given Ω_t.

Definition 1 *Let assumptions L4 and L6 hold. Then the univariate "MSE-loss probability measure", $f^*_{e_{t+h,t}}$, is defined by*

$$f^*_{e_{t+h,t}}\left(e;\hat{Y}_{t+h,t}\right) = \frac{\frac{1}{e}\cdot\left.\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right|_{Y_{t+h}=\hat{Y}_{t+h,t}+e}\cdot f_{e_{t+h,t}}\left(e;\hat{Y}_{t+h,t}\right)}{E_t\left[\frac{1}{Y_{t+h}-\hat{Y}_{t+h,t}}\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right]},\ e\neq 0$$

$$\equiv \frac{\left(\psi_{t+h,t}\left(\hat{Y}_{t+h,t}\right)/e\right)\cdot f_{e_{t+h,t}}\left(e;\hat{Y}_{t+h,t}\right)}{E_t\left[\frac{1}{Y_{t+h}-\hat{Y}_{t+h,t}}\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right]},\ e\neq 0.$$

for values of $\hat{Y}_{t+h,t}$ and e such that $\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)/\partial\hat{Y}\Big|_{Y_{t+h}=\hat{Y}_{t+h,t}+e}$ exists. If

$$\lim_{e\to 0^+}\frac{1}{e}\cdot\left.\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right|_{Y_{t+h}=\hat{Y}_{t+h,t}+e} = \lim_{e\to 0^-}\frac{1}{e}\cdot\left.\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right|_{Y_{t+h}=\hat{Y}_{t+h,t}+e}$$

and

$$\left|\lim_{e\to 0^+}\frac{1}{e}\cdot\left.\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right|_{Y_{t+h}=\hat{Y}_{t+h,t}+e}\right| < \infty$$

then

$$f^*_{e_{t+h,t}}\left(0;\hat{Y}_{t+h,t}\right) = \lim_{e\to 0}\frac{1}{e}\cdot\left.\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right|_{Y_{t+h}=\hat{Y}_{t+h,t}+e}\cdot\frac{f_{e_{t+h,t}}\left(0;\hat{Y}_{t+h,t}\right)}{E_t\left[\frac{1}{Y_{t+h}-\hat{Y}_{t+h,t}}\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right]}.$$

By construction the MSE-loss probability measure f^* is absolutely continuous with respect to the usual probability measure, f, (that is, $f^* << f$). The break in $f^*_{e_{t+h,t}}$ at $e=0$ does not occur for some common loss functions. For example,

$$\begin{aligned}
MSE &: \lim_{e\to 0}\frac{1}{e}\cdot\left.\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right|_{Y_{t+h}=\hat{Y}_{t+h,t}+e} = -2\\
Linex &: \lim_{e\to 0}\frac{1}{e}\cdot\left.\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right|_{Y_{t+h}=\hat{Y}_{t+h,t}+e} = -a^2\\
PropMSE &: \lim_{e\to 0}\frac{1}{e}\cdot\left.\frac{\partial L\left(Y_{t+h},\hat{Y}_{t+h,t}\right)}{\partial\hat{Y}}\right|_{Y_{t+h}=\hat{Y}_{t+h,t}+e} = -\frac{2}{\hat{Y}^2_{t+h,t}}
\end{aligned}$$

where the *PropMSE* loss function is $L(y, \hat{y}) = (y/\hat{y} - 1)^2$. For mean absolute error loss, the limits from both directions diverge to $-\infty$, making the MSE-loss density for MAE loss undefined at the point $e = 0$.

We can show that under the MSE-loss probability measure the optimal h-step ahead forecast errors are unbiased and exhibit zero serial correlation for all lags greater than $h - 1$:

Proposition 8 *Let assumptions L1, L4 and L6 hold. Then (i) the univariate "MSE-loss probability measure", $f^*_{e_{t+h,t}}$, defined above is a proper probability density function. If we further let assumption L2* hold, then (ii) the optimal forecast error, $e^*_{t+h,t} = Y_{t+h} - \hat{Y}^*_{t+h,t}$ has conditional (and unconditional) mean zero under the MSE-loss probability measure, and (iii) the optimal forecast error is serially uncorrelated under the MSE-loss probability measure for all lags greater than $h-1$. If we also assume that Y_{t+h} is covariance stationary under the new probability measure, then (iv) $V^*\left[e^*_{t+h,t}\right]$ is non-decreasing as a function of the forecast horizon.*

Note that the following equality follows directly from Proposition 8

$$\hat{Y}^*_{t+h,t} \equiv \arg\min_{\hat{y}} E_t\left[L\left(Y_{t+h}, \hat{y}\right)\right] = \arg\min_{\hat{y}} E^*_t\left[\left(Y_{t+h} - \hat{y}\right)^2\right] \tag{18}$$

i.e., the forecast that minimizes the expected loss, where the expectation is taken with respect to the true error density is equal to the forecast that minimizes the expected squared error, where the expectation is taken with respect to the MSE-loss probability density.

Figure 7 presents an example of how the MSE-loss error density differs from the objective error density. In this figure we show the MSE-loss and objective error densities for the linex loss/regime switching example considered above. We consider the one period horizon, and show how the densities change with the state probability vector, $\hat{\boldsymbol{\pi}}_{s_t,t}$.

4 Results under restrictions on the data generating process

In this section we consider the combination of asymmetric loss functions with a restricted class of DGPs, namely those with dynamics in the conditional mean but no dynamics in the remainder of the conditional distribution. This class of DGPs is still quite broad, and includes ARMA processes and non-linear regressions. A random variable within this class may be written as:

$$\begin{aligned} Y_{t+h} &= E\left[Y_{t+h}|\mathcal{F}_t\right] + \varepsilon_{t+h}, \text{ where } \varepsilon_{t+h}|\mathcal{F}_t \sim D_h \text{ and} \\ E\left[Y_{t+h}|\mathcal{F}_t\right] &= g\left(X_t\right), \end{aligned}$$

where g is some function of $X_t \in \mathcal{F}_t$. The restriction of dynamics only in the conditional mean implies that the innovation term, ε_{t+h}, is drawn from some distribution, D_h, which will generally depend on the forecast horizon, but is *independent* of $\mathcal{F}_t$ and so is not denoted with a subscript t. Note that this restriction implies that

$$E\left[\phi\left(\varepsilon_{t+h}\right)\cdot X_t\right] = E\left[\phi\left(\varepsilon_{t+h}\right)\right]E\left[X_t\right],$$

for all functions ϕ and any vector of elements $X_t \in \mathcal{F}_t$, and that $E_t\left[\varepsilon_{t+h}\right] = 0$.

For simplicity we concentrate here on loss functions that depend only upon the forecast error, i.e., $L\left(Y_{t+h}, \hat{Y}_{t+h,t}\right) = L\left(Y_{t+h} - \hat{Y}_{t+h,t}\right) = L\left(e_{t+h,t}\right)$. Many common loss functions are of this form, for example lin-lin, quad-quad and linex. However this restriction does rule out certain loss functions, for example those that focus on proportional errors, such as $L\left(y, \hat{y}\right) = \left(y/\hat{y} - 1\right)^2$.

As an example, suppose that Y_t has zero mean and is covariance stationary with Gaussian innovations. Wold's representation theorem then establishes that it can be represented as a linear combination of serially uncorrelated white noise terms:

$$Y_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j},$$

where $\varepsilon_t = Y_t - E(Y_t | y_{t-1}, y_{t-2}, ...)$ is white noise. The first order condition for the optimal forecast under linex loss is

$$E_t\left[\exp\left\{a\left(Y_{t+h} - \hat{Y}^*_{t+h,t}\right)\right\}\right] = 1,$$

so that

$$\exp\left\{\frac{a^2\sigma^2}{2}\left(\sum_{i=0}^{h-1}\theta_i^2\right) + a\left(\sum_{i=0}^{\infty}\theta_{h+i}\varepsilon_{t-i}\right) - a\hat{Y}^*_{t+h,t}\right\} = 1$$

and the optimal forecast and forecast error are given by

$$\begin{aligned} \hat{Y}^*_{t+h,t} &= \sum_{i=0}^{\infty}\theta_{h+i}\varepsilon_{t-i} + \frac{a\sigma^2}{2}\sum_{i=0}^{h-1}\theta_i^2, \\ e^*_{t+h,t} &= \sum_{i=0}^{h-1}\theta_i\varepsilon_{t+h-i} - \frac{a\sigma^2}{2}\sum_{i=0}^{h-1}\theta_i^2. \end{aligned}$$

This is consistent with the result of Christoffersen and Diebold (1997), who show that for this combination of loss function and DGP the optimal forecast is of the form:

$$\hat{Y}^*_{t+h,t} = E_t\left[Y_{t+h}\right] + \alpha_h,$$

where α_h is a bias term that depends only on the loss function and the forecast horizon. If the conditional distribution of $Y_{t+h}|\mathcal{F}_t$ has dynamics beyond those in the conditional mean, the bias term will depend not only on the forecast horizon and the loss function, but also on the higher-order dynamics. This would correspond to a violation of our assumption that D_h is independent of $\mathcal{F}_t$.

In the case without higher-order dynamics we obtain the following serial correlation properties of the optimal forecast error:

Proposition 9 *Let the loss function and DGP satisfy assumptions D2 and L5. Then* $Cov\left(e^*_{t+h,t}, e^*_{t+h-j,t-j}\right) = 0$ *for all* $j \geq h$ *and any* $h > 0$.

The above proposition shows that under a somewhat restrictive assumption on the DGP, and only one weak assumption on the loss function, the optimal forecast errors are serially uncorrelated at lags greater than or equal to the forecast horizon, for *any* loss function. This implies that given a sequence of realizations and forecasts, $\left\{\left(Y_{t+h}, \hat{Y}_{t+h,t}\right)\right\}_{t=1}^{T}$, we may test for forecast optimality *without knowledge of the forecaster's loss function* by testing the serial correlation properties of the forecast errors. For financial applications the assumption of constant higher-order conditional moments may be too strong, but in some macroeconomic applications the assumption that all dynamics are driven by the conditional mean may be palatable. In this case, tests of forecast optimality need *not* rely on the assumption of MSE loss, as in the papers listed in footnote 1, or on the assumption that the loss function is known up to an unknown parameter vector and that the forecast model is linear, as in Elliott, *et al.* (2002). Instead forecast optimality can be tested with a large degree of robustness to the loss function of the forecaster.

4.1 Forecast error variance and expected loss

We showed in Propositions 3 and 6 that although the unconditional expected loss is always a non-decreasing function of the forecast horizon, the unconditional forecast error variance may or may not be a non-decreasing function of the forecast horizon. We next establish conditions under which the unconditional forecast error variance is non-decreasing.

Proposition 10 *Let the loss function satisfy assumptions L2, L4 and L5, and the DGP satisfy D2. Then* $V\left[e^*_{t+h,t}\right]$ *is a weakly increasing function of* h.

Like Proposition 9, the above proposition may be used to test forecast optimality in the absence of information on the forecaster's loss function under the assumption of mean-only dynamics in the variable of interest. Given a time series of forecasts with a range of horizons, Proposition 10 suggests testing that the variance of the forecast error is weakly increasing in the forecast horizon.

The assumption of no dynamics beyond the conditional mean is quite restrictive, particularly for financial applications. This suggests that the equivalence between unconditional expected loss and unconditional error variance is peculiar to the MSE loss function, and will not generally be true for other loss functions in financial applications.

These results demonstrate that it is the *combination* of asymmetric loss and dynamics in the conditional distribution beyond those in the conditional mean that generate the violations reported in Section 2.2. Under MSE loss and an arbitrary DGP we showed that the standard properties hold. Under a weak assumption on the loss function and the restriction that the conditional density has no dynamics beyond the conditional mean we showed that while the optimal forecast is in general biased, the optimal forecast errors are serially uncorrelated for lags greater than $(h-1)$ and the unconditional forecast error variance is weakly increasing in h.

5 Conclusion

This paper demonstrated that the properties of optimal forecasts that are almost always tested in the empirical literature hold only under very restrictive assumptions. We demonstrated analytically how they are violated under more general assumptions about the loss function, extending the work of Granger (1969) and Christoffersen and Diebold (1997). The properties that optimal forecasts must possess were generalized to consider situations where the loss function may be asymmetric and the data generating process may be nonlinear but strictly stationary.

We introduced a change of measure, analogous to the change of measure from objective to risk-neutral commonly employed in asset pricing. Under the new probability measure, which we call the "MSE-loss probability measure", the optimal h-step forecast error for any general loss function has zero conditional mean and zero serial correlation for all lags greater than $h-1$, i.e., the same properties as an optimal forecast under MSE loss. This is a novel line of analysis, and one that may lead to new ways of testing forecast optimality.

We have deliberately constrained our analysis in this paper to ignore parameter estimation uncertainty. Our results are all the stronger since we have shown that simply changing the loss function and allowing for nonlinear dynamics can imply that all the standard properties an optimal forecast is usually thought to possess no longer remain valid.

Our analysis does not imply that forecast rationality is not testable. Rather, it suggests that researchers have to use economic arguments to establish the underlying loss function as suggested in a recent paper by Elliott, *et al.* (2002) or, alternatively, try to conduct tests that are robust to the shape of the loss function by exploiting (testable) restrictions on the dynamics of the data generating process. Two such results were presented in section 4 of this paper; the first on the autocorrelation structure of optimal forecast errors, and the second on the variance of optimal forecast errors as a function of the forecast horizon. Deriving testable implications of forecast optimality with limited knowledge of the data generating process and the forecaster's loss function is an interesting area for future research.

References

[1] Abarbanell, J. S. and V. L. Bernard, 1992, Tests of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behaviour. Journal of Finance, 46(3), 1181 - 1207.

[2] Batchelor, R. and D.A. Peel, 1998, Rationality Testing Under Asymmetric Loss, Economics Letters, 61, 49-54.

[3] Brown, B.Y. and S. Maital, 1981, What do economists know? An empirical study of experts' expectations. Econometrica 49, 491-504.

[4] Cargill, T.F. and R.A. Meyer, 1980, The Term Structure of Inflationary Expectations and Market Efficiency. Journal of Finance 35, 57-70.

[5] Chesher, A. and M. Irish, 1987, Residual Analysis in the Grouped and Censored Normal Linear Model, Journal of Econometrics, 34, 33-61.

[6] Christoffersen, P. and K. Jacobs, 2002, The Importance of the Loss Function in Option Valuation. Working paper, Faculty of Management, McGill University.

[7] Christoffersen, P.F. and F.X. Diebold, 1996, Further Results on Forecasting and Model Selection Under Asymmetric Loss, Journal of Applied Econometrics, 11, 561-72.

[8] Christoffersen, P.F. and F.X. Diebold, 1997, Optimal prediction under asymmetric loss. Econometric Theory 13, 808-817.

[9] Corradi, V. and N. R. Swanson, 2002, A Consistent Test for Nonlinear Out of Sample Predictive Accuracy. Journal of Econometrics, 110, 353-381.

[10] De Bondt, W. F. M. and R. H. Thaler, 1990. Do Security Analysts Overreact? American Economic Review, 80(2), 52-57.

[11] De Bondt, W.F.M. and M.M. Bange, 1992, Inflation Forecast Errors and Time Variation in Term Premia. Journal of Financial and Quantitative Analysis 27, 479-496.

[12] Diebold, F.X., 2001, Elements of Forecasting (2nd edition). Southwestern.

[13] Dokko, Y. and R. H. Edelstein, 1989, How Well do Economists Forecast Stock Market Prices? A study of the Livingston Surveys. American Economic Review 79, 865-871.

[14] Drost, F.C., and T.E. Nijman, 1993, Temporal aggregation of GARCH processes, Econometrica, 61, 909-928.

[15] Ehrbeck, T. and R. Waldmann, 1996, Why Are Professional Forecasters Biased? Agency Versus Behavioral Explanations, The Quarterly Journal of Economics, 111(1), 21-40.

[16] Elliott, G., I. Komunjer, and A. Timmermann, 2002, Estimating Loss Function Parameters, working paper, Department of Economics, University of California, San Diego.

[17] Engle, R.F., 1982, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica 50, 987-1007.

[18] Figlewski, S. and P. Wachtel, 1981, The Formation of Inflationary Expectations. Review of Economics and Statistics 63, 1-10.

[19] Gourieroux, C., A. Monfort, E. Renault and A. Trongnon, 1987, Generalized Residuals, Journal of Econometrics, 34, 5-32.

[20] Granger, C.W.J., 1969, Prediction with a generalized cost function, OR, 20, 199-207.

[21] Granger, C.W.J., 1999, Outline of Forecast Theory Using Generalized Cost Functions. Spanish Economic Review 1, 161-173.

[22] Granger, C.W.J., and P. Newbold, 1986, Forecasting Economic Time Series, Second Edition. Academic Press.

[23] Granger, C.W.J. and M.H. Pesaran, 2000, Economic and Statistical Measures of Forecast Accuracy. Journal of Forecasting 19, 537-560.

[24] Granger, C.W.J. and T. Terasvirta, 1993, Modelling Nonlinear Economic Relationships. Oxford University Press.

[25] Hamilton, J.D., 1989, A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica 57, 357-86.

[26] Hansen, P.R. and A. Lunde, 2003, Consistent Preordering with an Estimated Criterion Function, with an Application to the Evaluation and Comparison of Volatility Models, working paper 2003-01, Department of Economics, Brown University.

[27] Harrison, J.M. and D.M. Kreps, 1979, Martingales and Arbitrage in Multiperiod Securities Markets, Journal of Economic Theory, 20, 381-408.

[28] Hoque, A., J.R. Magnus and B. Pesaran, 1988, The Exact Multi-Period Mean-Square Forecast Error for the First-Order Autoregressive Model, Journal of Econometrics, 39, 327-346.

[29] Ingersoll, J.E., 1987, Theory of Financial Decision Making, Rowman and Littlefield, USA.

[30] Keane, M.P. and D.E. Runkle, 1990, Testing the Rationality of Price Forecasts: New Evidence from Panel Data. American Economic Review 80, 714-735.

[31] Keane, M.P. and D.E. Runkle, 1998, Are Financial Analysts' Forecasts of Corporate Profits Rational? Journal of Political Economy 106(4), 768-805.

[32] Lakonishok, J., 1980, Stock Market Return Expectations: Some General Properties. Journal of Finance 35, 921-931.

[33] Laster, D., P. Bennett and I.S. Geoum, 1999, Rational Bias in Macroeconomic Forecasts, The Quarterly Journal of Economics, 114(1), 293-318.

[34] Lehmann, E.L. and G. Casella, 1998, Theory of Point Estimation, Second Edition, Springer, New York.

[35] Machina, M.J. and C.W.J. Granger, 2003, Decision-based Loss Functions in Forecasting and Estimation. Mimeo, UCSD.

[36] Magnus, J.R. and B. Pesaran, 1989, The Exact Multi-Period Mean-Square Forecast Error for the First-Order Autoregressive Model with an Intercept, Journal of Econometrics, 42, 157-179.

[37] Magnus, J.R. and B. Pesaran, 1991, The Bias of Forecasts from a First-Order Autoregression, Econometric Theory, 7, 222-235.

[38] Michaely, R. and K.L. Womack, 1999, Conflict of Interest and the Credibility of Underwriter Analyst Recommendations, *Review of Financial Studies*, 12(4), 653-686.

[39] Mishkin, F.S., 1981, Are Markets Forecasts Rational? American Economic Review 71, 295-306.

[40] Muth, J. F., 1961, Rational Expectations and the Theory of Price Movements. Econometrica 29(3), 315-335.

[41] Pesando, J.E., 1975, A Note on the Rationality of the Livingston Price Expectations. Journal of Political Economy 83, 849-858.

[42] Pesaran, M.H. and S. Skouras, 2001, Decision-based Methods for Forecast Evaluation. In Clements, M.P. and D.F. Hendry (eds.) Companion to Economic Forecasting. Basil Blackwell.

[43] Prendergast, C. and L. Stole, 1996, Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning, Journal of Political Economy, 104(6), 1105-1134.

[44] Scharfstein, D.S. and J.C. Stein, 1990, Herd Behavior and Investment, American Economic Review, 80(3), 465-479.

[45] Schroeter, J.R. and S.L. Smith, 1986, A Reexamination of the Livingston Price Expectations. Journal of Money, Credit and Banking 18, 239-246.

[46] Sentana, E., 1998, Least Squares Predictions and Mean-Variance Analysis, CEMFI Working Paper 9711, Centre for Monetary and Financial Studies.

[47] Skouras, S., 2001, Decisionmetrics: A Decision-based Approach to Econometric Modeling. Mimeo, Santa Fe Institute.

[48] Varian, H. R., 1974, A Bayesian Approach to Real Estate Assessment. In Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage, eds. S.E. Fienberg and A. Zellner, Amsterdam: North Holland, 195-208.

[49] West, K.D., H.J. Edison and D. Cho, 1993, A Utility-based Comparison of Some Models of Exchange Rate Volatility. Journal of International Economics 35, 23-46.

[50] Zarnowitz, V., 1985, Rational Expectations and Macroeconomic Forecasts. Journal of Business and Economic Statistics 3, 293-311.

[51] Zellner, A., 1986, Bayesian Estimation and Prediction Using Asymmetric Loss Functions. Journal of the American Statistical Association, 81, 446-451.

Appendix

We will sometimes, though not generally, make use of one of the following further assumptions on the data generating process for Y:

Assumption D1: *The data generating process for Y is strictly stationary.*

Assumption D1*: *The data generating process for Y is covariance stationary.*

Assumption D2: *The DGP is such that $Y_{t+h} = E_t[Y_{t+h}] + \varepsilon_{t+h}$, $\varepsilon_{t+h}|\mathcal{F}_t \sim D_h$, where D_h is some distribution that may depend on h, but does not depend on $\mathcal{F}_t$.*

The following properties of the loss function are assumed at various points of the analysis:

Assumption L1: *The loss function is (at least) once differentiable with respect to its second argument, except on a set of measure zero.*

Assumption L2: *$\int L(y, \hat{y}) f_{t+h,t}(y)\, dy < \infty$ for all $\hat{y} \in Y$, for all t, h.*

Note that assumption L2 implies that $E[L(Y_{t+h}, \hat{y})] < \infty$ for all $\hat{y} \in \mathcal{Y}$:

$$\begin{aligned}
E[L(Y_{t+h}, \hat{y})] &= \int_{\mathbb{R}} \int_{\mathbb{R}^\nu} L(y, \hat{y}) f_{t+h,t}\left(y|Z^t\right) f\left(Z^t\right) dy dZ^t \\
&= \int_{\mathbb{R}^\nu} E\left[L(Y_{t+h}, \hat{y})|Z^t\right] f\left(Z^t\right) dZ^t \\
&\leq \int_{\mathbb{R}^\nu} \sup_{z^t} \left(E\left[L(Y_{t+h}, \hat{y})|z^t\right]\right) f\left(Z^t\right) dZ^t \\
&= \sup_{z^t} E\left[L(Y_{t+h}, \hat{y})|z^t\right] \\
&< \infty.
\end{aligned}$$

The first line follows by decomposing the unconditional density of Y_{t+h} into its conditional density given time t information (represented by the variable $Z^t \in \mathbb{R}^\nu$) and the unconditional density of the variable Z^t. We have made the dependence of $f_{t+h,t}$ on Z^t explicit here for clarity. The second line changes the order of integration. The third line uses the fact that $E\left[L(Y_{t+h}, \hat{y})|Z^t\right] \leq \sup_{z^t}\left(E\left[L(Y_{t+h}, \hat{y})|z^t\right]\right) \forall Z^t$. Since $\sup_{z^t}\left(E\left[L(Y_{t+h}, \hat{y})|z^t\right]\right)$ does not depend on Z^t, the fourth line takes it outside the integral, and uses the fact that the density of Z^t must integrate to one.

Assumption L2*: *An interior optimum of the problem*

$$\min_{\hat{y} \in \mathcal{Y} \subset \mathbb{R}} \int L(y, \hat{y}) f_{t+h,t}(y)\, dy$$

exists for all t and h, where Y is a compact subset of R.

Assumption L3: $\left|\int \partial L(y,\hat{y})/\partial\hat{y} f_{t+h,t}(y)\,dy\right| < \infty$ *for some* $\hat{y}$, *for all* t, h.

Assumption L4: $\partial L(y,\hat{y})/\partial\hat{y} \leq (\geq)\, 0$ *for* $y - \hat{y} \geq (\leq)\, 0$.

Assumption L2 simply ensures that the conditional expected loss from a forecast is finite, for some finite forecast. Assumptions L1 and L2* allow us to use the first-order condition of the minimization problem to study the optimal forecast. One set of sufficient conditions, see Lehmann and Casella (1998) for example, for Assumption L2* to hold are Assumption L2 and:

Assumption L5: *The loss function is solely a function of the forecast error.*

Assumption L5*: *The loss function is a non-monotone, convex function solely of the forecast error.*

We do not require that L is differentiable with respect to its second argument everywhere. Note also that we do not need to assume a unique optimum (though this is obtained if we impose Assumption L5*, with the convexity of the loss function being strict). Assumption L4 imposes that the loss function is weakly increasing as we move out from the point where $\hat{y} = y$. It is common to impose that $L(y,y) = 0$, i.e., the loss from a perfect forecast is zero, but this is obviously just a normalization and is not required here. Assumption L3 is required to interchange expectation and differentiation: $\partial E_t[L(Y_{t+h},\hat{y})]/\partial\hat{y} = E_t[\partial L(Y_{t+h},\hat{y})/\partial\hat{y}]$. The bounds on the integral in the left-hand side of this expression are unaffected by the choice of $\hat{y}$, and so two of the terms in Leibnitz's rule drop out, meaning we need only assume that the term on the right-hand side is finite.

In defining the MSE-loss probability measure we need to make the following assumption:

Assumption L6: $0 < \left|E_t\left[(Y_{t+h}-\hat{y})^{-1}\partial L(Y_{t+h},\hat{y})/\partial\hat{y}\right]\right| < \infty$ *for all* t, h *and* $\hat{y}$.

Proof of Proposition 1. Given $L\left(Y_{t+h},\hat{Y}_{t+h}\right) \equiv \left(Y_{t+h}-\hat{Y}_{t+h}\right)^2$, and assuming at least 2 finite conditional moments of Y_{t+h}, the first order condition implies that

$$\begin{aligned}
\frac{\partial E_t\left[L\left(Y_{t+h},\hat{Y}^*_{t+h,t}\right)\right]}{\partial\hat{Y}_{t+h,t}} &= -2\left(E_t[Y_{t+h}] - \hat{Y}^*_{t+h,t}\right) = 0, \text{ so} \\
\hat{Y}^*_{t+h,t} &= E_t[Y_{t+h}], \text{ and} \\
e^*_{t+h,t} &= Y_{t+h} - E_t[Y_{t+h}].
\end{aligned}$$

Thus the optimal forecast under MSE is conditionally and unconditionally unbiased for all forecast horizons and for all DGPs.

The remainder of the proof follows directly from the proofs of Propositions 6 and 7, presented below, when one observes the relation between the forecast error and the generalized forecast error (defined in Section 3), $\psi^*_{t+h,t}$, for the mean squared loss case: $e^*_{t+h,t} = -\frac{1}{2}\psi^*_{t+h,t}$, and noting that the MSE loss function satisfies assumptions L1 and L3, and also assumption L5* which then implies a unique interior optimum (see Lehmann and Casella, 1998, for example). ■

Proof of Proposition 2. The h-step-ahead forecast error has a conditional expectation of

$$E_t\left[e^*_{t+h,t}\right] = -\frac{1}{a}\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right)$$

which, since $\mathbf{P}$ is a probability matrix with an eigenvalue of unity, is different from zero even when $h \to \infty$. The unconditional expectation of the forecast error is

$$\begin{aligned}
E\left[e^*_{t+h,t}\right] &= E\left[E_t\left[e^*_{t+h,t}\right]\right] \\
&= \sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}E\left[-\frac{1}{a}\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right)|S_t = s_t\right] \\
&= -\frac{1}{a}\sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\log\left(\boldsymbol{\iota}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right) \\
&= -\frac{1}{a}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h,
\end{aligned}$$

where $\boldsymbol{\lambda}_h = \log\left(\mathbf{P}^h\boldsymbol{\varphi}\right)$ and $\boldsymbol{\iota}_{s_t} = \Pr\left[S_t|S_t = s_t\right]$ is a $k \times 1$ zero-one selection vector that is unity in the s_tth element and is zero otherwise.

Clearly the unconditional bias remains, in general, non-zero for all h. In the limit as $h \to \infty$,

$$E\left[e^*_{t+h,t}\right] \to -\frac{1}{a}\bar{\boldsymbol{\pi}}'\log\left(\boldsymbol{\iota}\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right) = -\frac{1}{a}\bar{\boldsymbol{\pi}}'\boldsymbol{\iota}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right) = -\frac{1}{a}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right)$$

which is also, in general, non-zero. ■

Proof of Proposition 3. From Proposition 2 we have

$$\begin{aligned}
Var\left(e^*_{t+h,t}\right) &= E\left[e^{*2}_{t+h,t}\right] - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&= E\left[\left(\sigma_{s_{t+h}}\nu_{t+h} - \frac{1}{a}\log\left(\hat{\boldsymbol{\pi}}_{s_t,t}'\mathbf{P}^h\boldsymbol{\varphi}\right)\right)^2\right] - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&= E\left[\sigma^2_{s_{t+h}}\nu^2_{t+h}\right] - \frac{2}{a}E\left[\sigma_{s_{t+h}}\nu_{t+h}\log\left(\hat{\boldsymbol{\pi}}_{s_t,t}'\mathbf{P}^h\boldsymbol{\varphi}\right)\right] \\
&\quad + \frac{1}{a^2}E\left[\log\left(\hat{\boldsymbol{\pi}}_{s_t,t}'\mathbf{P}^h\boldsymbol{\varphi}\right)^2\right] - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&= \sum_{s_{t+h}=1}^{k}\bar{\pi}_{(s_{t+h})}E\left[\sigma^2_{s_{t+h}}\nu^2_{t+h}|S_{t+h}=s_{t+h}\right] - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&\quad + \frac{1}{a^2}\sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}E\left[\log\left(\hat{\boldsymbol{\pi}}_{s_t,t}'\mathbf{P}^h\boldsymbol{\varphi}\right)\cdot\log\left(\hat{\boldsymbol{\pi}}_{s_t,t}'\mathbf{P}^h\boldsymbol{\varphi}\right)|S_t=s_t\right] \\
&= \sum_{s_{t+h}=1}^{k}\bar{\pi}_{(s_{t+h})}\sigma^2_{s_{t+h}} - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&\quad + \frac{1}{a^2}\sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\log\left(\boldsymbol{\varphi}'\mathbf{P}^{h\prime}\right)\boldsymbol{\iota}_{s_t}\boldsymbol{\iota}_{s_t}'\log\left(\mathbf{P}^h\boldsymbol{\varphi}\right) \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\sum_{s_t=1}^{k}\bar{\boldsymbol{\pi}}_{(s_t)}\boldsymbol{\iota}_{s_t}\boldsymbol{\iota}_{s_t}'\right)\boldsymbol{\lambda}_h - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&= \bar{\boldsymbol{\pi}}'\sigma^2 + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot I - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\lambda}_h.
\end{aligned}$$

Here $\bar{\pi}_{(i)}$ is the i^{th} element of the vector $\bar{\boldsymbol{\pi}}$, the outer product $\boldsymbol{\iota}_{s_t}\boldsymbol{\iota}_{s_t}'$ is a $k\times k$ matrix of all zeros, except for the $(s_t,s_t)^{th}$ element, which equals one. To examine the variance of the optimal h-step ahead forecast as $h\to\infty$, notice that

$$\boldsymbol{\lambda}_\infty \equiv \lim_{h\to\infty}\boldsymbol{\lambda}_h = \boldsymbol{\iota}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right).$$

Furthermore, for any vector $\bar{\boldsymbol{\pi}}$ such that $\bar{\boldsymbol{\pi}}'\boldsymbol{\iota}=1$,

$$\boldsymbol{\iota}'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot I - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\iota} = \boldsymbol{\iota}'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot I\right)\boldsymbol{\iota} - \boldsymbol{\iota}'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\iota} = \bar{\boldsymbol{\pi}}'\boldsymbol{\iota} - \left(\bar{\boldsymbol{\pi}}'\boldsymbol{\iota}\right)'\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\iota}\right) = 0.$$

As $h\to\infty$, the variance of the optimal h-step ahead forecast therefore converges to

$$\begin{aligned}
Var\left[e^*_{t+h,t}\right] &\to \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 + \frac{1}{a^2}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right)\boldsymbol{\iota}'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot I - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\iota}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right) \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2.
\end{aligned}$$

■

Proof of Corollary 1. Follows directly from the proof of Proposition 3. ■

Proof of Proposition 4. The autocovariance function for an h-step forecast is:

$$
\begin{aligned}
Cov\left[e^*_{t+h,t}, e^*_{t+h-j,t-j}\right] &= E\left[e^*_{t+h,t} \cdot e^*_{t+h-j,t-j}\right] - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&= E\left[\left(\sigma_{s_{t+h}}\nu_{t+h} - \frac{1}{a}\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right) \times \right.\right. \\
&\quad \left.\left. \sigma_{s_{t+h-j}}\nu_{t+h-j} - \frac{1}{a}\log\hat{\boldsymbol{\pi}}'_{s_{t-j},t-j}\mathbf{P}^h\boldsymbol{\varphi}\right)\right] - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&= E\left[\left(\sigma_{s_{t+h-j}}\sigma_{s_{t+h}}\nu_{t+h-j}\nu_{t+h}\right)\right] - \frac{1}{a}E\left[\sigma_{s_{t+h-j}}\nu_{t+h-j}\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right)\right] \\
&\quad -\frac{1}{a}E\left[\sigma_{s_{t+h}}\nu_{t+h}\log\left(\hat{\boldsymbol{\pi}}'_{s_{t-j},t-j}\mathbf{P}^h\boldsymbol{\varphi}\right)\right] \\
&\quad +\frac{1}{a^2}E\left[\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right)\log\left(\hat{\boldsymbol{\pi}}'_{s_{t-j},t-j}\mathbf{P}^h\boldsymbol{\varphi}\right)\right] - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2\mathbf{1}_{\{j=0\}} - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h + \frac{1}{a^2}\sum_{s_{t-j}=1}^{k}\sum_{s_t=1}^{k}\bar{\pi}_{(s_{t-j})}\pi_{s_t|s_{t-j}} \times \\
&\quad E\left[\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right)\log\left(\hat{\pi}'_{s_{t-j},t-j}\mathbf{P}^h\boldsymbol{\varphi}\right)|S_{t-j} = s_{t-j}, S_t = s_t\right] \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 1_{\{j=0\}} - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&\quad +\frac{1}{a^2}\sum_{s_{t-j}=1}^{k}\sum_{s_t=1}^{k}\bar{\pi}_{(s_{t-j})}\pi_{s_t|s_{t-j}}\log\left(\boldsymbol{\iota}'_{s_t}\mathbf{P}^h\boldsymbol{\varphi}\right)\log\left(\boldsymbol{\iota}'_{s_{t-j}}\mathbf{P}^h\boldsymbol{\varphi}\right) \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 1_{\{j=0\}} - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&\quad +\frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\sum_{s_{t-j}=1}^{k}\sum_{s_t=1}^{k}\bar{\pi}_{(s_{t-j})}\pi_{s_t|s_{t-j}}\boldsymbol{\iota}_{s_t}\boldsymbol{\iota}'_{s_{t-j}}\right)\boldsymbol{\lambda}_h \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 1_{\{j=0\}} - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&\quad +\frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\sum_{s_{t-j}=1}^{k}\bar{\pi}_{(s_{t-j})}\left(\sum_{s_t=1}^{k}\pi_{s_t|s_{t-j}}\boldsymbol{\iota}_{s_t}\right)\boldsymbol{\iota}'_{s_{t-j}}\right)\boldsymbol{\lambda}_h \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 1_{\{j=0\}} - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&\quad +\frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\sum_{s_{t-j}=1}^{k}\bar{\pi}_{(s_{t-j})}\mathbf{P}^{j'}\boldsymbol{\iota}_{s_{t-j}}\boldsymbol{\iota}'_{s_{t-j}}\right)\boldsymbol{\lambda}_h \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 1_{\{j=0\}} + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot\mathbf{P}^j - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\lambda}_h
\end{aligned}
$$

For fixed h, as $j \to \infty$, $Cov\left[e^*_{t+h,t}, e^*_{t+h-j,t-j}\right] \to \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}' \odot \boldsymbol{\iota}\bar{\boldsymbol{\pi}}'\right) - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\lambda}_h = 0.$ ■

Proof of Proposition 5. Assumptions L1 and L2* allow us to analyse the first-order condition for the optimal forecast, and assumption L3 permits the exchange of differentiation and expectation in the first-order condition, giving us

$$E_t\left[\psi^*_{t+h,t}\right] = E_t\left[\frac{\partial L\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)}{\partial \hat{Y}_{t+h,t}}\right] = 0,$$

by the optimality of $\hat{Y}^*_{t+h,t}$. $E\left[\psi^*_{t+h,t}\right] = 0$ follows from the law of iterated expectations. ■

Proof of Proposition 6. By strict stationarity of $\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)$ for all h and j (D1) we have

$$E\left[E_t\left[L\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)\right]\right] = E\left[E_{t-j}\left[L\left(Y_{t+h-j}, \hat{Y}^*_{t+h-j,t-j}\right)\right]\right]$$

and so the unconditional expected loss only depends on the forecast horizon, and not on the period when the forecast was made.

By the optimality of the forecast $\hat{Y}^*_{t+h,t}$ we also have, for $\forall j \geq 0$,

$$\begin{aligned}
E_t\left[L\left(Y_{t+h}, \hat{Y}^*_{t+h,t-j}\right)\right] &\geq E_t\left[L\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)\right] \\
E\left[L\left(Y_{t+h}, \hat{Y}^*_{t+h,t-j}\right)\right] &\geq E\left[L\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)\right] \\
E\left[L\left(Y_{t+h+j}, \hat{Y}^*_{t+h+j,t}\right)\right] &\geq E\left[L\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)\right]
\end{aligned}$$

where the second line follows using the law of iterated expectations and the third line follows from strict stationarity. Hence the unconditional expected loss is a non-decreasing function of the forecast horizon.

To show that the conditional expected loss may be an increasing or a decreasing function of the forecast horizon we need only construct an example. We will use the 2-state regime switching/linex loss example from Section 2.2. Assume that $\hat{\boldsymbol{\pi}}_{s_t,t} = [0.95, 0.05]'$. Then from equations (6) and (7) we know that optimum forecasts and resulting conditional expected losses are: $\hat{Y}^*_{t+1,t} = 0.5376$, $\hat{Y}^*_{t+2,t} = 0.6616$, $E_t\left[L\left(Y_{t+1}, \hat{Y}^*_{t+1,t}\right)\right] = 3.1685$ and $E_t\left[L\left(Y_{t+2}, \hat{Y}^*_{t+2,t}\right)\right] = 3.7390$. If, on the other hand, $\hat{\boldsymbol{\pi}}_{s_t,t} = [0.05, 0.95]'$ then the optimal forecasts and resulting conditional expected losses are: $\hat{Y}^*_{t+1,t} = 1.8714$, $\hat{Y}^*_{t+2,t} = 1.7927$, $E_t\left[L\left(Y_{t+1}, \hat{Y}^*_{t+1,t}\right)\right] = 8.1050$ and $E_t\left[L\left(Y_{t+2}, \hat{Y}^*_{t+2,t}\right)\right] = 7.9995$. So if we start from a point where there is a high probability of being in the low volatility state, then the conditional expected loss is increasing with h. But if we start from a point where there is a high probability of being in the high volatility state, then the conditional expected loss is decreasing with h. ■

Proof of Corollary 2. Follows using similar steps as in the proofs of Propositions 3 and 4. Details are available from authors upon request. ■

Proof of Proposition 7. Since $\sigma\left(Y_t, Y_{t-1}, ...\right) \subseteq \Omega_t$ by assumption we know that $\psi^*_{t+h-j,t-j} = \partial L\left(Y_{t+h-j}, \hat{Y}^*_{t+h-j,t-j}\right)/\partial \hat{y}$ is an element of $\mathcal{F}_t$ for all $j \geq h$. Assumptions L1 and L2* again allow us to analyze the first-order condition for the optimal forecast, and assumption L3 permits the exchange of differentiation and expectation in the first-order condition. We thus obtain have

$$E\left[\psi^*_{t+h,t}|\mathcal{F}_t\right] = E\left[\left.\frac{\partial L\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)}{\partial \hat{Y}}\right| \Omega_t\right] = 0,$$

which implies $E\left[\psi^*_{t+h,t} \cdot \gamma\left(X_t\right)\right] = 0$ for all $X_t \in \mathcal{F}_t$ and all functions γ. Thus $\psi^*_{t+h,t}$ is uncorrelated with any function of any element of $\mathcal{F}_t$. This implies that

$$E\left[\psi^*_{t+h,t} \cdot \psi^*_{t+h-j,t-j}\right] = 0 \text{ for all } j \geq h$$

and so $\psi^*_{t+h,t}$ is uncorrelated with $\psi^*_{t+h-j,t-j}$. ■

Proof of Corollary 3. Follows using similar steps as in the proofs of Propositions 3 and 4. Details are available from authors upon request. ■

Proof of Proposition 8. We first need to show that $f^*_{e_{t+h}} \geq 0$ for all possible values of e, and that
$\int f^*_{e_{t+h,t}}\left(u; \hat{Y}_{t+h,t}\right) du = 1$. By assumption L4 we have

$$\frac{1}{e} \cdot \left.\frac{\partial L\left(Y_{t+h}, \hat{Y}_{t+h,t}\right)}{\partial \hat{Y}_{t+h,t}}\right|_{Y_{t+h}=\hat{Y}_{t+h,t}+e} < 0 \text{ for all } e \neq 0, \text{ where } \left.\frac{\partial L\left(Y_{t+h}, \hat{Y}_{t+h,t}\right)}{\partial \hat{Y}_{t+h,t}}\right|_{Y_{t+h}=\hat{Y}_{t+h,t}+e} \text{ exists.}$$

Thus the numerator is non-positive, and the denominator is negative (and finite by assumption L6), so $f^*_{e_{t+h,t}}\left(e; \hat{Y}_{t+h,t}\right) \geq 0$, if $f_{e_{t+h,t}}\left(e; \hat{Y}_{t+h,t}\right) \geq 0$. By the construction of $f^*_{e_{t+h,t}}$ it is clear that it integrates to 1.

To prove (ii), note that

$$\begin{aligned}
E^*_t\left[e^*_{t+h,t}\right] &\equiv \int e f^*_{e_{t+h,t}}\left(e; \hat{Y}_{t+h,t}\right) de \\
&= E_t\left[\frac{\psi^*_{t+h,t}}{e^*_{t+h,t}}\right]^{-1} \cdot \int \left.\frac{\partial L\left(Y_{t+h}, \hat{Y}^*_{t+h,t}\right)}{\partial \hat{Y}_{t+h,t}}\right|_{Y_{t+h}=\hat{Y}^*_{t+h,t}+e} \cdot f_{e_{t+h,t}}\left(e; \hat{Y}_{t+h,t}\right) de \\
&= 0.
\end{aligned}$$

since the second part of the second line equals zero by the first-order condition for an optimal forecast. The unconditional mean is also zero by the law of iterated expectations.

In the proof of part (iii) we make reference to the bivariate MSE-loss probability measure, but do not need to explicitly define it in order to obtain the result. Since $E^*\left[e^*_{t+h,t}\right]=0$ we need only show that $E^*\left[e^*_{t+h,t}\cdot e^*_{t+h+j,t+j}\right]=0$ for $j\geq h$.

$$\begin{aligned} E_t^*\left[e^*_{t+h,t}\cdot e^*_{t+h+j,t+j}\right] &= E_t^*\left[e^*_{t+h,t}\cdot E^*_{t+j}\left[e^*_{t+h+j,t+j}\right]\right] \text{ for } j\geq h \\ &= 0, \end{aligned}$$

by part (ii). $E_t^*\left[e^*_{t+h,t}\cdot e^*_{t+h+j,t+j}\right]=0$ follows by the law of iterated expectations.

For part (iv) note that $V_t^*\left[e^*_{t+h,t}\right]=E_t^*\left[e^{*2}_{t+h,t}\right]$, and $V^*\left[e^*_{t+h,t}\right]=E^*\left[e^{*2}_{t+h,t}\right]$. Further, note that $E_t^*\left[e_{t+h,t}\right]=0=E_t^*\left[Y_{t+h}-\hat{Y}^*_{t+h,t}\right]$ is the first-order condition of $\arg\min_{\hat{y}} E_t^*\left[\left(Y_{t+h}-\hat{y}\right)^2\right]$, so

$$\begin{aligned} E_t^*\left[\left(Y_{t+h}-\hat{Y}^*_{t+h,t}\right)^2\right] &\leq E_t^*\left[\left(Y_{t+h}-\hat{Y}^*_{t+h,t-j}\right)^2\right] \forall j\geq 0 \\ V^*\left[e^*_{t+h,t}\right] &= E^*\left[\left(Y_{t+h}-\hat{Y}^*_{t+h,t}\right)^2\right]\leq E^*\left[\left(Y_{t+h}-\hat{Y}^*_{t+h,t-j}\right)^2\right] \\ &= E^*\left[\left(Y_{t+h}-\hat{Y}^*_{t+h+j,t}\right)^2\right]=V^*\left[e^*_{t+h+j,t}\right] \end{aligned}$$

by the optimality of $\hat{Y}^*_{t+h,t}$, the law of iterated expectations and the assumption of stationarity under the MSE-loss probability measure. ■

Proof of Proposition 9. Under assumption L5 Granger (1969) and Christoffersen and Diebold (1997) show that the optimal forecast may be written as

$$\hat{Y}^*_{t+h,t}=E_t\left[Y_{t+h}\right]+\alpha_h$$

and so the optimal forecast error is $e^*_{t+h,t}=Y_{t+h}-\hat{Y}^*_{t+h,t}=\varepsilon_{t+h}-\alpha_h$. Since α_h is constant for fixed h,

$$\begin{aligned} Cov\left[e^*_{t+h,t},e^*_{t+h-j,t-j}\right] &= E\left[\varepsilon_{t+h}\cdot\varepsilon_{t+h-j}\right] \\ &= E\left[E_t\left[\varepsilon_{t+h}\right]\cdot\varepsilon_{t+h-j}\right] \text{ for } \forall j\geq h \\ &= 0. \end{aligned}$$

■

Proof of Proposition 10. Consider $h > 0$ and $j > 0$. Let

$$\begin{aligned} Y_{t+h+j} &= E_t\left[Y_{t+h+j}\right] + \eta_{t+h+j},\ \eta_{t+h+j}|\mathcal{F}_t \sim D_{h+j} \\ Y_{t+h+j} &= E_{t+j}\left[Y_{t+h+j}\right] + \varepsilon_{t+h+j},\ \varepsilon_{t+h+j}|\mathcal{F}_t \sim D_h \end{aligned}$$

From Christoffersen and Diebold (1997) we know that under the above assumptions $\hat{Y}^*_{t+h,t} = E_t\left[Y_{t+h}\right] + \alpha_h$, so

$$\begin{aligned} e^*_{t+h+j,t} &= \eta_{t+h+j} - \alpha_{h+j} \\ e^*_{t+h+j,t+j} &= \varepsilon_{t+h+j} - \alpha_h \end{aligned}$$

where α_h and α_{h+j} are constants. Thus $V_t\left[e^*_{t+h+j,t}\right] = V_t\left[\eta_{t+h+j}\right] \equiv \sigma^2_{h+j}$, and $V_t\left[e^*_{t+h+j,t+j}\right] \equiv \sigma^2_h$. Note also that $V\left[e^*_{t+h+j,t}\right] = E\left[E_t\left[\eta^2_{t+h+j}\right]\right] = \sigma^2_{h+j}$, and similarly $V\left[e^*_{t+h+j,t+j}\right] = \sigma^2_h$. Now we seek to show that $\sigma^2_{h+j} \geq \sigma^2_h$.

$$\begin{aligned} V\left[e^*_{t+h+j,t}\right] &= V_t\left[Y_{t+h+j} - E_t\left[Y_{t+h+j}\right]\right] \\ &= V_t\left[\varepsilon_{t+h+j} + \left(E_{t+j}\left[Y_{t+h+j}\right] - E_t\left[Y_{t+h+j}\right]\right)\right] \\ &= \sigma^2_h + V_t\left[E_{t+j}\left[Y_{t+h+j}\right]\right] + 2Cov_t\left[\varepsilon_{t+h+j}, E_{t+j}\left[Y_{t+h+j}\right] - E_t\left[Y_{t+h+j}\right]\right] \\ &\geq \sigma^2_h \\ &= V\left[e^*_{t+h,t}\right] \end{aligned}$$

where the first equality follows from the equality of the conditional and unconditional variance of the forecast error in this scenario; the third equality follows from the fact that $E_t\left[Y_{t+h+j}\right]$ is constant given $\mathcal{F}_t$; the weak inequality follows from the non-negativity of $V_t\left[E_{t+1}\left[Y_{t+2}\right]\right]$ and that $E_{t+j}\left[\varepsilon_{t+h+j} \cdot \phi\left(X^{t+j}\right)\right] = 0$; the final equality follows from the fact that D_h does not change with t. The cases where $h = 0$ and/or $j = 0$ are trivial. Thus $V\left[e^*_{t+h+j,t}\right] \geq V\left[e^*_{t+h,t}\right]\ \forall\, h,\, j \geq 0$. ■

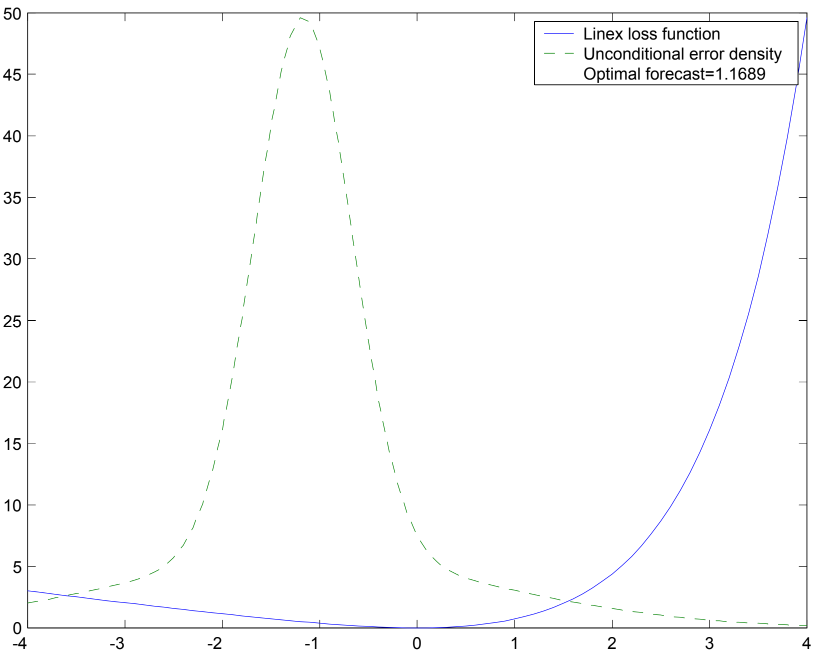

Figure 1: *Linear-exponential loss function and unconditional optimal forecast error density, two-state regime switching example.*

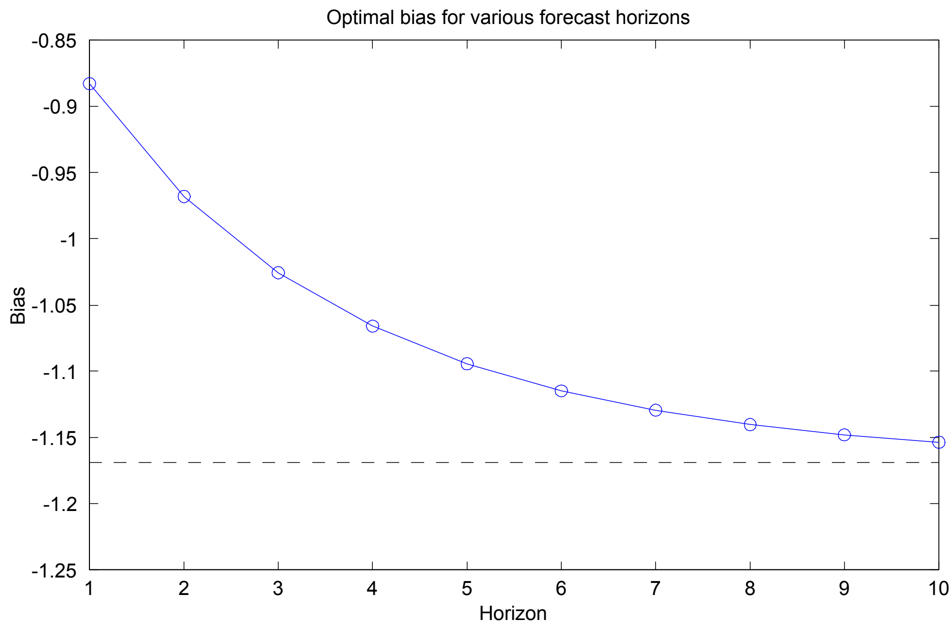

Figure 2: *Bias in the optimal forecast for various forecast horizons, two-state regime switching example.*

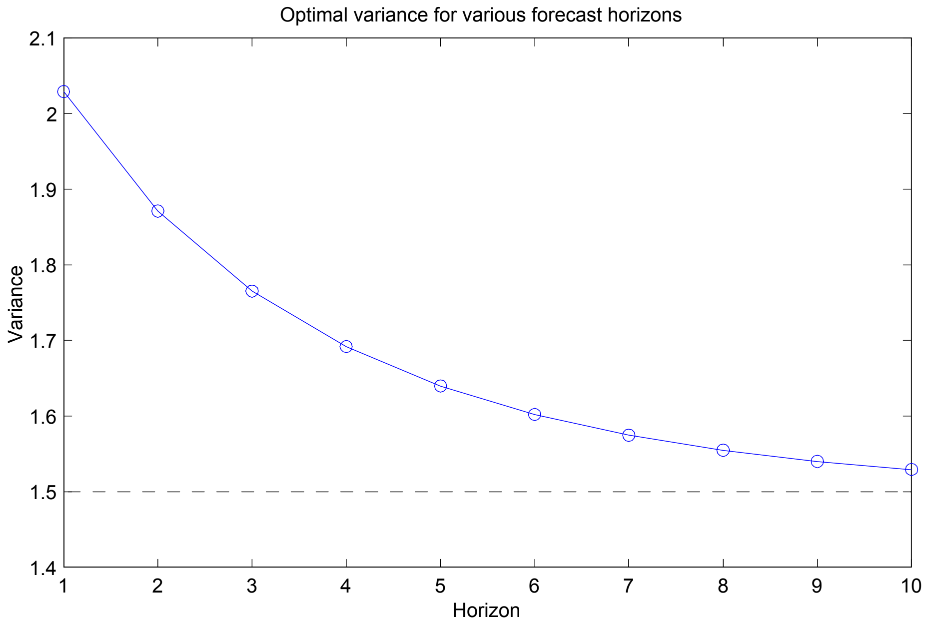

Figure 3: *Variance of the optimal h-step forecast error for various forecast horizons, two-state regime switching example.*

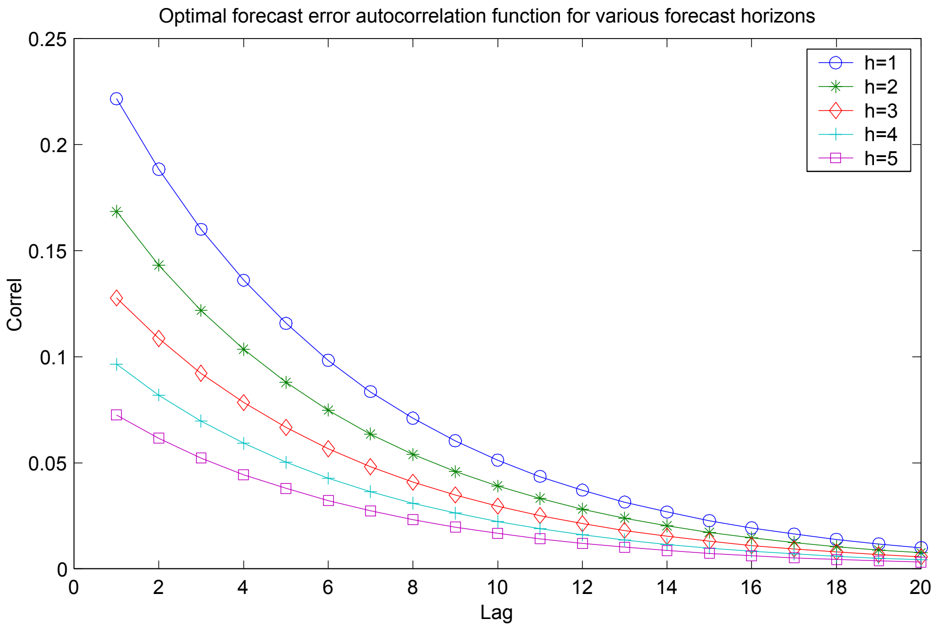

Figure 4: *Autocorrelation in the optimal h-step forecast error for various forecast horizons, two-state regime switching example.*

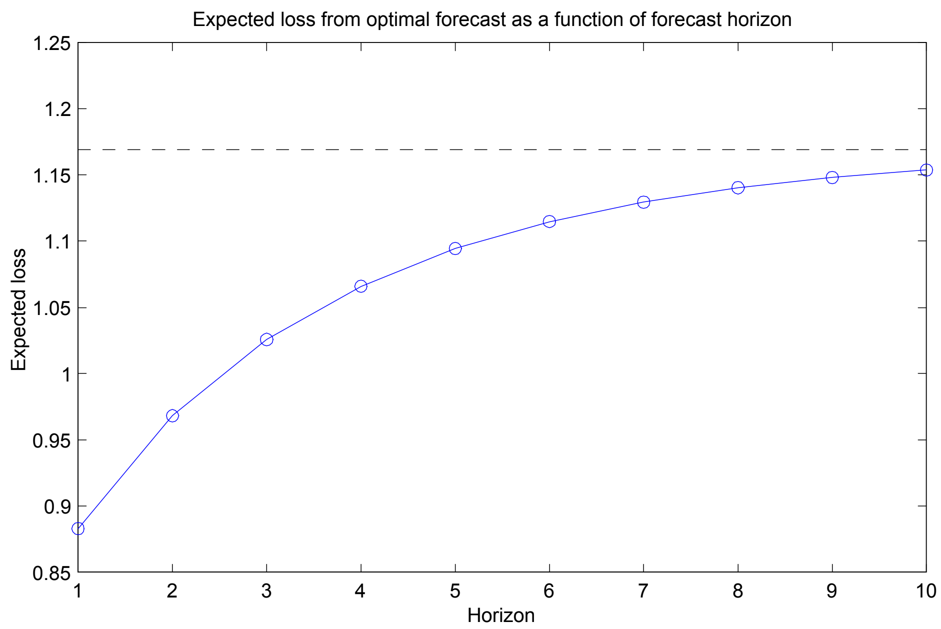

Figure 5: *Expected loss from the optimal forecast for various forecast horizons, two-state regime switching example.*

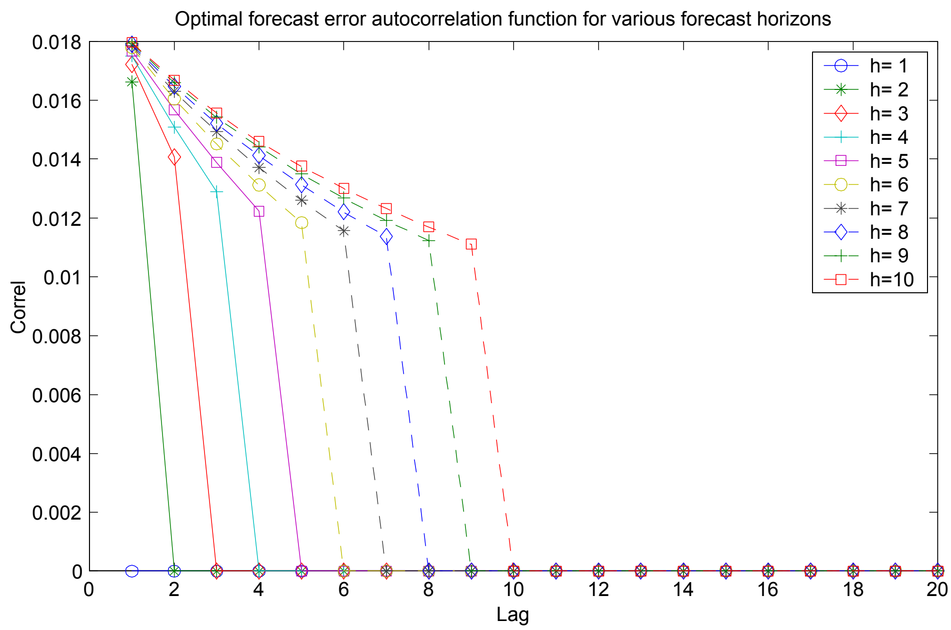

Figure 6: *Autocorrelation in the generalised optimal forecast error for various forecast horizons, two-state regime switching example.*

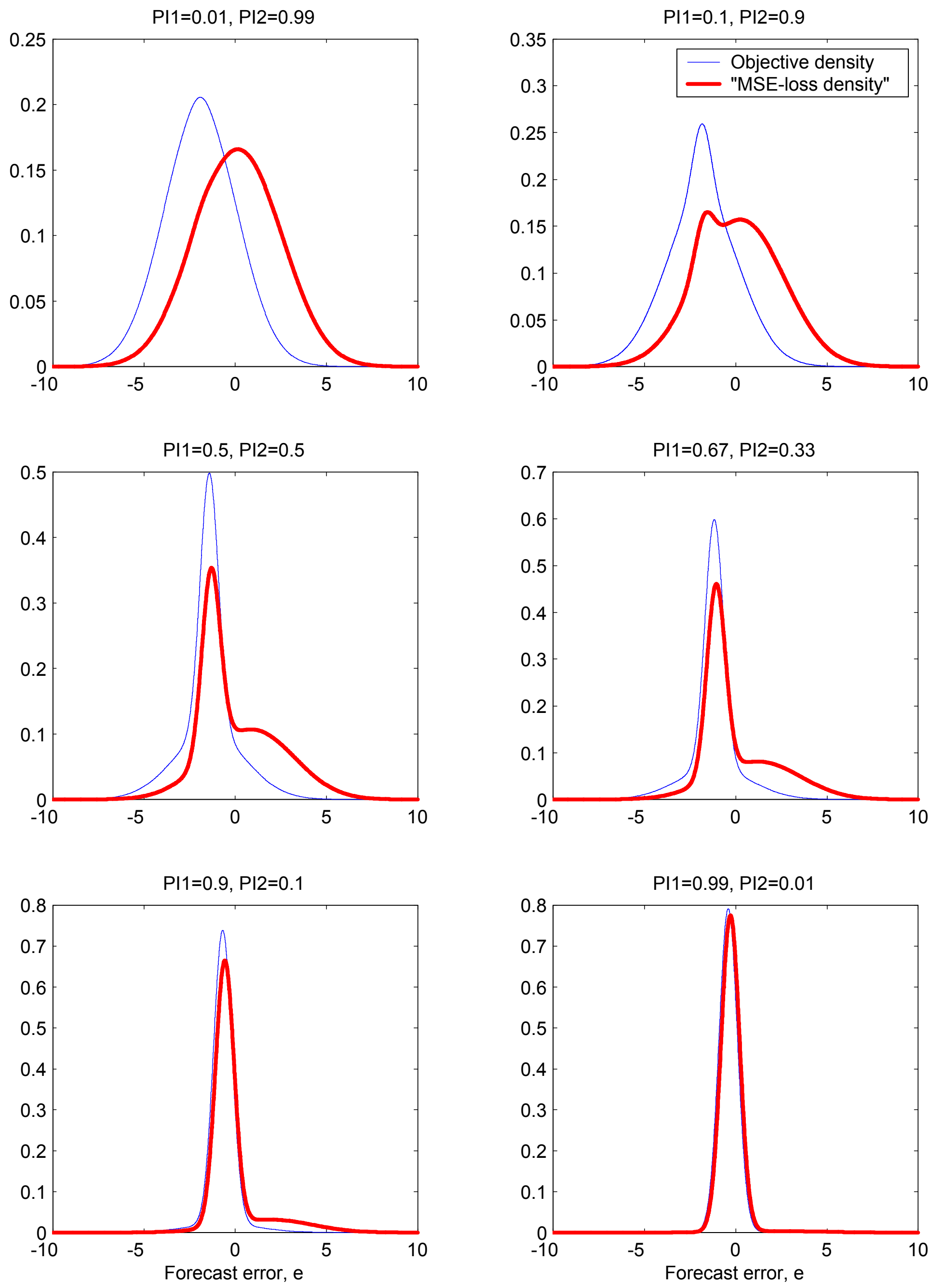

Figure 7: *Objective and "MSE-loss" error densities for the regime switching example, one-step forecast horizon, for various values of the state probability vector,* $\hat{\boldsymbol{\pi}}_{s_t,t}$.