Financial Econometrics Research Centre

WORKING PAPERS SERIES

WP04-05

Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity

Andrew Patton and Allan Timmermann

Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity*

Andrew J. Patton
London School of Economics

Allan Timmermann
University of California, San Diego

19 June, 2004

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 nonlinear data generating processes and thus may be very misleading as a benchmark for an optimal forecast. We establish instead that a suitable transformation of the forecast error - known as the generalized forecast error - possesses an equivalent set of properties.

Keywords: properties of optimal forecasts, loss function, nonlinear data generating process.

J.E.L. Codes: C53, C22, C52

---

*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, Stanley Zin 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. Patton: London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. Email: a.patton@lse.ac.uk. Timmermann: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0508, U.S.A. Email: atimmerm@ucsd.edu.

1 Introduction

Properties of optimal forecasts under mean squared error (MSE) loss include unbiasedness of the forecast, lack of serial correlation in one-step-ahead forecast errors, serial correlation of (at most) 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. 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 properties has not previously been demonstrated.[1]

In this paper we provide new analytical results on the violation of the standard optimality properties of optimal forecasts under linear-exponential (linex) loss and a Markov switching data generating process. Linex loss is frequently used in the literature on asymmetric loss, c.f. Batchelor and Peel (1998), Varian (1974), Zellner (1986) and Elliott and Timmermann (2003). Similarly, following Hamilton (1989), Markov switching processes have been widely used as a way to identify nonlinear dynamics in macroeconomic and financial time series. We show 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 thus 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 (DGP).[2]

Our results are of particular interest due to the revived interest in forecasting and econometric modeling under asymmetric loss, c.f. Christoffersen and Diebold (1996, 1997), Diebold (2004), Granger and Newbold (1986), Granger and Pesaran (2000) and Skouras (2001). When coupled with what is now common findings of nonlinearities and dynamics in higher order moments, in the form of ARCH effects for example, in most macroeconomic and financial time series this means

[1] Under asymmetric loss functions such as lin-lin and linex and assuming a conditionally Gaussian process, Christoffersen and Diebold (1997) 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.

[2] Hoque, *et al.* (1988), and Magnus and Pesaran (1989 and 1991) discuss violations of the standard properties of optimal forecasts caused by estimation error, rather than by a choice of a 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.

that the typical situation facing an economic forecaster may be well represented by our stylized example.

The insights provided in the paper also have important empirical implications. For example, it is common to test for rationality by testing for serial correlation in forecast errors.[3] Our paper shows that this implication of rationality fails to hold under even minor deviations from symmetric loss when combined with nonlinear dynamics in the DGP, and that substantial serial correlation can be expected in the forecast errors from optimal forecasts under very plausible assumptions. This suggests that many of the conclusions in the empirical literature concerning suboptimality of forecasts may be premature.

The main contribution of the paper is to provide tractable, analytical results for the multi-step forecast errors which allows us to address all of the standard properties an optimal forecast has even under asymmetric loss and nonlinearities in the DGP. Due to difficulties with time-aggregation of many nonlinear processes, such as GARCH processes, this has not previously been done. We also demonstrate, again through analytical results, that the "generalized forecast error" (which is a simple transformation of the forecast error) inherits some of the standard properties such as lack of bias and serial correlation. Finally, we provide an empirical example that demonstrates the importance of these effects on stock return data.

The outline of the paper is as follows. Section 2 shows how optimal forecast errors under asymmetric loss and nonlinearity can have a non-zero bias and a decreasing variance. Section 3 establishes violation of the standard properties on (absence of) serial correlation in forecast errors, while Section 4 shows which properties actually do hold for some transformation of the forecast error known as the generalized forecast error. Section 5 provides an empirical illustration and Section 6 concludes. An appendix contains proofs.

2 Unbiasedness and non-decreasing forecast error variance

We are interested in studying optimal $h-$period ahead forecasts of some univariate time-series process, $\{Y_\tau\}$, computed at time t conditional on some information set, $\mathcal{F}_t$. Generic forecasts are

[3] The literature on forecast rationality testing is very extensive. For a list of references to papers that use the Survey of Professional Forecasters data set maintained by the Federal Reserve Bank of Philadelphia, see http://www.phil.frb.org/econ/spf/spfbib.html.

denoted by $\hat{Y}_{t+h,t}$ and forecast errors are $e_{t+h,t} = Y_{t+h} - \hat{Y}_{t+h,t}$. In particular, the optimal $h-$period forecast computed at time t is denoted by $\hat{Y}^*_{t+h,t}$ and its forecast error is $e^*_{t+h,t}$.

2.1 Optimality Properties under mean squared error loss

Under mean squared error (MSE) loss and provided that the target variable and the predictions are joint covariance stationary, optimal forecasts satisfy the following four properties (c.f. Diebold and Lopez (1996)):

1. Forecasts are unbiased (so the forecast error ($e^*_{t+h,t}$) has zero mean).
2. The variance of the forecast error ($e^*_{t+h,t}$) is a non-decreasing function of the forecast horizon, h.
3. h-period forecast errors ($e^*_{t+h,t}$) are not correlated with information dated by h periods or more ($\mathcal{F}_t$). Hence the h-period forecast error is at most serially correlated of order $h-1$.
4. Single-period forecast errors ($e^*_{t+1,t}$) are serially uncorrelated.

These properties are easily derived under MSE loss. Suppose that Y_t has zero mean and is covariance stationary. Wold's representation theorem then establishes that it can be represented as a linear combination of serially uncorrelated white noise terms:

$$Y_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}, \tag{1}$$

where $\varepsilon_t = Y_t - E(Y_t|y_{t-1}, y_{t-2}, ...)$ is white noise, $WN(0, \sigma^2)$. The optimal $h-$period forecast thus becomes

$$\hat{Y}^*_{t+h,t} = \sum_{i=0}^{\infty} \theta_{h+i} \varepsilon_{t-i}, \tag{2}$$

with associated forecast error

$$e^*_{t+h,t} = \sum_{i=0}^{h-1} \theta_i \varepsilon_{t+h-i}. \tag{3}$$

It follows directly that

$$\begin{aligned} E[e^*_{t+h,t}] &= 0, \\ Var(e^*_{t+h,t}) &= h\sigma^2, \\ Cov(e^*_{t+h,t}, e_{t+h-j,t-j}) &= \begin{cases} \sum_{i=j}^{h-1} \theta_i \theta_{i-j} & \text{for } j < h \\ 0 & \text{for } j \geq h \end{cases}. \end{aligned} \tag{4}$$

Which account for properties 1, 2 and 3-4, respectively.

2.2 Asymmetric loss and a nonlinear process

We next provide analytical results with reasonable assumptions about the forecaster's loss function and a non-linear data generating process (DGP) which show that each of the properties (1) - (4) cease to be valid when the assumption of MSE loss is relaxed.

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

$$L\left(Y_{t+h}-\hat{Y}_{t+h,t};a\right)=\exp\left\{a(Y_{t+h}-\hat{Y}_{t+h,t})\right\}-a(Y_{t+h}-\hat{Y}_{t+h,t})-1,\ \ a\neq 0. \tag{5}$$

This loss function has been used extensively to demonstrate the effect of asymmetric loss, see e.g. Varian (1974), Zellner (1986) and Christoffersen and Diebold (1997). An optimal forecast is defined as the value that, conditional on the information set at time t, $\mathcal{F}_t$, minimizes the expected loss:

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

where $Y_{t+h}|\mathcal{F}_t$ has density $f_{t+h,t}$. Assuming 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

$$E_t\left[\frac{\partial L\left(Y_{t+h}-\hat{Y}^*_{t+h,t};a\right)}{\partial \hat{Y}_{t+h,t}}\right]=a-aE_t\left[\exp\left\{a\left(Y_{t+h}-\hat{Y}^*_{t+h,t}\right)\right\}\right]=0. \tag{6}$$

We derive analytical expressions for the optimal forecast and the expected loss using a popular nonlinear DGP, namely a regime switching model of the type proposed by Hamilton (1989).[4] Thus suppose that $\{Y_t\}$ is generated by a simple Gaussian mixture model driven by some underlying

[4] 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* standard 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).

state process, S_t:

$$\begin{aligned} Y_{t+1} &= \mu + \sigma_{s_{t+1}} v_{t+1} \\ v_{t+1} &\sim i.i.d.\ N(0,1) \\ s_{t+1} &= 1, ..., k < \infty \end{aligned} \tag{7}$$

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.[5] 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

$$\mathbf{P}(s_{t+1}|s_t) = \mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & \vdots \\ \vdots & \vdots & ... & p_{k-1k} \\ p_{k1} & \cdots & p_{kk-1} & p_{kk} \end{bmatrix}, \tag{8}$$

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. $\hat{\boldsymbol{\pi}}_{s_{t+h},t} = (\Pr(S_{t+h} = 1|\mathcal{F}_t), ..., \Pr(S_{t+h} = k|\mathcal{F}_t))'$, 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 adapted to $\mathcal{F}_t$.

2.3 Unbiasedness

We will consider an idealized h-step-ahead forecasting problem 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: neither estimation error nor irrationality is driving our results.

[5] 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.

Using the conditional normality of v_{t+h}, the expected loss is[6]

$$\begin{aligned}
E_t\left[L\left(e_{t+h,t};a\right)\right] &= E_t\left[\exp\left\{a\left(Y_{t+h}-\hat{Y}_{t+h,t}\right)\right\}\right]-aE_t\left[Y_{t+h}\right]+a\hat{Y}_{t+h,t}-1 \\
&= \sum_{s_{t+h}=1}^{k}\hat{\pi}_{s_{t+h},t}E_t\left[\exp\left\{a\left(Y_{t+h}-\hat{Y}_{t+h,t}\right)\right\}|S_{t+h}=s_{t+h}\right] \\
&\quad -a\sum_{s_{t+h}=1}^{k}\hat{\pi}_{s_{t+h},t}E_t\left[Y_{t+h}|S_{t+h}=s_{t+h}\right]+a\hat{Y}_{t+h,t}-1 \\
&= \hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\exp\left\{a\boldsymbol{\mu}-a\hat{Y}_{t+h,t}+\frac{a^2}{2}\boldsymbol{\sigma}^2\right\}-a\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\mu}+a\hat{Y}_{t+h,t}-1. \qquad (9)
\end{aligned}$$

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:

$$\hat{Y}^*_{t+h,t}=\mu+\frac{1}{a}\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right), \qquad (10)$$

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 (unbiasedness) in our setup:

Proposition 1 *The unconditional and conditional bias in the optimal forecast error arising under linex loss (5) for the Markov switching process (7)-(8) is given by:*

$$\begin{aligned}
E_t\left[e^*_{t+h,t}\right] &= -\frac{1}{a}\log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right) \qquad (11)\\
E\left[e^*_{t+h,t}\right] &= -\frac{1}{a}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h\rightarrow-\frac{1}{a}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right) \;\; \text{as } h\rightarrow\infty
\end{aligned}$$

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.

[6] 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:

$$\begin{aligned}
\boldsymbol{\mu} &= [0, 0]' \\
\boldsymbol{\sigma}^2 &= [0.25, 2]' \\
P &= \begin{bmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{bmatrix} \text{ so} \\
\bar{\boldsymbol{\pi}} &= \left[\frac{2}{3}, \frac{1}{3}\right]'.
\end{aligned}$$

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 empirical results frequently obtained when this model is estimated on financial data, see Section 5 for example. For this particular parameterization the optimal unconditional bias in $e^*_{t+1,t}$ is -1.17, indicating that it is optimal to over-predict. Figure 1 shows the unconditional 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 for symmetric distributions the amount of probability mass either side of the forecast would be 50%. In Figure 2 we plot the optimal unconditional bias as a function of the forecast horizon (using the steady-state weights as initial probabilities). The unconditional bias for this case is an increasing (in absolute value) function of h and asymptotes to -1.17.

2.4 Non-decreasing variance

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

Proposition 2 *The variance of the forecast error arising under linex loss (5) for the Markov switching process (7)-(8) associated with the optimum forecast is given by*

$$Var\left(e^*_{t+h,t}\right) = \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right) \odot \mathbf{I} - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\lambda}_h. \tag{12}$$

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, $\bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2$.

Conversely, under MSE loss, the unconditional forecast error variance is constant across forecast horizons, h:

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

It is also common to consider the mean squared error of the forecast. This is closely related to the forecast error variance but differs by the squared bias. The corresponding result for the MSFE is as follows:

Corollary 1 *The mean-square forecast error arising under linex loss (5) for the Markov switching process (7)-(8) associated with the optimum forecast is given by*

$$MSFE\left(e^*_{t+h,t}\right) = \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 + \frac{1}{a^2}\boldsymbol{\lambda}'_h\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot\mathbf{I}\right)\boldsymbol{\lambda}_h \tag{13}$$

The MSFE need not be a decreasing function of the forecast horizon, h. In the limit as h goes to infinity, the MSFE 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 2 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.[7] The reason for this surprising result lies in the mismatch of the forecast objective function, L, and the variance of the forecast error, $Var(e_{t+h,t})$, which does not occur when using quadratic loss (see next section).

[7] Using the same numerical example it can be shown that the MSFE 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.

A similar mismatch of the objective function and the performance metric has been discussed by Christoffersen and Jacobs (2004), 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 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 $\boldsymbol{\lambda}_h = \boldsymbol{\iota}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right)$ 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.

3 Serial Correlation in 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. This particular property of an optimal forecast is usually tested by regressing the observed forecast error on its own lagged values of order h and higher:

$$e_{t+h,t} = \alpha_0 + \sum_{i=0}^{n}\beta_i e_{t-i,t-i-h} + u_{t+h,t}.$$

Here it is assumed that $u_{t+h,t}$ has mean zero and is serially uncorrelated. Under MSE loss and forecast rationality, $\alpha_0 = \beta_0 = ... = \beta_n = 0$.

Outside the standard setting with MSE loss this property need no longer hold. In particular, we have

Proposition 3 *The h-step-ahead forecast error arising under linex loss (5) for the Markov switching process (7)-(8) is serially correlated with autocovariance*

$$Cov\left[e^*_{t+h,t}, e^*_{t+h-j,t-j}\right] = \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2\mathbf{1}_{\{j=0\}} + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot P^j - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\lambda}_h. \tag{14}$$

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).[8]

It is easily verified that, under MSE loss, the autocovariance of the optimal forecast errors under the regime switching process is zero:

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

Thus optimal forecast errors under MSE loss 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.

4 Properties of a "generalized forecast error"

We demonstrated in the previous sections that *all* the properties of optimal forecasts established under MSE loss can be violated under asymmetric loss and a nonlinear DGP. In this section we establish the properties that a simple transformation of the optimal forecast error–known as the generalized forecast error–must have.

[8] 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.

Following Granger (1999) and Patton and Timmermann (2004), the generalized forecast error, $\psi^*_{t+h,t}$, is defined from the first-order condition of the loss function, (6)

$$\psi^*_{t+h,t} \equiv \frac{\partial L\left(Y_{t+h} - \hat{Y}^*_{t+h,t}; a\right)}{\partial \hat{Y}_{t+h,t}} \tag{15}$$

which in our case simplifies to

$$\psi^*_{t+h,t} = a - a\exp\left\{a\sigma_{s_{t+h}}\nu_{t+h} - \log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}\mathbf{P}^h\boldsymbol{\varphi}\right)\right\}. \tag{16}$$

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

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

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.

Turning to the second optimality property, while we saw that the variance of the optimal forecast error need not be non-decreasing with the forecast horizon, thus violating the second standard optimality property, the unconditional expected loss will always be non-decreasing in the forecast horizon:

Proposition 4 *Under linex loss (5) and the regime switching process (7)-(8) the unconditional expected loss is*

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

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.

For the third property, we saw that in a non-linear framework, optimal $h-$step forecast errors need not follow an $MA(h-1)$ process. However, the generalized forecast error will display serial correlation of, at most, $(h-1)^{th}$ order:

Proposition 5 *The generalized forecast error from an optimal h-step forecast made at time t under the regime switching process (7)-(8) and assuming linex loss (5) has the following autocovariance function:*

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

where $\boldsymbol{\varphi}^4\equiv\exp\left\{2a^2\boldsymbol{\sigma}^2\right\}$.

Using the numerical example above, Figure 6 presents the autocorrelation function for the optimal generalized forecast error. As implied by Proposition 5, the autocorrelations are non-zero for $j<h$ and drop sharply to zero when $j\geq h$.

5 Empirical illustration

In this section we present an illustration of the above results for actual financial data. The intention and results of this empirical illustration is to show that the sample model employed in the paper was a reasonable case to consider. Further, they emphasize that using the standard properties of optimal forecast errors (unbiasedness, increasing variance, restricted serial correlation) in the presence of asymmetric loss and time-varying volatility could lead to the rejection of a perfectly optimal forecast.

Using maximum likelihood methods we estimated the two-state regime switching model discussed above on weekly returns for Exxon over the period January 1970 to December 2003, and obtained the following parameter estimates:

$$\begin{aligned}\hat{\boldsymbol{\mu}}&=[0.1323,0.1323]'\\ \hat{\boldsymbol{\sigma}}&=[0.9698,1.6711]'\\ \hat{P}&=\begin{bmatrix}0.9756&0.0244\\0.1014&0.8986\end{bmatrix}.\end{aligned}$$

Thus the annualized return on Exxon over this period was 6.9%. The steady state probabilities implied by the estimated transition matrix $\widehat{\bar{\boldsymbol{\pi}}} = [0.8061, 1939]'$ indicate that returns were in the low volatility regime four times more often than in the high volatility regime.

If we assume a linex loss function for the forecast user, and again assume that the parameter of this function is set to $a = 1$, then Proposition 1 shows that the unconditional mean of the optimal Exxon return forecast errors is -0.7292. Given the asymmetry of the loss function it is not surprising that it is optimal to over-predict, leading to forecast errors that are negative on average.

Using the expression given in Proposition 2 we obtain the unconditional variance of the optimal forecast errors for different horizons, and plot these in Figure 7. We again see, under linex loss and a regime switching model for Exxon stock returns, that the unconditional variance is a decreasing function of the forecast horizon. The MSFE of the optimal forecast errors (not shown) is similarly monotonically downward sloping. Taking the standard properties of optimal forecasts as a guide, this figure would suggest that it is about 7% *easier* (variance of 1.31 versus 1.41) to predict the one-week return on Exxon in the period between nine and ten weeks from today, than the return on Exxon over the coming week. As we discussed above, this interpretation is misleading when the forecast user's loss function is not the MSE loss function. If we instead use expected loss to measure forecast difficulty, we find that it is about 8% *harder* (expected loss of 0.72 versus 0.67) to forecast the 10-step ahead one-week return than it is to forecast the return over the coming week.

Using Proposition 3 we obtain the autocorrelation function of the optimal forecast errors in this illustration, and plot this in Figure 8 for $h = 1, .., 5$. The one-step ahead weekly return forecast, for example, has first-order serial correlation of almost 0.07, and even at the twentieth lag the serial correlation is still greater than zero. This implies that the forecast error is predictable using lagged forecast errors, albeit only weakly in this case. It does *not* imply that the forecast is suboptimal, however: the generalized forecast error exhibits the serial correlation properties presented in Proposition 5.

6 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, and are not generally robust to even slight departures from these assumptions. We showed analytically how they are violated under more general assumptions about the loss function, extending the seminal 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. A small illustration using weekly stock return data confirmed that our findings have much empirical relevance. Thus, although our results were confined to a particular loss function and data generating process, they are powerful as they rely only on empirically reasonable assumptions.

A number of implications follow from this paper. Most importantly, our results suggest the need to develop new and more general methods for forecast evaluation that are robust to deviations from mean squared error loss. Most economic time series have some dynamics in the conditional variance. This means that even small deviations from mean squared error loss will overturn the standard properties that an optimal forecast must have. One approach−proposed by Elliott, Komunjer and Timmermann (2002)−is to assume that the loss function belongs to a family of parsimoniously parameterized functions that nests mean squared error loss as a special case. This approach can be used even with the small sample sizes typically obtainable with forecasting data. Furthermore, the parameters of the loss function can be estimated by GMM and forecast optimality can be tested by means of a standard J-test when the model is overidentified.

In situations where even the form of the loss function is unknown, a less parametric approach is required. For this situation Patton and Timmermann (2004) show that if the conditional dynamics of the data generating process can be restricted to the first and second moments, then it is possible to develop tests for forecast optimality that hold for a broad range of loss functions and do not require estimating the shape of the (unknown) loss function.

References

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

[2] Christoffersen, P. and K. Jacobs, 2004, The Importance of the Loss Function in Option Valuation. Journal of Financial Economics, 72, 291-318.

[3] 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.

[4] Christoffersen, P.F. and F.X. Diebold, 1997, Optimal Prediction under Asymmetric Loss. Econometric Theory 13, 808-817.

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

[6] Diebold, F.X., 2004, Elements of Forecasting (3rd edition). Southwestern.

[7] Diebold, F.X. and J. Lopez, 1996, Forecast Evaluation and Combination, in G.S. Maddala and C.R. Rao (eds.), Handbook of Statistics, Amsterdam: North-Holland, 241-268.

[8] Drost, F.C., and T.E. Nijman, 1993, Temporal Aggregation of GARCH Processes, Econometrica, 61, 909-928.

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

[10] Elliott, G. and A. Timmermann, 2003, Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions. Forthcoming, Journal of Econometrics.

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

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

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

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

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

[16] 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.

[17] 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.

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

[19] Patton, A.J. and A. Timmermann, 2004, Testable Implications of Forecast Optimality, work-in-progress.

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

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

[22] 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.

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

Appendix

Proof of Proposition 1. 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 one 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.

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 2. From Proposition 2 we have

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

$$
\begin{aligned}
&= \sum_{s_{t+h}=1}^{k} \bar{\pi}_{(s_{t+h})}\sigma^2_{s_{t+h}} - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&\quad + \frac{1}{a^2}\sum_{s_t=1}^{k} \bar{\pi}_{(s_t)} \log\left(\boldsymbol{\varphi}'\mathbf{P}^{h\prime}\right)\boldsymbol{\iota}_{s_t}\boldsymbol{\iota}_{s_t}' \log\left(\mathbf{P}^h\boldsymbol{\varphi}\right) \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\sum_{s_t=1}^{k}\bar{\boldsymbol{\pi}}_{(s_t)}\boldsymbol{\iota}_{s_t}\boldsymbol{\iota}_{s_t}'\right)\boldsymbol{\lambda}_h - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot I - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\lambda}_h.
\end{aligned}
$$

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

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

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

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

So, 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 2. ■

Proof of Proposition 3. The autocovariance function for an h-step forecast is:

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

$$
\begin{aligned}
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 1_{\{j=0\}} - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&\quad + \frac{1}{a^2}\sum_{s_{t-j}=1}^{k}\sum_{s_t=1}^{k}\bar{\pi}_{(s_{t-j})}\pi_{(s_t|s_{t-j})}\log\left(\boldsymbol{\iota}_{s_t}'\mathbf{P}^h\boldsymbol{\varphi}\right)\log\left(\boldsymbol{\iota}_{s_{t-j}}'\mathbf{P}^h\boldsymbol{\varphi}\right) \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 1_{\{j=0\}} - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&\quad + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\sum_{s_{t-j}=1}^{k}\bar{\pi}_{(s_{t-j})}\left(\sum_{s_t=1}^{k}\pi_{(s_t|s_{t-j})}\boldsymbol{\iota}_{s_t}\right)\boldsymbol{\iota}_{s_{t-j}}'\right)\boldsymbol{\lambda}_h \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 1_{\{j=0\}} - \frac{1}{a^2}\boldsymbol{\lambda}_h'\bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\boldsymbol{\lambda}_h \\
&\quad + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\sum_{s_{t-j}=1}^{k}\bar{\pi}_{(s_{t-j})}\mathbf{P}^{j\prime}\boldsymbol{\iota}_{s_{t-j}}\boldsymbol{\iota}_{s_{t-j}}'\right)\boldsymbol{\lambda}_h \\
&= \bar{\boldsymbol{\pi}}'\boldsymbol{\sigma}^2 1_{\{j=0\}} + \frac{1}{a^2}\boldsymbol{\lambda}_h'\left(\left(\bar{\boldsymbol{\pi}}\boldsymbol{\iota}'\right)\odot\mathbf{P}^j - \bar{\boldsymbol{\pi}}\bar{\boldsymbol{\pi}}'\right)\boldsymbol{\lambda}_h
\end{aligned}
$$

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

Proof of Proposition 4. $E\left[L\left(Y_{t+h},\hat{Y}_{t+h,t}^{*}\right)\right]$

$$
\begin{aligned}
&= E\left[\exp\left\{a\sigma_{s_{t+h}}\nu_{t+h} - \log\left(\hat{\boldsymbol{\pi}}_{s_t,t}'\mathbf{P}^h\boldsymbol{\varphi}\right)\right\}\right] + E\left[\log\left(\hat{\boldsymbol{\pi}}_{s_t,t}'\mathbf{P}^h\boldsymbol{\varphi}\right)\right] - 1 \\
&= \sum_{s_t=1}^{k}\sum_{s_{t+h}=1}^{k}\bar{\pi}_{(s_t)}\pi_{(s_{t+h}|s_t)}E\left[\exp\left\{a\sigma_{s_{t+h}}\nu_{t+h} - \log\left(\hat{\boldsymbol{\pi}}_{s_t,t}'\mathbf{P}^h\boldsymbol{\varphi}\right)\right\}|S_{t+h},S_t\right] \\
&\quad + \sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}E\left[\log\left(\hat{\boldsymbol{\pi}}_{s_t,t}'\mathbf{P}^h\boldsymbol{\varphi}\right)|S_t\right] - 1 \\
&= \sum_{s_t=1}^{k}\sum_{s_{t+h}=1}^{k}\bar{\pi}_{(s_t)}\pi_{(s_{t+h}|s_t)}\exp\left\{-\log\left(\boldsymbol{\iota}_{s_t}'\mathbf{P}^h\boldsymbol{\varphi}\right) + \frac{a^2}{2}\sigma_{s_{t+h}}^2\right\} + \sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\boldsymbol{\iota}_{s_t}'\log\left(\mathbf{P}^h\boldsymbol{\varphi}\right) - 1 \\
&= \sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\boldsymbol{\iota}_{s_t}'\left(\mathbf{P}^h\boldsymbol{\varphi}\right)^{-1}\sum_{s_{t+h}=1}^{k}\pi_{(s_{t+h}|s_t)}\exp\left\{\frac{a^2}{2}\sigma_{s_{t+h}}^2\right\} + \sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\boldsymbol{\iota}_{s_t}'\log\left(\mathbf{P}^h\boldsymbol{\varphi}\right) - 1 \\
&= \sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\boldsymbol{\iota}_{s_t}'\left(\mathbf{P}^h\boldsymbol{\varphi}\right)^{-1}\left(\mathbf{P}^h\boldsymbol{\varphi}\right) + \sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\boldsymbol{\iota}_{s_t}'\log\left(\mathbf{P}^h\boldsymbol{\varphi}\right) - 1 \\
&= \sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\boldsymbol{\iota}_{s_t}'\log\left(\mathbf{P}^h\boldsymbol{\varphi}\right) \\
&= \bar{\boldsymbol{\pi}}'\log\left(\mathbf{P}^h\boldsymbol{\varphi}\right)
\end{aligned}
$$

As $h\to\infty$ the expected loss becomes: $E\left[L\left(Y_{t+h},\hat{Y}_{t+h,t}^{*};a\right)\right] = \bar{\boldsymbol{\pi}}'\log\left(\mathbf{P}^h\boldsymbol{\varphi}\right)\to\bar{\boldsymbol{\pi}}'\log\left(\boldsymbol{\iota}\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right) = \bar{\boldsymbol{\pi}}'\boldsymbol{\iota}\log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right) = \log\left(\bar{\boldsymbol{\pi}}'\boldsymbol{\varphi}\right).$ ■

Proof of Proposition 5. $Cov\left[\psi^*_{t+h,t},\psi^*_{t+h-j,t-j}\right]$

$$\begin{aligned}
&= a^2 - a^2 E\left[\exp\left\{a\sigma_{s_{t+h}}\nu_{t+h} - \log\left(\hat{\pi}'_{s_t,t}P^h\varphi\right)\right\}\right] \\
&\quad -a^2 E\left[\exp\left\{a\sigma_{s_{t+h-j}}\nu_{t+h-j} - \log\left(\hat{\pi}'_{s_{t-j},t-j}P^h\varphi\right)\right\}\right] \\
&\quad +a^2 E\left[\exp\left\{a\sigma_{s_{t+h}}\nu_{t+h} + a\sigma_{s_{t+h-j}}\nu_{t+h-j} - \log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}P^h\varphi\right) - \log\left(\hat{\boldsymbol{\pi}}'_{s_{t-j},t-j}P^h\varphi\right)\right\}\right] \\
&= aE\left[\psi^*_{t+h,t}\right] + aE\left[\psi^*_{t+h-j,t-j}\right] - a^2 \\
&\quad +a^2 E\left[\exp\left\{a\sigma_{s_{t+h}}\nu_{t+h} + a\sigma_{s_{t+h-j}}\nu_{t+h-j} - \log\left(\hat{\boldsymbol{\pi}}'_{s_t,t}P^h\varphi\right) - \log\left(\hat{\boldsymbol{\pi}}'_{s_{t-j},t-j}P^h\varphi\right)\right\}\right] \\
&= -a^2 + a^2 E\left[\exp\left\{a\sigma_{s_{t+h}}\nu_{t+h} + a\sigma_{s_{t+h-j}}\nu_{t+h-j} - \log\left(\hat{\pi}'_{s_t,t}P^h\varphi\right) - \log\left(\hat{\pi}'_{s_{t-j},t-j}P^h\varphi\right)\right\}\right]
\end{aligned}$$

If $j=0$ we get the variance of the generalized forecast error: $V\left[\psi^*_{t+h,t}\right]$

$$\begin{aligned}
&= -a^2 + a^2\sum_{s_t=1}^{k}\sum_{s_{t+h}=1}^{k}\bar{\pi}_{(s_t)}\pi_{(s_{t+h},t)}\exp\left\{-2\log\left(\boldsymbol{\iota}'_{s_t}P^h\boldsymbol{\varphi}\right) + 2a^2\sigma^2_{s_{t+h}}\right\} \\
&= -a^2 + a^2\sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\left(\boldsymbol{\iota}'_{s_t}P^h\boldsymbol{\varphi}\right)^{-2}\sum_{s_{t+h}=1}^{k}\pi_{(s_{t+h},t)}\exp\left\{2a^2\sigma^2_{s_{t+h}}\right\} \\
&= -a^2 + a^2\sum_{s_t=1}^{k}\bar{\pi}_{(s_t)}\left(\boldsymbol{\iota}'_{s_t}P^h\boldsymbol{\varphi}\right)^{-2}\left(\boldsymbol{\iota}'_{s_t}P^h\boldsymbol{\varphi}^4\right)
\end{aligned}$$

where $\boldsymbol{\varphi}^4 \equiv \exp\left\{2a^2\boldsymbol{\sigma}^2\right\} = \boldsymbol{\varphi}\odot\boldsymbol{\varphi}\odot\boldsymbol{\varphi}\odot\boldsymbol{\varphi}$

For $0<j<h$ we get: $Cov\left[\psi^*_{t+h,t},\psi^*_{t+h-j,t-j}\right]$

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

$$
\begin{aligned}
&= -a^2 + a^2 \sum_{s_{t-j}=1}^{k} \bar{\pi}_{(s_{t-j})} \left(\boldsymbol{\iota}'_{s_{t-j}} P^h \boldsymbol{\varphi}\right)^{-1} \sum_{s_t=1}^{k} \pi_{s_t,t-j} \left(\boldsymbol{\iota}'_{s_t} P^h \boldsymbol{\varphi}\right)^{-1} \cdot \\
&\quad \sum_{s_{t+h-j}=1}^{k} \pi_{\left(s_{t+h-j},t\right)} \exp\left\{\frac{a^2}{2}\sigma^2_{s_{t+h-j}}\right\} \sum_{s_{t+h}=1}^{k} \pi_{(s_{t+h},t+h-j)} \exp\left\{\frac{a^2}{2}\sigma^2_{s_{t+h}}\right\} \\
&= -a^2 + a^2 \sum_{s_{t-j}=1}^{k} \bar{\pi}_{(s_{t-j})} \left(\boldsymbol{\iota}'_{s_{t-j}} P^h \boldsymbol{\varphi}\right)^{-1} \sum_{s_t=1}^{k} \pi_{s_t,t-j} \left(\boldsymbol{\iota}'_{s_t} P^h \boldsymbol{\varphi}\right)^{-1} \cdot \\
&\quad \sum_{s_{t+h-j}=1}^{k} \pi_{\left(s_{t+h-j},t\right)} \exp\left\{\frac{a^2}{2}\sigma^2_{s_{t+h-j}}\right\} \boldsymbol{\iota}'_{s_{t+h-j}} P^j \varphi \\
&= -a^2 + a^2 \sum_{s_{t-j}=1}^{k} \bar{\pi}_{(s_{t-j})} \left(\boldsymbol{\iota}'_{s_{t-j}} P^h \boldsymbol{\varphi}\right)^{-1} \cdot \sum_{s_t=1}^{k} \pi_{(s_t,t-j)} \left(\boldsymbol{\iota}'_{s_t} P^h \boldsymbol{\varphi}\right)^{-1} \cdot \left(\boldsymbol{\varphi}' \odot \left(\boldsymbol{\iota}'_{s_t} P^{h-j}\right)\right) P^j \boldsymbol{\varphi}
\end{aligned}
$$

And for $j \geq h$ we get: $Cov\left[\psi^*_{t+h,t}, \psi^*_{t+h-j,t-j}\right]$

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

$$
\begin{aligned}
&= -a^2 + a^2 \sum_{s_{t-j}=1}^{k} \bar{\pi}_{(s_{t-j})} \left(\boldsymbol{\iota}'_{s_{t-j}} P^h \boldsymbol{\varphi}\right)^{-1} \cdot \sum_{s_{t+h-j}=1}^{k} \pi_{\left(s_{t+h-j},t-j\right)} \cdot \exp\left\{\frac{a^2}{2}\sigma^2_{s_{t+h-j}}\right\} \cdot \\
&\quad \sum_{s_t=1}^{k} \pi_{(s_t,t+h-j)} \left(\boldsymbol{\iota}'_{s_t} P^h \boldsymbol{\varphi}\right)^{-1} \sum_{s_{t+h}=1}^{k} \pi_{s_{t+h},t} \exp\left\{\frac{a^2}{2}\sigma^2_{s_{t+h}}\right\} \\
&= -a^2 + a^2 \sum_{s_{t-j}=1}^{k} \bar{\pi}_{(s_{t-j})} \left(\boldsymbol{\iota}'_{s_{t-j}} P^h \boldsymbol{\varphi}\right)^{-1} \cdot \sum_{s_{t+h-j}=1}^{k} \pi_{\left(s_{t+h-j},t-j\right)} \cdot \exp\left\{\frac{a^2}{2}\sigma^2_{s_{t+h-j}}\right\} \\
&= -a^2 + a^2 \sum_{s_{t-j}=1}^{k} \bar{\pi}_{(s_{t-j})} \left(\boldsymbol{\iota}'_{s_{t-j}} P^h \boldsymbol{\varphi}\right)^{-1} \left(\boldsymbol{\iota}'_{s_{t-j}} P^h \boldsymbol{\varphi}\right) = 0
\end{aligned}
$$

The autocovariance is zero for all lags greater than or equal to the forecast horizon. ■

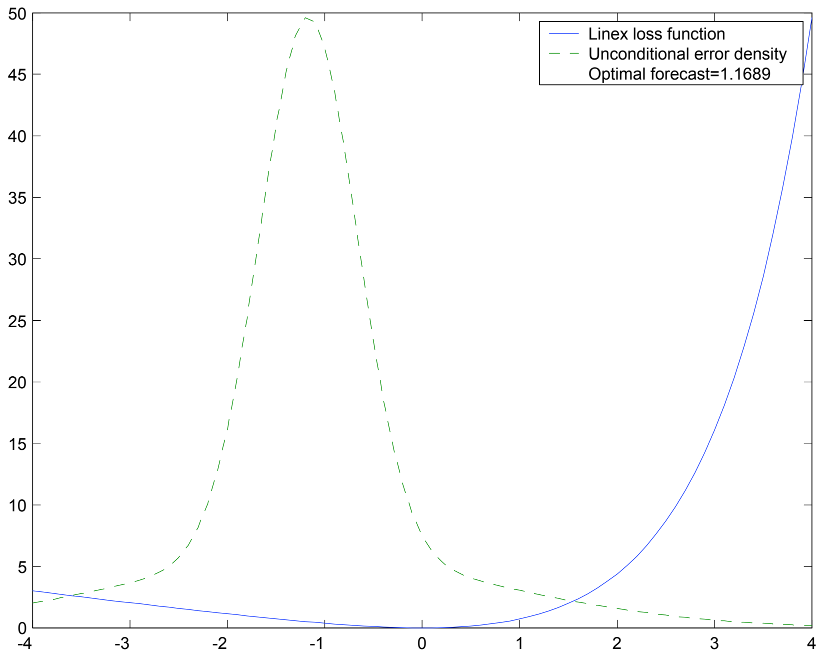

Figure 1: *Linear-exponential loss function and unconditional optimal forecast error density, two-state regime switching example.*

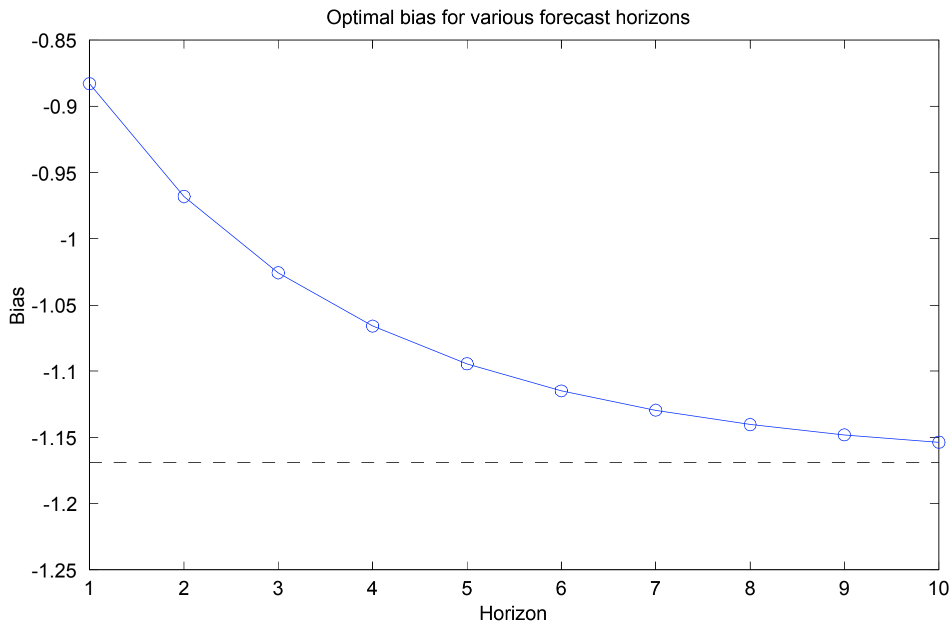

Figure 2: *Bias in the optimal forecast for various forecast horizons, two-state regime switching example.*

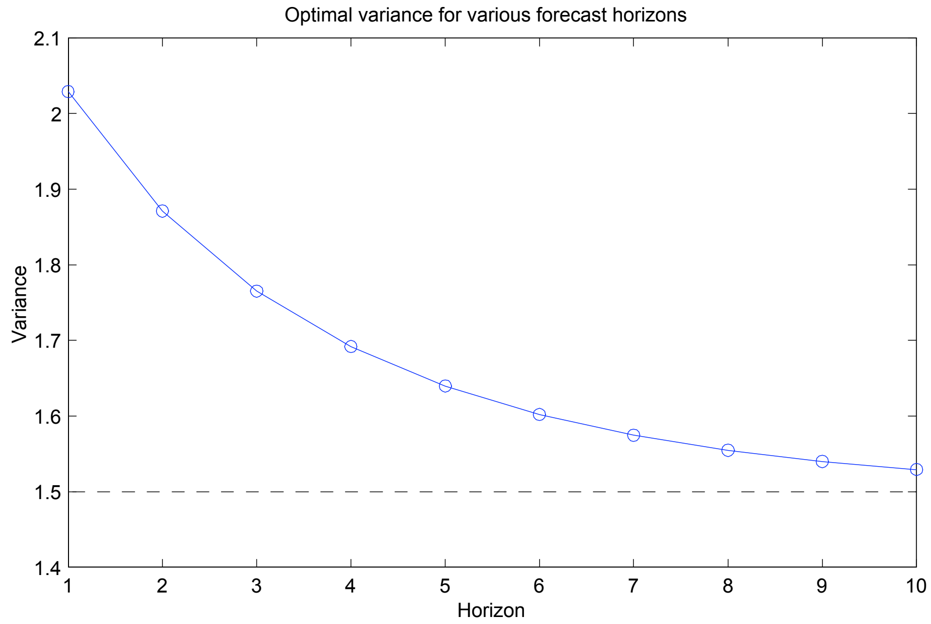

Figure 3: *Variance of the optimal h-step forecast error for various forecast horizons, two-state regime switching example.*

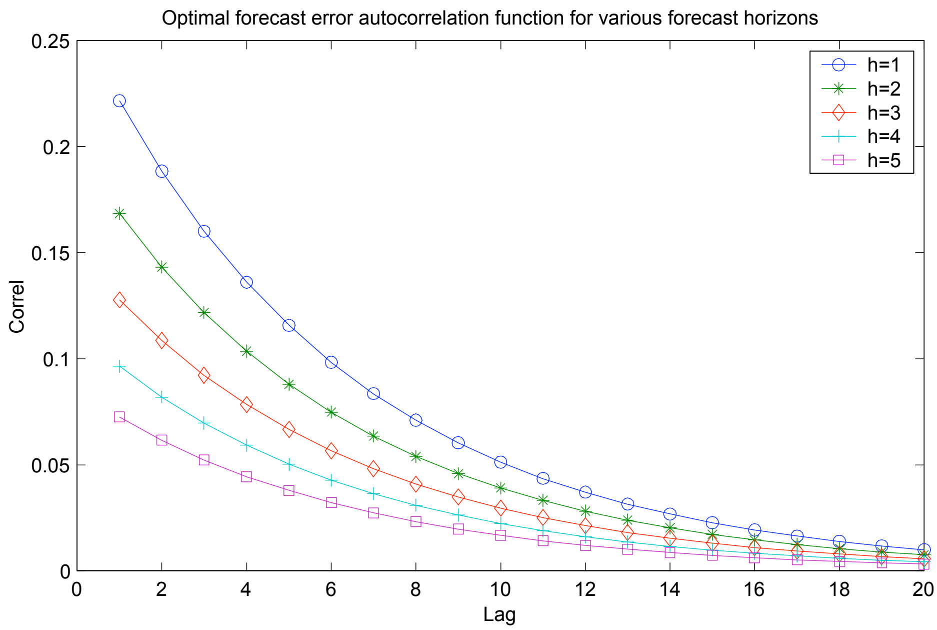

Figure 4: *Autocorrelation in the optimal h-step forecast error for various forecast horizons, two-state regime switching example.*

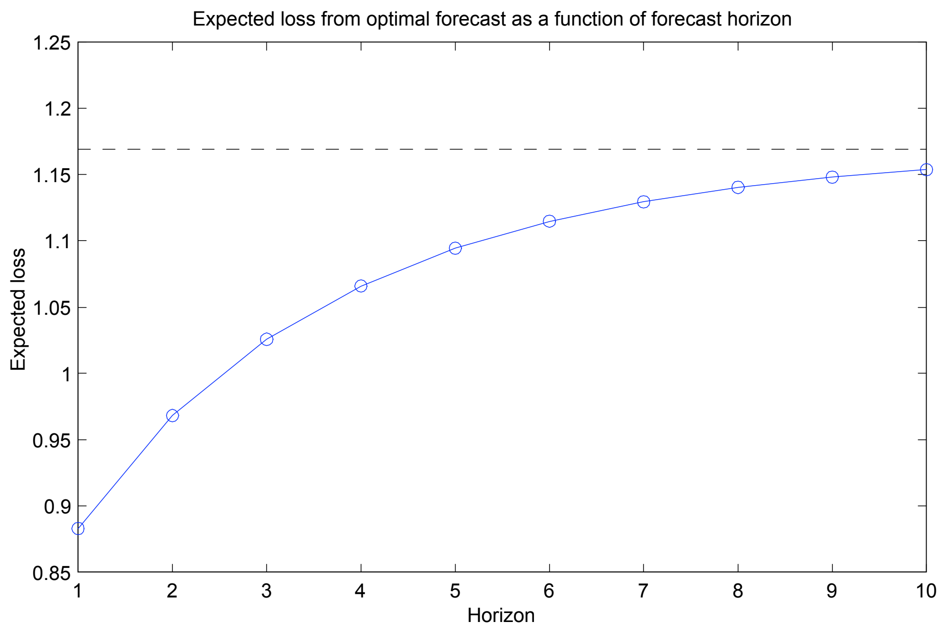

Figure 5: *Expected loss from the optimal forecast for various forecast horizons, two-state regime switching example.*

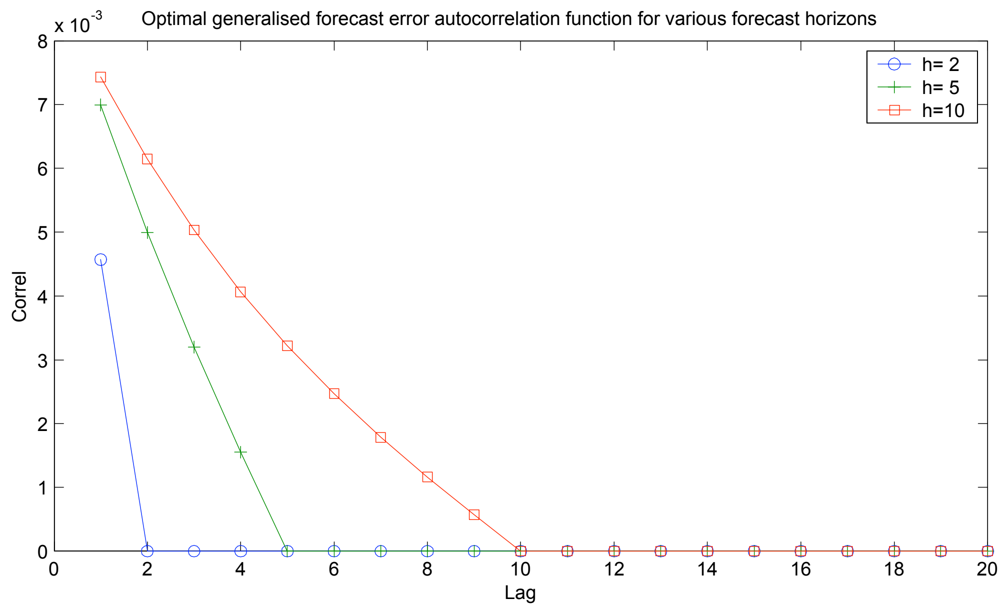

Figure 6: *Autocorrelation in the optimal generalised forecast error for various forecast horizons, two-state regime switching example.*

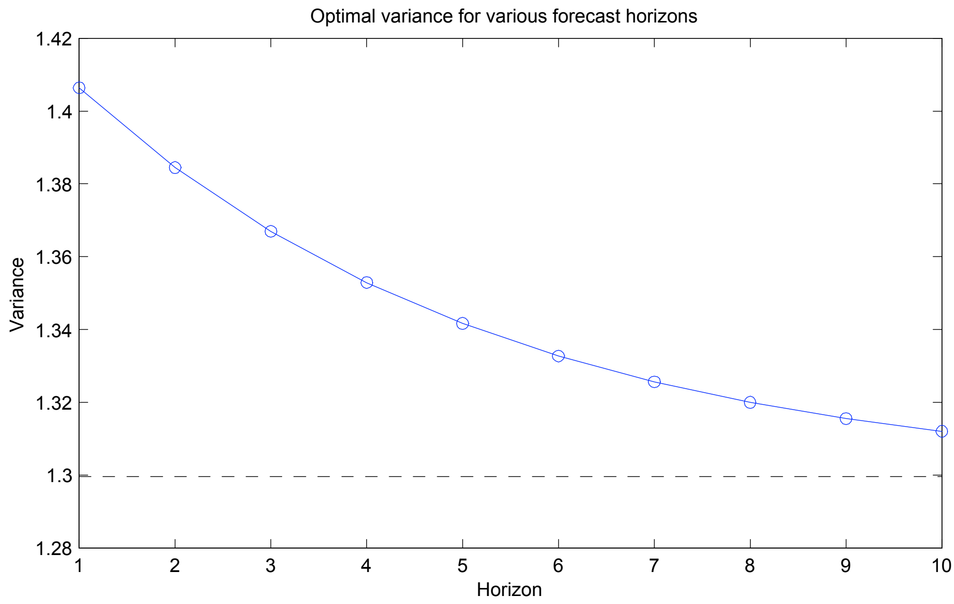

Figure 7: *Variance of the optimal h-step forecast error for various forecast horizons, Exxon stock returns, Jan 1970 to Dec 2003.*

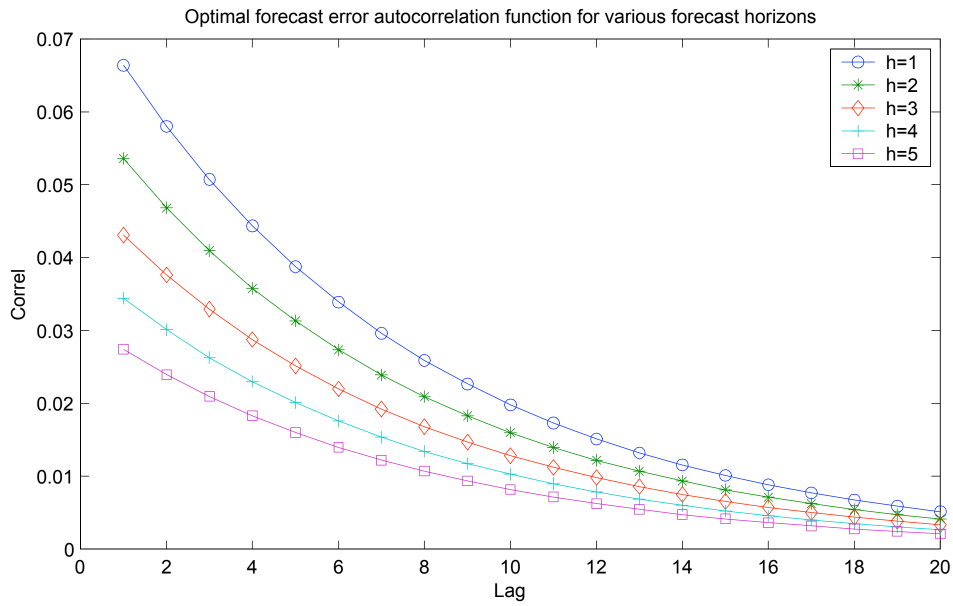

Figure 8: *Autocorrelation in the optimal h-step forecast error for various forecast horizons, Exxon stock returns, Jan 1970 to Dec 2003.*

List of other working papers:

2004

1. Xiaohong Chen, Yanqin Fan and Andrew Patton, Simple Tests for Models of Dependence Between Multiple Financial Time Series, with Applications to U.S. Equity Returns and Exchange Rates, WP04-19
2. Valentina Corradi and Walter Distaso, Testing for One-Factor Models versus Stochastic Volatility Models, WP04-18
3. Valentina Corradi and Walter Distaso, Estimating and Testing Sochastic Volatility Models using Realized Measures, WP04-17
4. Valentina Corradi and Norman Swanson, Predictive Density Accuracy Tests, WP04-16
5. Roel Oomen, Properties of Bias Corrected Realized Variance Under Alternative Sampling Schemes, WP04-15
6. Roel Oomen, Properties of Realized Variance for a Pure Jump Process: Calendar Time Sampling versus Business Time Sampling, WP04-14
7. Richard Clarida, Lucio Sarno, Mark Taylor and Giorgio Valente, The Role of Asymmetries and Regime Shifts in the Term Structure of Interest Rates, WP04-13
8. Lucio Sarno, Daniel Thornton and Giorgio Valente, Federal Funds Rate Prediction, WP04-12
9. Lucio Sarno and Giorgio Valente, Modeling and Forecasting Stock Returns: Exploiting the Futures Market, Regime Shifts and International Spillovers, WP04-11
10. Lucio Sarno and Giorgio Valente, Empirical Exchange Rate Models and Currency Risk: Some Evidence from Density Forecasts, WP04-10
11. Ilias Tsiakas, Periodic Stochastic Volatility and Fat Tails, WP04-09
12. Ilias Tsiakas, Is Seasonal Heteroscedasticity Real? An International Perspective, WP04-08
13. Damin Challet, Andrea De Martino, Matteo Marsili and Isaac Castillo, Minority games with finite score memory, WP04-07
14. Basel Awartani, Valentina Corradi and Walter Distaso, Testing and Modelling Market Microstructure Effects with an Application to the Dow Jones Industrial Average, WP04-06
15. Andrew Patton and Allan Timmermann, Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity, WP04-05
16. Andrew Patton, Modelling Asymmetric Exchange Rate Dependence, WP04-04
17. Alessio Sancetta, Decoupling and Convergence to Independence with Applications to Functional Limit Theorems, WP04-03
18. Alessio Sancetta, Copula Based Monte Carlo Integration in Financial Problems, WP04-02
19. Abhay Abhayankar, Lucio Sarno and Giorgio Valente, Exchange Rates and Fundamentals: Evidence on the Economic Value of Predictability, WP04-01

2002

1. Paolo Zaffaroni, Gaussian inference on Certain Long-Range Dependent Volatility Models, WP02-12
2. Paolo Zaffaroni, Aggregation and Memory of Models of Changing Volatility, WP02-11
3. Jerry Coakley, Ana-Maria Fuertes and Andrew Wood, Reinterpreting the Real Exchange Rate - Yield Diffential Nexus, WP02-10
4. Gordon Gemmill and Dylan Thomas , Noise Training, Costly Arbitrage and Asset Prices: evidence from closed-end funds, WP02-09
5. Gordon Gemmill, Testing Merton's Model for Credit Spreads on Zero-Coupon Bonds, WP02-08
6. George Christodoulakis and Steve Satchell, On th Evolution of Global Style Factors in the MSCI Universe of Assets, WP02-07
7. George Christodoulakis, Sharp Style Analysis in the MSCI Sector Portfolios: A Monte Caro Integration Approach, WP02-06

8. George Christodoulakis, Generating Composite Volatility Forecasts with Random Factor Betas, WP02-05
9. Claudia Riveiro and Nick Webber, Valuing Path Dependent Options in the Variance-Gamma Model by Monte Carlo with a Gamma Bridge, WP02-04
10. Christian Pedersen and Soosung Hwang, On Empirical Risk Measurement with Asymmetric Returns Data, WP02-03
11. Roy Batchelor and Ismail Orgakcioglu, Event-related GARCH: the impact of stock dividends in Turkey, WP02-02
12. George Albanis and Roy Batchelor, Combining Heterogeneous Classifiers for Stock Selection, WP02-01

2001

1. Soosung Hwang and Steve Satchell , GARCH Model with Cross-sectional Volatility; GARCHX Models, WP01-16
2. Soosung Hwang and Steve Satchell, Tracking Error: Ex-Ante versus Ex-Post Measures, WP01-15
3. Soosung Hwang and Steve Satchell, The Asset Allocation Decision in a Loss Aversion World, WP01-14
4. Soosung Hwang and Mark Salmon, An Analysis of Performance Measures Using Copulae, WP01-13
5. Soosung Hwang and Mark Salmon, A New Measure of Herding and Empirical Evidence, WP01-12
6. Richard Lewin and Steve Satchell, The Derivation of New Model of Equity Duration, WP01-11
7. Massimiliano Marcellino and Mark Salmon, Robust Decision Theory and the Lucas Critique, WP01-10
8. Jerry Coakley, Ana-Maria Fuertes and Maria-Teresa Perez, Numerical Issues in Threshold Autoregressive Modelling of Time Series, WP01-09
9. Jerry Coakley, Ana-Maria Fuertes and Ron Smith, Small Sample Properties of Panel Time-series Estimators with I(1) Errors, WP01-08
10. Jerry Coakley and Ana-Maria Fuertes, The Felsdtein-Horioka Puzzle is Not as Bad as You Think, WP01-07
11. Jerry Coakley and Ana-Maria Fuertes, Rethinking the Forward Premium Puzzle in a Non-linear Framework, WP01-06
12. George Christodoulakis, Co-Volatility and Correlation Clustering: A Multivariate Correlated ARCH Framework, WP01-05
13. Frank Critchley, Paul Marriott and Mark Salmon, On Preferred Point Geometry in Statistics, WP01-04
14. Eric Bouyé and Nicolas Gaussel and Mark Salmon, Investigating Dynamic Dependence Using Copulae, WP01-03
15. Eric Bouyé, Multivariate Extremes at Work for Portfolio Risk Measurement, WP01-02
16. Erick Bouyé, Vado Durrleman, Ashkan Nikeghbali, Gael Riboulet and Thierry Roncalli, Copulas: an Open Field for Risk Management, WP01-01

2000

1. Soosung Hwang and Steve Satchell , Valuing Information Using Utility Functions, WP00-06
2. Soosung Hwang, Properties of Cross-sectional Volatility, WP00-05
3. Soosung Hwang and Steve Satchell, Calculating the Miss-specification in Beta from Using a Proxy for the Market Portfolio, WP00-04
4. Laun Middleton and Stephen Satchell, Deriving the APT when the Number of Factors is Unknown, WP00-03
5. George A. Christodoulakis and Steve Satchell, Evolving Systems of Financial Returns: Auto-Regressive Conditional Beta, WP00-02
6. Christian S. Pedersen and Stephen Satchell, Evaluating the Performance of Nearest Neighbour Algorithms when Forecasting US Industry Returns, WP00-01

1999

1. Yin-Wong Cheung, Menzie Chinn and Ian Marsh, How do UK-Based Foreign Exchange Dealers Think Their Market Operates?, WP99-21
2. Soosung Hwang, John Knight and Stephen Satchell, Forecasting Volatility using LINEX Loss Functions, WP99-20
3. Soosung Hwang and Steve Satchell, Improved Testing for the Efficiency of Asset Pricing Theories in Linear Factor Models, WP99-19
4. Soosung Hwang and Stephen Satchell, The Disappearance of Style in the US Equity Market, WP99-18
5. Soosung Hwang and Stephen Satchell, Modelling Emerging Market Risk Premia Using Higher Moments, WP99-17
6. Soosung Hwang and Stephen Satchell, Market Risk and the Concept of Fundamental Volatility: Measuring Volatility Across Asset and Derivative Markets and Testing for the Impact of Derivatives Markets on Financial Markets, WP99-16
7. Soosung Hwang, The Effects of Systematic Sampling and Temporal Aggregation on Discrete Time Long Memory Processes and their Finite Sample Properties, WP99-15
8. Ronald MacDonald and Ian Marsh, Currency Spillovers and Tri-Polarity: a Simultaneous Model of the US Dollar, German Mark and Japanese Yen, WP99-14
9. Robert Hillman, Forecasting Inflation with a Non-linear Output Gap Model, WP99-13
10. Robert Hillman and Mark Salmon , From Market Micro-structure to Macro Fundamentals: is there Predictability in the Dollar-Deutsche Mark Exchange Rate?, WP99-12
11. Renzo Avesani, Giampiero Gallo and Mark Salmon, On the Evolution of Credibility and Flexible Exchange Rate Target Zones, WP99-11
12. Paul Marriott and Mark Salmon, An Introduction to Differential Geometry in Econometrics, WP99-10
13. Mark Dixon, Anthony Ledford and Paul Marriott, Finite Sample Inference for Extreme Value Distributions, WP99-09
14. Ian Marsh and David Power, A Panel-Based Investigation into the Relationship Between Stock Prices and Dividends, WP99-08
15. Ian Marsh, An Analysis of the Performance of European Foreign Exchange Forecasters, WP99-07
16. Frank Critchley, Paul Marriott and Mark Salmon, An Elementary Account of Amari's Expected Geometry, WP99-06
17. Demos Tambakis and Anne-Sophie Van Royen, Bootstrap Predictability of Daily Exchange Rates in ARMA Models, WP99-05
18. Christopher Neely and Paul Weller, Technical Analysis and Central Bank Intervention, WP99-04
19. Christopher Neely and Paul Weller, Predictability in International Asset Returns: A Re-examination, WP99-03
20. Christopher Neely and Paul Weller, Intraday Technical Trading in the Foreign Exchange Market, WP99-02
21. Anthony Hall, Soosung Hwang and Stephen Satchell, Using Bayesian Variable Selection Methods to Choose Style Factors in Global Stock Return Models, WP99-01

1998

1. Soosung Hwang and Stephen Satchell, Implied Volatility Forecasting: A Compaison of Different Procedures Including Fractionally Integrated Models with Applications to UK Equity Options, WP98-05
2. Roy Batchelor and David Peel, Rationality Testing under Asymmetric Loss, WP98-04
3. Roy Batchelor, Forecasting T-Bill Yields: Accuracy versus Profitability, WP98-03
4. Adam Kurpiel and Thierry Roncalli , Option Hedging with Stochastic Volatility, WP98-02
5. Adam Kurpiel and Thierry Roncalli, Hopscotch Methods for Two State Financial Models, WP98-01