Properties of Optimal Forecasts

Andrew Patton
London School of Economics

Allan Timmermann
University of California, San Diego
Testing rationality and market efficiency

- Tests of market efficiency and investor rationality are usually done by testing properties of forecast errors.
Testing rationality and market efficiency

- Tests of market efficiency and investor rationality are usually done by testing properties of forecast errors:


- All of these papers rely on properties of optimal forecasts derived assuming squared error loss:
The standard set-up

- The standard results on optimal forecasts were derived under the assumption that:

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

\[ \hat{Y}^*_t \equiv \arg \min_{\hat{y}} E_t \left[ (Y_{t+h} - \hat{y})^2 \right] \]

- Which implies that:

\[ \hat{Y}^*_{t+h,t} = E_t [Y_{t+h}] \]
Standard properties of optimal forecasts

- The properties of optimal forecasts in the standard set-up are:

1. Optimal forecasts are unbiased

2. Optimal h-step forecast errors are serially correlated only to lag (h-1)
   - So 1-step forecasts have zero serial correlation

3. Optimal unconditional forecast error variance is an increasing function of the forecast horizon
Is MSE the right loss function?

- The assumption of MSE loss in economics has been questioned in the (mostly) recent literature:

- For example: financial analysts’ forecasts have been found to be biased upwards
  - A result of analyst irrationality, or simply that the analyst is penalised more heavily for under-predictions than over-predictions?
What we do in this paper

- We extend the work of Christoffersen and Diebold (1997) and Granger (1969, 1999) to analytically consider the time series properties of optimal forecasts under asymmetric loss and nonlinear DGPs.

- We show that all the standard properties may be violated in quite reasonable situations

  - Thus the previous work on market efficiency and investor rationality may be disregarded if you do not believe in MSE loss
What we do in this paper (cont’d)

- We provide some general results on properties of optimal forecasts when the loss function is known, which may then be used in testing rationality.

- We also provide some testable implications of forecast optimality that hold without knowledge of the forecaster’s loss function.

- Finally, we introduce a change of measure, from the objective to the “MSE-loss probability measure”, under which the optimal forecast has the same properties as under MSE loss.
Notation and some assumptions

\[ Y_{t+h} \]  the scalar random variable to be forecast

\[ \hat{Y}_{t+h,t} \]  a forecast made at time \( t \)

\[ \hat{Y}_{t+h,t}^* \]  the optimal forecast made at time \( t \)

\[ L = L\left(Y_{t+h}, \hat{Y}_{t+h,t}\right) \]  the loss function

\[ e_{t+h,t} \equiv Y_{t+h} - \hat{Y}_{t+h,t} \]  the forecast error

\[ \Omega_t \]  time \( t \) information set \( \equiv \sigma(Y_{t-j}; j \geq 0) \)

\[ \hat{Y}_{t+h,t}^* \equiv \arg \min_{\hat{y}} E\left[L\left(Y_{t+h}, \hat{y}\right) \mid \Omega_t \right] \]
Properties in non-standard situations

1. Forecast error has zero conditional mean
   - *Granger (1969) and Christoffersen and Diebold (1997) showed that bias may be optimal under asym loss*

2. The optimal $h$-step forecast error exhibits zero serial correlation beyond the $(h - 1)^{th}$ lag.
   - *Right idea, but wrong object: standard forecast error is not (generally) the variable with zero serial correl.*

3. Unconditional forecast error variance is increasing in $h$.
   - *Variance is not (generally) right measure of forecast accuracy*
A counter-example

- We now present a realistic situation where all the standard properties of optimal forecasts and forecast errors break down.

- Our results are all analytical. We assume that the agent knows his loss function, *and* the DGP (including the parameters of the DGP)
  \[ \rightarrow \text{This agent is as optimal as can possibly be...} \]

- We will define the loss function and DGP as follows:
Counter-example: loss function

- For tractability, we focus on the linear-exponential loss function of Varian (1975)

\[
L(Y_{t+h}, \hat{Y}_{t+h,t}; a) = \exp\{a(Y_{t+h} - \hat{Y}_{t+h,t})\} - a(Y_{t+h} - \hat{Y}_{t+h,t}) - 1 \\
= \exp\{ae_{t+h,t}\} - ae_{t+h,t} - 1
\]
We consider a regime switching process, popular in both macroeconomics and finance:

\[ Y_{t+1} = \mu + \sigma_{st+1} \nu_{t+1}, \quad \nu_{t+1} \sim \text{iid } \mathcal{N}(0,1) \]

\[ S_{t+1} = \{1,2,...,k\} \]

\[ \Pr[S_{t+1} = j | S_t = i] = P_{[i,j]} \]
Counter-example: DGP

- For presentation purposes, I will focus on a particular case of the RS model:

\[ Y_{t+1} = \mu + \sigma_{st+1} \nu_{t+1}, \quad \nu_{t+1} \sim iid \ N(0,1) \]

\[ \mu = 0 \]

\[ \sigma = [0.5,2]' \]

\[ P = \begin{bmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{bmatrix} \]

\[ \Pi = \begin{bmatrix} 2 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}' \]
1. First property: Bias

- Optimal $h$-step forecast in this special case is:
  \[ \hat{Y}_{t+h,t} = \mu + \frac{1}{a} \log(\hat{\pi}_{s,t} P^h \phi) \]

  \[ \phi = \exp\left\{0.5a^2 \sigma^2\right\} \]
1. First property: Bias

- Optimal $h$-step forecast in this special case is:

$$\hat{Y}_{t+h,t} = \mu + \frac{1}{a} \log(\hat{\pi}_{s,t} P^h \varphi)$$

$$\varphi = \exp\left\{0.5a^2 \sigma^2\right\}$$

- which implies conditional and unconditional bias of:

$$E_t[e_{t+h,t}^*] = -\frac{1}{a} \log(\hat{\pi}_{s,t} P^h \varphi)$$

$$E[e_{t+h,t}^*] = -\frac{1}{a} \pi' \log(P^h \varphi)$$

$$\rightarrow -\frac{1}{a} \log(\pi' \varphi) \text{ as } h \rightarrow \infty$$
1. Bias (cont’d)

**Optimal bias for various forecast horizons**

![Graph showing the optimal bias for various forecast horizons.](attachment:graph.png)
2. Second property: Serial correlation

- The $j^{th}$-order serial correlation for the $h$-step forecast is given by:

$$
\text{Cov}\left[ e^*_t, e^*_{t+h-j} \right] = \pi' \sigma^2 1\{j = 0\} + \frac{1}{a} \lambda_h \left( (\pi') \otimes P^j - \pi \pi' \right) \lambda_h
$$

$$
\rightarrow 0 \text{ as } j \rightarrow \infty
$$

where $\lambda_h = \log(P^h \phi)$

- Notice that the only break-point is at $j=0 \Rightarrow$ serial correlation for $j > h-1$ may also be non-zero...
2. Serial correlation (cont’d)

Optimal forecast error autocorrelation function for various forecast horizons

Serial correlation

Lag (j)

0 5 10 15 20

0 0.05 0.1 0.15 0.2 0.25

h=1
h=2
h=3
h=4
h=5
2. Serial correlation (cont’d)

This is the shape that we would expect from standard results:

---

Serial correlation

Lag

0 5 10 15 20
2. Serial correlation - intuition

- Some intuition for this result may be gleaned from a result of Christoffersen and Diebold, who show that:

\[ \hat{Y}_{t+h,t}^* = E_t[Y_{t+h}] + \alpha_{t+h,t} \]

where \( \alpha_{t+h,t} \) depends only on the time-varying moments of order higher than 1.

- If \( \alpha_{t+h,t} \) exhibits persistence, via its dependence on persistent second moments for example, then the forecast error may also exhibit persistence, ie serial correlation.
3. Third property: Forecast error variance

- The conditional forecast error variance can be increasing or decreasing in $h$ under MSE loss – GARCH is a common example here.

- We instead focus on unconditional forecast error variance as a function of $h$, which is non-decreasing for MSE loss. In the RS example it is:

$$V[e^*_{t+h,t}] = \bar{\pi}' \sigma^2 + \frac{1}{a} \lambda'_h ((\bar{\pi}' I) \odot I - \bar{\pi} \bar{\pi}') \lambda_h$$

$$\rightarrow \bar{\pi}' \sigma^2 = V[Y_{t+h,t}] \text{ as } h \rightarrow \infty$$
3. Forecast error variance (cont’d)

Optimal forecast error variance for various forecast horizons

- Variance decreases as the horizon increases.
- The variance ranges from approximately 2.1 to 1.5 as the horizon changes from 1 to 10.

The graph shows a clear downward trend, indicating that the forecast error variance decreases as the forecast horizon increases.
3. Forecast error variance - intuition

- The main intuition here is that, in general, forecast error variance is \textit{not} the right way to measure how difficult it is to forecast.

- Given some loss function $L$, the right way to measure forecast accuracy is \textit{expected loss}.

- It happens that under MSE loss forecast error variance and expected loss coincide:
Expected forecast error loss and variance under MSE loss

- Under MSE loss:

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

\[
E[L(Y_{t+h}, \hat{Y}_{t+h,t}^*)] = E[e_{t+h,t}^2] = V[e_{t+h,t}^*]
\]

- And so it happens that here expected loss and error variance coincide. In general this is not the case.
A recap

- What we’ve shown up to this point:
  1. Optimal forecast errors may have non-zero mean
  2. Optimal forecast errors may be serially correlated
  3. The forecast error variance may *decrease* with the forecast horizon

- But where are these “violations” coming from?
Causes of the violations

- Our counter-example involves *both*:
  - an asymmetric loss function, and
  - a non-linear DGP

- Earlier, we showed that the usual results hold under:
  - Squared error loss, and
  - any stationary DGP

- *What about the case of:*
  - asymmetric loss and
  - a simple DGP, such as an ARMA?
Asymmetric loss and DGP with mean-only dynamics

- Let $Y_{t+h} = E[Y_{t+h} | \Omega_t] + \varepsilon_{t+h}, \varepsilon_{t+h} | \Omega_t \sim D_h$
Asymmetric loss and DGP with mean only dynamics

- Let $Y_{t+h} = E[Y_{t+h} | \Omega_t] + \varepsilon_{t+h}, \varepsilon_{t+h} | \Omega_t \sim D_h$

- Let $L(Y_{t+h}, \hat{Y}_{t+h,t}) = L(Y_{t+h} - \hat{Y}_{t+h,t}) = L(e_{t+h,t})$
Asymmetric loss and DGP with mean only dynamics

- Let \( Y_{t+h} = E[Y_{t+h} \mid \Omega_t] + \varepsilon_{t+h}, \varepsilon_{t+h} \mid \Omega_t \sim D_h \)

- Let \( L(Y_{t+h}, \hat{Y}_{t+h,t}) = L(Y_{t+h} - \hat{Y}_{t+h,t}) = L(e_{t+h,t}) \)

Then: \( \hat{Y}^*_{t+h,t} = E_t[Y_{t+h}] + \alpha_h \)

and

1. Optimal forecast is biased (bias is only a function of \( h \), see Christoffersen and Diebold, 1997)
Asymmetric loss and DGP with mean only dynamics

- Let \( Y_{t+h} = E[Y_{t+h} | \Omega_t] + \varepsilon_{t+h}, \varepsilon_{t+h} | \Omega_t \sim D_h \)

- Let \( L(Y_{t+h}, \hat{Y}_{t+h,t}) = L(Y_{t+h} - \hat{Y}_{t+h,t}) = L(e_{t+h,t}) \)

Then: \( \hat{Y}^*_{t+h,t} = E_t[Y_{t+h}] + \alpha_h \)

and

1. Optimal forecast is biased (bias is only a function of \( h \), see Christoffersen and Diebold, 1997)
2. The \( h \) - step forecast error has MA(\( h-1 \)) ACF
3. The forecast error variance is weakly increasing in \( h \)
$h$-step forecast error has $MA(h-1)\ ACF$

Proof:

\[ Y_{t+h} = E_t[Y_{t+h}] + \varepsilon_{t+h} \]
\[ \hat{Y}_{t+h,t} = E_t[Y_{t+h}] + \alpha_h \]
\[ e^*_t = Y_{t+h} - \hat{Y}_{t+h,t} \]
\[ = \varepsilon_{t+h} - \alpha_h \]
\[ Cov[e^*_{t+h,t}, e^*_{t+h-j,t-j}] = Cov[\varepsilon_{t+h}, \varepsilon_{t+h-j}] \]
\[ = 0 \quad \forall \quad j \geq h \]

since \( \varepsilon_{t+h} | \Omega_t \sim D_h \)
Interpretation

- This shows that the serial correlation properties are robust to the loss function under restrictions on the DGP.

- This implies that if we can assume that there are only conditional mean dynamics, we can test for forecast optimality without any knowledge of the forecaster’s loss function. This extends existing literature:

1. Assume MSE, allow arbitrary DGP

2. Elliott *et al.* (2002): assume loss function up to unknown parameter vector, assume linear forecast model
Error variance is weakly increasing in $h$

Proof:

$$Y_{t+h+j} = E_t[Y_{t+h+j} | \Omega_t] + \eta_{t+h+j}, \quad \eta_{t+h+j} \sim D_{h+j}$$

$$Y_{t+h+j} = E_{t+j}[Y_{t+h+j}] + \varepsilon_{t+h+j}, \quad \varepsilon_{t+h+j} \sim D_h$$

$$e_{t+h+j,t}^* = \eta_{t+h+j} - \alpha_{h+j}$$

$$e_{t+h+j,t+j}^* = \varepsilon_{t+h+j} - \alpha_h$$

$$V_t[e_{t+h+j,t}^*] = \sigma^2_{h+j} = V[e_{t+h+j,t}^*]$$

$$V_t[e_{t+h+j,t+j}^*] = \sigma^2_h = V[e_{t+h+j,t+j}^*]$$

Want to show $\sigma^2_{h+j} \geq \sigma^2_h$
Error variance is weakly increasing in $h$

\[
\sigma^2_{h+j} = V_t[e^*_{t+h+j,t}]
\]

\[
= V_t[Y_{t+h+j} - E_t[Y_{t+h+j}]]
\]

\[
= V_t[\varepsilon_{t+h+j} + E_{t+j}[Y_{t+h+j}] - E_t[Y_{t+h+j}]]
\]

\[
= V_t[\varepsilon_{t+h+j}] + V_t[E_{t+j}[Y_{t+h+j}]] + 2\text{Cov}_t[\varepsilon_{t+h+j}, E_{t+j}[Y_{t+h+j}]]
\]

\[
= V_t[\varepsilon_{t+h+j}] + V_t[E_{t+j}[Y_{t+h+j}]]
\]

\[
\geq V_t[\varepsilon_{t+h+j}]
\]

\[
= \sigma^2_h
\]
Some intuition

- What’s behind the results violating standard properties?
  - A mis-match of the loss function and MSE
  - Dynamics in the process beyond the mean

- The standard results all follow from the use of the squared error as the loss function, and when a different loss is employed we find “violations”

- So what are the properties optimal forecasts in general situations?
The “generalised forecast error”

- Granger (1999) proposes looking at a generalised forecast error. We modify his definition slightly.

- The generalised forecast error comes out of the first-order condition for forecast optimality:

\[
\hat{Y}_{t+h,t}^* = \arg \min_{\hat{y}} E\left[ L(Y_{t+h}, \hat{y}) \mid \Omega_t \right]
\]

\[
FOC : \frac{\partial E\left[ L(Y_{t+h}, \hat{Y}_{t+h,t}^*) \mid \Omega_t \right]}{\partial \hat{y}} = 0
\]
The generalised forecast error

- A natural alternative to the standard forecast error is thus:

\[ \psi_{t+h,t} = \frac{\partial L(Y_{t+h}, \hat{Y}_{t+h,t})}{\partial \hat{Y}} \]

- Notice that under MSE the generalised and standard forecast errors are related by:

\[ \psi_{t+h,t}^* = -2e_{t+h,t}^* \]

- The properties assigned to \( e_{t+h,t}^* \) are actually properties of \( \psi_{t+h,t}^* \) more generally
Properties of optimal forecast errors under general conditions

- By using the generalised forecast error and the arbitrary loss function $\mathcal{L}$ we can provide properties of optimal forecasts more generally:

1. $E_t[\psi_{t+h,t}^*] = E[\psi_{t+h,t}^*] = 0$.

2. The generalised forecast error from an optimal $h$-step forecast has the same ACF as some MA($h-1$) process.

3. Unconditional expected loss is non-decreasing in $h$. 
1. Mean of generalised forecast error

- By the first-order condition for an optimal forecast we have:

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

so \( E \left[ \psi_{t+h,t}^* \right] = 0 \) by the L.I.E.

(assuming that we can interchange the differentiation and expectation operators.)
2. Serial correlation

- Instead of referring to an MA($h - 1$) process, we show that the generalised forecast errors are uncorrelated for lags $> h - 1$, i.e., it has the same ACF as some MA($h - 1$) process.

\[
E[\psi_{t+h,t}^* | \Omega_t] = 0 \Rightarrow E[\psi_{t+h,t}^* \cdot \gamma(\psi_{t+h-j,t-j}^*)] = 0
\]

for all $j \geq h$ and any function $\gamma$

\[
\Rightarrow (\psi_{t+h,t}^*, \psi_{t+h-j,t-j}^*) \text{ are uncorrelated for all } j \geq h
\]
3. Expected loss

- The unconditional expected loss from an optimal forecast is non-decreasing in the forecast horizon.

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

$$E[L(Y_{t+h}, \hat{Y}^*_{t+h,t}) | \Omega_t] \leq E[L(Y_{t+h}, \hat{Y}^*_{t+h,t-j}) | \Omega_t]$$

$$E[L(Y_{t+h}, \hat{Y}^*_{t+h,t})] \leq E[L(Y_{t+h}, \hat{Y}^*_{t+h,t-j})] \text{ by the L.I.E.}$$

$$= E[L(Y_{t+h+j}, \hat{Y}^*_{t+h+j,t})]$$
Properties under a different measure

- Here we propose retaining the object of interest, but changing its probability distribution

- This is akin to moving from the objective to the risk-neutral measure in asset pricing.
  - After a change of measure, assets may be priced as though agents are risk neutral

- Following our change of measure, the optimal forecast errors have the same properties as under MSE loss
  - So bias and serial correlation may be tested, for example
A change of measure - assumptions

- Suppose:

\[
\frac{\partial L(Y_{t+h}, \hat{Y}_{t+h,t})}{\partial \hat{y}} \geq 0 \text{ if } Y_{t+h} - \hat{Y}_{t+h,t} < 0
\]

\[
\frac{\partial L(Y_{t+h}, \hat{Y}_{t+h,t})}{\partial \hat{y}} \leq 0 \text{ if } Y_{t+h} - \hat{Y}_{t+h,t} > 0
\]

\[
0 < \left| E_t \left[ \frac{1}{e_{t+h}} \frac{\partial L(Y_{t+h}, \hat{Y}_{t+h,t})}{\partial \hat{y}} \right] \right| < \infty
\]
A change of measure - formula

Notice that:

\[ f_{e_{t+h},t}(e; \hat{Y}_{t+h}) = f_{t+h}(Y_{t+h} - \hat{Y}_{t+h}) \forall e, \hat{Y}_{t+h} \]
A change of measure - formula

- Notice that:

\[ f_{e_{t+h,t}}(e; \hat{Y}_{t+h,t}) = f_{t+h,t}(Y_{t+h} - \hat{Y}_{t+h,t}) \forall e, \hat{Y}_{t+h,t} \]

Let the “MSE-loss probability measure” be defined as:

\[
\begin{align*}
\hat{f}_{e_{t+h,t}}^*(e; \hat{Y}_{t+h,t}) &= \frac{1}{e} \cdot \left. \frac{\partial L(Y, \hat{Y}_{t+h,t})}{\partial \hat{Y}} \right|_{Y=\hat{Y}_{t+h,t} + e} \cdot f_{e_{t+h,t}}(e; \hat{Y}_{t+h,t}) \\
&= E_t \left[ \frac{1}{e_{t+h,t}} \cdot \frac{\partial L(Y_{t+h}, \hat{Y}_{t+h,t})}{\partial \hat{Y}} \right]
\end{align*}
\]
Change of measure - validity

- Must show that the new measure is a valid probability measure.

- By assumption

\[
\frac{1}{Y_{t+h} - \hat{Y}_{t+h,t}} \cdot \frac{\partial L(Y_{t+h}, \hat{Y}_{t+h,t})}{\partial \hat{y}} \leq 0 \forall Y_{t+h}, \hat{Y}_{t+h,t}
\]

- So the denominator is negative, and numerator is weakly negative, thus entire expression is weakly positive

- By construction it integrates to 1, so it is a valid pdf.
Mean under MSE loss measure

Proposition: Under the MSE-loss probability measure the optimal forecast error has conditional mean zero.

Proof:

\[ E_t^*[e_{t+h,t}^*] = A^{-1} \int e e \left. \frac{\partial L(Y, \hat{Y}_{t+h,t}^*)}{\partial \hat{y}} \right|_{Y=\hat{Y}_{t+h,t}^* + e} \]
Mean under MSE-loss measure

Proposition: Under the MSE-loss probability measure the optimal forecast error has conditional mean zero.

Proof:

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

\[ = A^{-1} \cdot \int \frac{\partial L(Y, \hat{Y}_{t+h,t}^*)}{\partial \hat{y}} \left| _{Y=\hat{Y}_{t+h,t}^*+e} \right. f_{e_{t+h,t}} (e; \hat{Y}_{t+h,t}^*) \]

\[ = 0 \]

by the first-order condition for forecast optimality.
Serial correl under MSE-loss measure

Proposition: The optimal \( h \)-step forecast error has zero serial correlation beyond lag \( h - 1 \).

Proof:

\[
\text{Cov}^* [e_{t+h,t}^*, e_{t+h-j,t-j}^*] = \text{E}^* [e_{t+h,t}^* \cdot e_{t+h-j,t-j}^*] \\
= \text{E}^* [\text{E}_t^* [e_{t+h,t}^*] \cdot e_{t+h-j,t-j}^*] \forall j \geq h \\
= 0
\]
Objective and MSE-loss error densities

Pi = [2/3, 1/3]'
Summary of Results: Implications

- Tests of forecast optimality/forecaster rationality that are based on the standard forecast errors (generally) implicitly assume MSE loss.

- If the forecast user/provider has a different loss function, the forecasts may be perfectly optimal and still violate standard properties.

- Our results simply show that without some knowledge of the forecaster’s loss function testing forecast optimality is an extremely difficult task.
Summary: Testing optimality

- If the forecaster’s loss function is known, the results in this paper may be used to construct tests of forecast optimality
  - Combine our results with the tests of Diebold-Mariano (1995) or West (1996)

- If the forecaster’s loss function is known up to an unknown parameter, the work of Elliott, Komunjer and Timmermann (2002) may be used instead

- If the DGP is known to only have conditional mean dynamics we showed that forecast optimality may be tested with much robustness to the unknown loss function.