Properties of Optimal Forecasts

Andrew Patton

Department of Economics
Duke University

OMI-SoFiE Summer School, 2013
Papers to be covered


Outline

1. Types of economic forecasts

2. Definition of an optimal forecast

3. Properties of optimal forecasts
   - MSE
   - MAE and Lin-Lin
   - Linex
   - General loss functions
   - Unknown loss functions
Thoughts on economic forecasting

- **Wall Street indices predicted nine out of the last five recessions**
  - Paul Samuelson, Nobel laureate in economics

- **Prediction is very difficult, especially if it’s about the future.**
  - Nils Bohr, Nobel laureate in physics

- **The only function of economic forecasting is to make astrology look respectable.**
  - John Kenneth Galbraith
Serious thoughts on economic forecasting

- *It is not so much their wrongness as their pretensions to rightness that have brought economic predictions and the theories that underlie them into well-deserved contempt.*
  - Peter Medawar (1982)

- Comparison between weather forecasting and economic forecasting (also from Medawar 1982):

<table>
<thead>
<tr>
<th>Weather forecasting</th>
<th>Economic forecasting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables observable</td>
<td>Often unobservable</td>
</tr>
<tr>
<td>Functional relations often exactly known</td>
<td>Usually not known</td>
</tr>
<tr>
<td>Uninfluenced by politics or fashion</td>
<td>Influenced by both</td>
</tr>
<tr>
<td>Wholly non-reflexive</td>
<td>Highly reflexive</td>
</tr>
</tbody>
</table>
Types of forecasts

- Economic and financial forecasts can take one of three different types
  - Point, interval and density (discussed below)
  - Understanding the type of forecast is important for:

1. **Constructing** optimal forecasts
   - What (if any) structure does the optimal forecast have?

2. **Evaluating** proposed forecasts
   - How can we tell if a given forecast is different from the optimal forecast?

3. **Comparing** competing forecasts
   - How can we tell whether one forecast is “better” than another?

4. **Combining** multiple forecasts
   - Is a combination of individual forecasts better than any individual forecast?
Types of forecasts: (1) Point forecasts

- A **point forecast** is the most common type of economic forecast.
- In the univariate case it is a single number (in multivariate case it is a vector).
  - Eg: Conditional mean, median, variance, or some other function
- Ideally, a point forecast is obtained by minimizing the conditional expectation of the future loss:
  \[
  \hat{Y}_{t+h|t}^* = \arg\min_{\hat{y} \in \mathcal{Y}} E \left[ L \left( Y_{t+h}, \hat{y} \right) | \mathcal{F}_t \right]
  \]

- These lectures will focus on point forecasts (many more details to come...)

Types of forecasts: (2) Interval forecasts

- **An interval forecast** takes the form: “The variable $Y_{t+h}$ will lie between $L$ and $U$ with probability $p$”.

- These forecasts are often an attempt to succinctly convey the **uncertainty** around a point forecast

  - Eg: Latest polling (April 2013) in Australian federal politics has the Labor party with 45% of the vote, with a **margin of error** $\pm 3\%$. So an interval forecast for the Labor share of the vote would be $[42, 48]$ with $p = 0.95$.

- Interval forecasts can also be used to focus attention on the **tail** of the distribution

  - Eg: a 5% **Value-at-Risk** forecast can be interpreted as an interval forecast for future returns, with the interval being $(-\infty, \text{VaR}]$ and the $p = 0.05$. 

Patton (Duke)
Types of forecasts: (3) Density forecasts

- A **density forecast** takes the form: “The variable $Y_{t+h}$ will be drawn from distribution $F$.”
  - Eg: the return on the FTSE 100 tomorrow will be drawn from $N(0.02, 0.94^2)$

- Clearly this is the most comprehensive of the three forecasts
  - Given a density forecast and a loss function, the optimal point forecast can be obtained
  - Given a density forecast, any interval forecast can be obtained

- Density forecasts can be used to convey even more information about uncertainty
Bank of England “fan chart” for CPI inflation
May 2013. Color bands cover 30%, 60% and 90% of probability
Bank of England “fan chart” for GDP growth

May 2013. Color bands cover 30%, 60% and 90% of probability.
Outline

1. Types of economic forecasts

2. Definition of an optimal forecast

3. Properties of optimal forecasts
   - MSE
   - MAE and Lin-Lin
   - Linex
   - General loss functions
   - Unknown loss functions
Definition of an optimal forecast

- Given a loss function $L$, the optimal (point) forecast is obtained by minimizing the conditional expectation of the future loss:

$$\hat{Y}_{t+h|t}^* = \arg\min_{\hat{y}} E \left[ L \left( Y_{t+h}, \hat{y} \right) | \mathcal{F}_t \right]$$

- This definition has several components that are worth noting:
  - The target variable, $Y$: which variable are we forecasting?
  - The forecast horizon, $h$. Note that this depends on the choice of unit of time
    - Using weekly data, a one-week-ahead forecast is $h = 1$, while using daily data it would be $h = 5$
  - The range of possible values for the forecast, $\mathcal{Y}$
Definition of an optimal forecast, cont’d

\[ \hat{Y}_{t+h|t}^* \equiv \arg \min_{\hat{y} \in \mathcal{Y}} E \left[ L (Y_{t+h}, \hat{y}) \mid \mathcal{F}_t \right] \]

- The **loss function**, \( L \): this function maps the realization of the target variable and the forecast to a “loss” or “cost.”
  - Usually normalized to be zero when \( \hat{Y} = Y \)
  - Common loss functions include the MSE (quadratic) and MAE (absolute):
    - MSE: \( L (y, \hat{y}) = (y - \hat{y})^2 \)
    - MAE: \( L (y, \hat{y}) = |y - \hat{y}| \)

- **Information set**, \( \mathcal{F}_t \). What are we assuming that the forecaster uses to construct the forecast?
Definition of an optimal forecast, cont’d

\[ \hat{Y}_{t+h|t} \equiv \arg \min_{\hat{y} \in \mathcal{Y}} E \left[ L \left( Y_{t+h}, \hat{y} \right) \mid \mathcal{F}_t \right] \]

- **Summary measure** for future loss: Standing at time \( t \), \( L \left( Y_{t+h}, \hat{y} \right) \) is a random variable. Common to minimize expected loss, but not the only choice
  - Minimize **median** loss – less sensitive to extremes
  - Minimize **maximum** loss – very averse to loss (note: this may not exist)
  - Minimize some (high) **quantile** of loss – related to robust estimation and prediction
Outline

1 Types of economic forecasts

2 Definition of an optimal forecast

3 Properties of optimal forecasts
   - MSE
   - MAE and Lin-Lin
   - Linex
   - General loss functions
   - Unknown loss functions
Optimal forecasts under MSE

- The most common loss function is the quadratic, or MSE.

- Obtaining the optimal forecast for this loss function is straightforward:

$$
\hat{Y}_{t+h|t}^* \equiv \arg \min_{\hat{y} \in \mathcal{Y}} E \left[ (Y_{t+h} - \hat{y})^2 | \mathcal{F}_t \right]
$$

$$
\text{FOC} \ 0 \ = \ \frac{\partial}{\partial \hat{y}} E \left[ \left( Y_{t+h} - \hat{Y}_{t+h|t}^* \right)^2 | \mathcal{F}_t \right] = -2E \left[ \left( Y_{t+h} - \hat{Y}_{t+h|t}^* \right) | \mathcal{F}_t \right]
$$

so

$$
\hat{Y}_{t+h|t}^* \ = \ E \left[ Y_{t+h} | \mathcal{F}_t \right]
$$

- This helps explain the common usage of “forecast”, “prediction” and “expectation” as synonyms.

- Also note, this links MSE prediction to OLS estimation.
Properties of optimal forecasts under MSE

- Given that $\hat{Y}_{t+h|t} = E[Y_{t+h}|F_t]$ under MSE we can derive some (testable) properties of this forecast or its error, $e^*_{t+h|t} \equiv Y_{t+h} - \hat{Y}_{t+h|t}$

1. $E\left[e^*_{t+h|t}\right] = 0$. Optimal forecasts are unbiased.

2. $\text{Corr}\left[e^*_{t+h|t}, Z_t\right] = 0 \ \forall \ Z_t \in F_t$. Optimal forecast errors are unpredictable.

3. $V\left[e^*_{t+h_S|t}\right] \leq V\left[e^*_{t+h_L|t}\right]$ for all $h_S \leq h_L$. It is more difficult to predict further into the future.
Forecast optimality testing under MSE I

- Each of the three above implications can be tested empirically.
- The first two implications are often tested via a simple OLS regression:

\[ e_{t+h|t} = \alpha_0 + \alpha_1 e_{t|t-h} + \alpha_2 Z_t + u_{t+h} \]

\[ H_0 : \quad \alpha_0 = \alpha_1 = \alpha_2 = 0 \]

\[ \text{vs. } H_1 : \quad \alpha_j \neq 0 \text{ for some } j = 0, 1, 2 \]

- where \( Z_t \) is anything known at time \( t \) (and might be a vector)

- Note that this approach tests only a necessary condition for optimality

- A forecast may be sub-optimal, but still satisfy \( H_0 \) above.

- Tests that will (asymptotically) detect any deviation from optimality can be constructed using the methods of Bierens (1990) and de Jong (1996). We won’t consider these here.
Forecast optimality testing under MSE II

- Another well-known regression-based test of optimality under MSE is the “Mincer-Zarnowitz” regression:

\[
Y_{t+h} = \beta_0 + \beta_1 \hat{Y}_{t\mid t-h} + u_{t+h}
\]

\[H_0 : \beta_0 = 0 \cap \beta_1 = 1\]

vs.

\[H_1 : \beta_0 \neq 0 \cup \beta_1 \neq 1\]

- This method can be “augmented” by including \(Z_t\) variables on the RHS:

\[
Y_{t+h} = \beta_0 + \beta_1 \hat{Y}_{t\mid t-h} + \beta_2 Z_t + u_{t+h}
\]

\[H_0 : \beta_0 = 0 \cap \beta_1 = 1 \cap \beta_2 = 0\]

vs.

\[H_1 : \beta_0 \neq 0 \cup \beta_1 \neq 1 \cup \beta_2 \neq 0\]
When forecasts for multiple horizons \((h)\) are available, the \textbf{third implication} of optimality can also be tested

\[
V \left[ e^*_{t|t-h_S} \right] \leq V \left[ e^*_{t|t-h_L} \right] \quad \text{for all } h_S \leq h_L
\]

Patton and Timmermann (2012) propose a variety of tests of \textbf{bounds} such as the one above, based on existing methods for testing (multiple) \textbf{inequality restrictions}.

The tests of Wolak (1989) and White (2000) can be used here.
Forecast optimality testing under MSE IV

- PT also present other bounds that are implied by optimality under MSE:

\[
V \left[ \hat{Y}_{t|h_S} \right] \geq V \left[ \hat{Y}_{t|h_L} \right] \quad \text{for all } h_S \leq h_L
\]

- i.e., short-horizon forecasts should be more volatile than long-horizon forecasts. (Note that this does not require data on the target variable)

\[
Cov \left[ Y_t, \hat{Y}_{t|h_S} \right] \geq V \left[ Y_t, \hat{Y}_{t|h_L} \right] \quad \text{for all } h_S \leq h_L
\]

- i.e., short horizon forecasts should be more correlated with the target variable

\[
V \left[ \hat{Y}_{t|h_S} - \hat{Y}_{t|h_L} \right] \leq 2 \times Cov \left[ Y_t, \hat{Y}_{t|h_S} - \hat{Y}_{t|h_L} \right]
\]

- i.e., the forecast revision variance is bounded by (twice) its covariance with the target
Outline

1. Types of economic forecasts
2. Definition of an optimal forecast
3. Properties of optimal forecasts
   - MSE
   - MAE and Lin-Lin
   - Linex
   - General loss functions
   - Unknown loss functions
Another common loss function is the absolute error loss function

Obaining the optimal forecast for this loss function is a little harder, as the loss function is non-differentiable:

\[
\hat{Y}_{t+h|t}^* \equiv \arg \min_{\hat{y} \in \mathcal{Y}} E \left[ |Y_{t+h} - \hat{y}| \mid \mathcal{F}_t \right]
\]

\[
= \arg \min_{\hat{y} \in \mathcal{Y}} \int_{-\infty}^{\infty} |y - \hat{y}| f_{t+h|t}(y) \, dy
\]

\[
= \arg \min_{\hat{y} \in \mathcal{Y}} - \int_{-\infty}^{\hat{y}} (y - \hat{y}) f_{t+h|t}(y) \, dy + \int_{\hat{y}}^{\infty} (y - \hat{y}) f_{t+h|t}(y) \, dy
\]

Now we’ve removed the non-differentiability of the objective function, and we use Leibniz’s rule to obtain the first-order condition:

\[
\frac{\partial}{\partial x} \int_{a(x)}^{b(x)} f(x, y) \, dy = \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x, y) \, dy + f(b, y) \frac{\partial b}{\partial x} - f(a, y) \frac{\partial a}{\partial x}
\]
Optimal forecasts under MAE II

\[
0 = \frac{\partial}{\partial \hat{Y}} E \left[ \left( Y_{t+h} - \hat{Y}_{t+h|t}^* \right) \mid \mathcal{F}_t \right]
\]

\[
= - \left( \int_{-\infty}^{\hat{Y}_{t+h|t}^*} (-1) f_{t+h|t}(y) \, dy + (0) f_{t+h|t} \left( \hat{Y}_{t+h|t}^* \right) (1) - 0 \right) \\
+ \left( \int_{\hat{Y}_{t+h|t}^*}^{\infty} (-1) f_{t+h|t}(y) \, dy + 0 - (0) f_{t+h|t} \left( \hat{Y}_{t+h|t}^* \right) (1) \right)
\]

\[
= \int_{-\infty}^{\hat{Y}_{t+h|t}^*} f_{t+h|t}(y) \, dy - \int_{\hat{Y}_{t+h|t}^*}^{\infty} f_{t+h|t}(y) \, dy
\]

\[
\equiv F_{t+h|t} \left( \hat{Y}_{t+h|t}^* \right) - \left( 1 - F_{t+h|t} \left( \hat{Y}_{t+h|t}^* \right) \right)
\]

so \( F_{t+h|t} \left( \hat{Y}_{t+h|t}^* \right) = \frac{1}{2} \)

or \( \hat{Y}_{t+h|t}^* = \text{Median}_{t} [Y_{t+h}] \)

Note this links MAE prediction to quantile regression.
Properties of optimal forecasts under MAE

- If \( F_{t+h|t} \) is asymmetric then \( \text{Median}_t [Y_{t+h}] \neq E_t [Y_{t+h}] \), and so the optimal forecast under MAE will differ from that under MSE.

- This means that the properties of the optimal forecast (and its error) are different, in general.

- Using standard tests that were developed for MSE-optimal forecasts may lead to the rejection of a fully optimal MAE-based forecast.

- Below we derive MAE analogs of the results for MSE prediction errors.

- We will make use of the fact that

\[
\text{Median}_t \left[ e^*_{t+h|t} \right] = 0
\]

\[
\Leftrightarrow \Pr_t \left[ e^*_{t+h|t} \leq 0 \right] = \frac{1}{2}
\]

\[
\Leftrightarrow E_t \left[ 1 \left\{ e^*_{t+h|t} \leq 0 \right\} - \frac{1}{2} \right] = 0
\]
Properties of optimal forecasts under MAE II

- Given $\hat{Y}_{t+h|t}^* = \text{Median}_t [Y_{t+h}]$, we know:

1. $\text{Median} \left[ e_{t+h|t}^* \right] = 0$. Optimal forecasts are (unconditional) **median unbiased**

2. $E \left[ \left( 1 \left\{ e_{t+h|t}^* \leq 0 \right\} - \frac{1}{2} \right) \cdot Z_t \right] = 0 \ \forall \ Z_t \in \mathcal{F}_t$. Optimal forecast errors are **median unpredictable**.

3. $E \left[ \left| e_{t+h_L|t}^* \right| \right] \leq E \left[ \left| e_{t+h_S|t}^* \right| \right]$ for all $h_S \leq h_L$. It is more difficult to predict further into the future.

- Importantly, “difficulty” must be measured using the forecaster’s loss function (not variance)
Tests of forecast optimality can be conducted using **quantile (median)** regression.

That is, estimate the following using quantile regression:

\[ e_{t+h|t} = \alpha_0 + \alpha_1 Z_t + u_{t+h} \]

and then test

\[ H_0 : \alpha_0 = \alpha_1 = 0 \]

vs.

\[ H_1 : \alpha_j \neq 0 \text{ for some } j = 0, 1 \]

Estimation and inference for quantile regression is more involved than OLS estimation, see Komunjer (2005, 2013) for details.

The third implication can be tested using the bounds approach of PT (2012).
Optimal forecasts under Lin-Lin loss

Properties of Optimal Forecasts

Lin-lin loss functions

- \( a = 0.5 \)
- \( a = 0.33 \)
- \( a = 0.25 \)
- \( a = 0.05 \)
Optimal forecasts under Lin-Lin loss

- The “lin-lin” loss function:

\[
L(y, \hat{y}; a) = \begin{cases} 
(1-a) |y - \hat{y}|, & y \leq \hat{y} \\
 a |y - \hat{y}|, & y > \hat{y}
\end{cases}, \text{ for } a \in (0, 1)
\]

- This loss function nests MAE at \(a = 0.5\).

- Using the same approach as for MAE it is possible to show that under this loss function the FOC yields

\[
F_{t+h|t} \left( \hat{Y}^*_t \right) = a \in (0, 1)
\]

so \( \hat{Y}^*_t = \text{Quantile}_{t}^{(a)} [Y_{t+h}] \)

- Similar results on properties of the lin-lin forecast error can be obtained, and quantile regression can again be used to test optimality.
The “linex” loss function is a popular asymmetric loss function, which has a neat result when combined with Normal random variables:

\[ L(y, \hat{y}; a) = \frac{2}{a^2} \left[ \exp \{ a(y - \hat{y}) \} - a(y - \hat{y}) - 1 \right] \]

\[ \rightarrow (y - \hat{y})^2 \quad \text{as} \quad a \rightarrow 0 \]
Linex loss

Properties of Optimal Forecasts

OMI-SoFiE Summer School, 2013
Optimal forecasts under Linex loss I

- The “linex” loss function is a popular asymmetric loss function, which has a neat result when combined with Normal random variables:

\[
L(y, \hat{y}; a) = \frac{2}{a^2} \left[ \exp \{ a (y - \hat{y}) \} - a (y - \hat{y}) - 1 \right] \\
\rightarrow (y - \hat{y})^2 \text{ as } a \to 0
\]

- If \( Y_{t+h} | \mathcal{F}_t \sim N(\mu_{t+h|t}, \sigma_{t+h|t}^2) \) then recall that

\[
E \left[ \exp \{ a (Y_{t+h} - \hat{y}) \} \right] = E \left[ \exp \{ aY_{t+h} \} \right] \exp \{-a\hat{y}\} \\
= \exp \left\{ a\mu_{t+h|t} + \frac{1}{2} a^2 \sigma_{t+h|t}^2 \right\} \exp \{-a\hat{y}\} \\
= \exp \left\{ a \left( \mu_{t+h|t} - \hat{y} \right) + \frac{1}{2} a^2 \sigma_{t+h|t}^2 \right\}
\]
So we find (dropping the $2/a^2$ and the $-1$):

$$\hat{Y}_{t+h|t}^* \equiv \arg\min_{\hat{y}\in\mathcal{Y}} (E \{ \exp \{ a (Y_{t+h} - \hat{y}) \} \} - aE \{ Y_{t+h} - \hat{y} \} | \mathcal{F}_t)$$

$$= \arg\min_{\hat{y}\in\mathcal{Y}} \exp \left\{ a \left( \mu_{t+h|t} - \hat{y} \right) + \frac{1}{2} a^2 \sigma_{t+h|t}^2 \right\} - a \left( \mu_{t+h|t} - \hat{y} \right)$$

$$\text{FOC } 0 = -a \exp \left\{ a \left( \mu_{t+h|t} - \hat{Y}_{t+h|t}^* \right) + \frac{1}{2} a^2 \sigma_{t+h|t}^2 \right\} + a$$

so $$\hat{Y}_{t+h|t}^* = \mu_{t+h|t} + \frac{a}{2} \sigma_{t+h|t}^2$$

So the optimal forecast is a function of both the conditional mean and the conditional variance.

And when $a = 0$, corresponding to MSE loss, we obtain the usual result under MSE.
Optimal forecasts under Linex loss III

- Note that testing optimality under linex loss is more difficult than MSE or Lin-Lin: the optimal forecast is a function of two conditional moments, which may have quite different dynamics to each other.

- Without putting some more structure on the problem (eg, on the DGP of $Y_{t+h}$) we cannot base tests on the forecast error, $e_{t+h|t} = Y_{t+h} - \hat{Y}_{t+h|t}$.

- This motivates considering properties of optimal forecasts under general loss...
Optimal forecasts under general loss I

- Not every loss function will yield a closed-form expression for the optimal forecast
  \[ \hat{Y}_{t+h|t}^\ast \equiv \arg\min_{\hat{y} \in \mathcal{Y}} \mathbb{E} \left[ L \left( Y_{t+h}, \hat{y} \right) | \mathcal{F}_t \right] \]

- In fact, the set of loss functions for which we can obtain such an expression is (likely) much lower in reality than in textbooks
  - We like getting things in closed-form...

- Despite the lack of closed-form representation for the optimal forecast, we can often still study it
  - Related to the analysis of extremum estimators: the solution to the optimization problem implicitly defines the optimal forecast.
Examples of loss functions
Testing optimality under general loss I

- Consider the general definition of a loss function:

\[
\hat{Y}_{t+h|t}^* \equiv \arg \min_{\hat{Y} \in \mathcal{Y}} E \left[ L \left( Y_{t+h}, \hat{Y} \right) | \mathcal{F}_t \right]
\]

- Without a closed-form expression for \( \hat{Y}_{t+h|t}^* \), which we used above to obtain testable implications of forecast optimality, we must take a different route.

- Note that the first-order condition for optimality (assuming that the loss function is differentiable) is:

\[
\text{FOC} \quad 0 = E \left[ \psi_{t+h|t}^* | \mathcal{F}_t \right]
\]

where

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

- The variable \( \psi_{t+h|t}^* \) is known as the "generalized forecast error" (Granger 1999).
Patton and Timmermann (2010) propose testable implications of optimality that hold for general loss functions:

1. $E \left[ \psi^*_t | h_t \right] = 0.$

2. $\text{Corr} \left[ \psi^*_t | h_t , Z_t \right] = 0$ for any $Z_t \in F_t$

3. $E \left[ L \left( Y_t , \hat{Y}^*_t | t-h_S \right) \right] \leq E \left[ L \left( Y_t , \hat{Y}^*_t | t-h_L \right) \right]$ for all $h_S \leq h_L.$

When the loss function is MSE, these simplify to the familiar properties discussed above.
Testing optimality under general loss III

- Again, the first two properties can be tested using a regression-based approach:

\[
\psi_{t+h|t} = \alpha_0 + \alpha_1 Z_t + u_{t+h}
\]

\[H_0 : \alpha_0 = \alpha_1 = 0\]

vs. \[H_1 : \alpha_j \neq 0 \text{ for some } j = 0, 1\]

where \(Z_t\) is anything known at time \(t\) (and might be a vector)

- The third property can be tested using the bounds approach of PT (2012)
Outline

1 Types of economic forecasts

2 Definition of an optimal forecast

3 Properties of optimal forecasts
   - MSE
   - MAE and Lin-Lin
   - Linex
   - General loss functions
   - Unknown loss functions
Forecast optimality under *unknown* loss

- When **constructing** a set of forecasts, the loss function is of course **known**
- But in many real-world settings, the **forecast producer** is separate from the **forecast consumer**
  - In this case, the loss function used to obtain the forecast may be **unknown**
- It may be of interest to know whether there exists any loss function that makes a given sequence of forecasts optimal
  - This is testing forecast optimality under unknown loss
  - Making progress on such a tough problem requires making some assumptions
  - This problem was considered by Elliott, Komunjer and Timmermann (2005) and Patton and Timmermann (2007)
Patton and Timmermann (2007)

Patton and Timmermann establish general conditions under which forecast optimality can be tested with only minimal assumptions on the loss function.

The two main results in PT involve a trade-off between assumptions made about the loss function and assumptions made about the DGP:

1. The target variable is conditionally homoskedastic and the loss function is a function solely of the forecast error.

2. The target variable is conditionally location-scale and the loss function is a homogeneous function solely of the forecast error.
PT’s assumptions: DGP

- **Assumption D1:**

  \[ Y_{t+h} = \mu_{t+h|t} + \epsilon_{t+h} \]

  where \( \epsilon_{t+h}|\mathcal{F}_t \sim F_{\epsilon, h} \left( 0, \sigma^2_{\epsilon, h} \right) \)

  where \( F_{\epsilon, h} \left( 0, \sigma^2_{\epsilon, h} \right) \) is some distribution with mean zero and variance \( \sigma^2_{\epsilon, h} \), which may depend on \( h \), but does not depend on \( \mathcal{F}_t \).

- **Assumption D2:**

  \[ Y_{t+h} = \mu_{t+h|t} + \sigma_{t+h|t} \eta_{t+h} \]

  where \( \eta_{t+h}|\mathcal{F}_t \sim F_{\eta, h} \left( 0, 1 \right) \)

  where \( F_{\eta, h} \left( 0, 1 \right) \) is some distribution with mean zero and unit variance, which may depend on \( h \), but does not depend on \( \mathcal{F}_t \).
PT’s assumptions: Loss function

- **Assumption L1:** *The loss function depends solely on the forecast error.* That is, \( L(y, \hat{y}) = L(y - \hat{y}) = L(e) \), \( \forall (y, \hat{y}) \in \mathbb{R} \times \mathcal{Y} \).

This class includes almost all commonly-used loss functions: MSE, MAE, the EKT class, and Linex. It excludes the “MSE-prop” (or “HMSE”) loss function:

\[
L = (y/\hat{y} - 1)^2
\]

- **Assumption L2:** *The loss function is a homogeneous function solely of the forecast error.* That is, \( L(ae) = g(a) L(e) \) for some positive function \( g \).

This class includes most commonly-used loss functions: MSE, MAE and the EKT class. It excludes “MSE-prop”, and also “Linex” loss:

\[
L = \exp \{a(y - \hat{y}) \} - a(y - \hat{y}) - 1
\]
PT’s Proposition 1

**Proposition 1:** Let the DGP and loss function satisfy D1 and L1. Then:

1. The optimal forecast takes the form:
   \[
   \hat{Y}_{t+h|t} = \mu_{t+h|t} + \alpha_h^*
   \]
   where \(\alpha_h^*\) is a constant that depends on \(L\) and \(F_{\varepsilon,h}\), but not on \(\mathcal{F}_t\). (This is due to Granger 1969.)

2. The optimal forecast error \(e_{t+h|t}^*\) is independent of all \(Z_t \in \mathcal{F}_t\), since
   \[
   e_{t+h|t}^* \equiv Y_{t+h} - \hat{Y}_{t+h|t} = \varepsilon_{t+h} + \alpha_h^*
   \]
   where \(\varepsilon_{t+h|\mathcal{F}_t} \sim F_{\varepsilon,h} \left(0, \sigma^2_{\varepsilon,h} \right)\)
PT’s Proposition 1, implications

- Though simple, PT’s proposition 1 has big implications: forecast optimality can be tested nonparametrically - only minimal assumptions are required on the loss function (though a strong assumption is needed for the DGP)

- Forecast optimally can be tested, for example, using a standard regression

\[ e_{t+h|t} = \alpha + \beta'Z_t + u_{t+h} \]

\[ H_0 : \beta = 0 \]
\[ H_a : \beta \neq 0 \]

- Under MSE the null would also test \( \alpha = 0 \), but under D1+L1 an optimal forecast may have \( \alpha \neq 0 \)
PT’s Proposition 2

**Proposition 2:** Let the DGP and loss function satisfy D2 and L2, and define 
\[ d_{t+h|t}^* \equiv e_{t+h|t}^* / \sigma_{t+h|t}. \] Then:

(1) The optimal forecast takes the following form:

\[ \hat{Y}_{t+h|t} = \mu_{t+h|t} + \sigma_{t+h|t} \cdot \gamma_h \]

where \( \gamma_h^* \) is a constant, depending only on \( F_{\eta,h} \) and the loss function \( L \).

(2) \( d_{t+h|t}^* \) is independent of all \( Z_t \in \mathcal{F}_t \). In particular,

\[ \text{Cov} \left[ d_{t+h|t}^{*r}, d_{t+h-j|t-j}^{*s} \right] = 0 \quad \forall j \geq h \]

for all \( (r, s) \) such that the covariance exists.
PT’s Proposition 2, proof 1

Assumption D2 yields

\[ Y_{t+h} = \mu_{t+h|t} + \sigma_{t+h|t} \eta_{t+h}, \quad \text{where} \quad \eta_{t+h|\mathcal{F}_t} \sim F_{\eta,h}(0,1) \]

and w.l.o.g., represent a forecast as

\[ \hat{Y}_{t+h} = \mu_{t+h|t} + \sigma_{t+h|t} \cdot \hat{\gamma}_{t+h|t} \]

Then...
PT's Proposition 2, proof II

\[
\hat{\gamma}^*_{t+h|t} \equiv \arg\min_{\hat{y}} \int L(y - \hat{y}) \, dF_{t+h|t}(y)
\]

\[
= \arg\min_{\hat{y}} \int \frac{1}{g \left( \frac{1}{\sigma_{t+h|t}} \right)} \left( \frac{1}{\sigma_{t+h|t}} (y - \hat{y}) \right) \, dF_{t+h|t}(y)
\]

\[
= \arg\min_{\hat{y}} \int L \left( \frac{1}{\sigma_{t+h|t}} (y - \hat{y}) \right) \, dF_{t+h|t}(y)
\]

\[
= \mu_{t+h|t} + \sigma_{t+h|t} \times \left\{ \arg\min_{\hat{\gamma}} \int L \left( \frac{1}{\sigma_{t+h|t}} \left( \mu_{t+h|t} + \sigma_{t+h|t} \eta_{t+h} \right. \right. \\
- \mu_{t+h|t} - \sigma_{t+h|t} \cdot \hat{\gamma}_{t+h|t} \left. \right) \right) \, dF_{\eta,h}(\eta) \right\}
\]

\[
= \mu_{t+h|t} + \sigma_{t+h|t} \times \left\{ \arg\min_{\hat{\gamma}} \int L \left( \eta_{t+h} - \hat{\gamma} \right) \, dF_{\eta,h}(\eta) \right\}
\]

\[
\equiv \mu_{t+h|t} + \sigma_{t+h|t} \times \gamma^*_h
\]

since \( F_{\eta,h} \) does not depend on \( t \) by assumption D2. Then:
The optimal standardised forecast error, $d^*_{t+h|t}$, is

$$d^*_{t+h|t} \equiv \frac{Y_{t+h} - \hat{Y}^*_{t+h|t}}{\sigma_{t+h|t}}$$

$$= \frac{\mu_{t+h|t} + \sigma_{t+h|t}\eta_{t+h} - \mu_{t+h|t} - \sigma_{t+h|t}\gamma^*_h}{\sigma_{t+h|t}}$$

$$= \eta_{t+h} - \gamma^*_h$$

which is independent of all elements in $\mathcal{F}_t$ though generally not mean zero. □
PT’s Proposition 2, implications

- The first part of PT’s Prop 2 provides a “representation” result for an optimal forecast under assumptions (D2,L2)

  - the optimal forecast takes a simple linear form, and the impact of the unknown (infinite dimensional) loss function reduces to a single unknown scalar.

- Implementing a test of the second part of Prop 2 (that \( \text{Cov} \left[ d^*_{t+h,t}, Z_t \right] = 0 \ \forall \ Z_t \in F_t \)) is complicated by the fact that \( \sigma_{t+h,t} \) is required to obtain \( d^*_{t+h,t} \).

  - Can be overcome if the DGP follows an ARCH-in-mean structure:
    \[
    \mu_{t+h} = \beta \sigma_{t+h,t} 
    \]

  - Can also be handled by making a parametric assumption for \( \sigma_{t+h,t} \) (eg GARCH), or possibly by using a nonparametric estimator of \( \sigma_{t+h,t} \) (eg RV)
PT’s Proposition 3

**Proposition 3:** Let the DGP and loss function satisfy D1 and L1 or assumptions D2 and L2. Then:

1. The optimal forecast is such that, for all \( t \),

\[
F_{t+h|t} \left( \hat{Y}^*_{t+h|t} \right) = q_h
\]

where \( q_h \in (0, 1) \) depends only on the forecast horizon and the loss function. If \( F_{t+h|t} \) is continuous and strictly increasing then we obtain:

\[
\hat{Y}^*_{t+h|t} = F^{-1}_{t+h|t}(q_h)
\]

2. \( I_{t+h|t}^* \equiv 1 \left\{ Y_{t+h} \leq \hat{Y}^*_{t+h|t} \right\} \) is independent of all \( Z_t \in \mathcal{F}_t \).
PT’s Proposition 3, proof I

**Proof:** (1) Under assumptions D1 and L1, or assumptions D2 and L2, we know from above that

\[ \hat{Y}^*_{t+h|t} = \mu_{t+h|t} + \sigma_{t+h|t} \cdot \gamma^*_h \]

with \( \sigma_{t+h|t} \) constant under assumption D1. \( \gamma^*_h \) depends only upon the loss function and the forecast horizon. Now notice that

\[
F_{t+h,t} \left( \hat{Y}^*_{t+h|t} \right) \equiv \Pr \left[ Y_{t+h} \leq \hat{Y}^*_{t+h|t} | \mathcal{F}_t \right] \\
= \Pr \left[ \mu_{t+h|t} + \sigma_{t+h|t} \eta_{t+h} \leq \mu_{t+h|t} + \sigma_{t+h|t} \cdot \gamma^*_h | \mathcal{F}_t \right] \\
= \Pr \left[ \eta_{t+h} \leq \gamma^*_h | \mathcal{F}_t \right] \\
\equiv q^*_h \quad \forall \ t
\]

Thus \( \hat{Y}^*_{t+h|t} \) is the \( q^*_h \) conditional quantile of \( Y_{t+h|\mathcal{F}_t} \forall \ t \). Note that \( q^*_h \) is only a function of the loss function and the forecast horizon.
(2) Since \( I_{t+h,t}^* \equiv 1 \left\{ Y_{t+h} \leq \hat{Y}_{t+h|t}^* \right\} \) is a binary random variable and

\[
\Pr \left[ I_{t+h,t}^* = 1 \mid \mathcal{F}_t \right] = \Pr \left[ Y_{t+h} \leq \hat{Y}_{t+h|t}^* \mid \mathcal{F}_t \right] \\
= \Pr \left[ \eta_{t+h} \leq \gamma_{h|t}^* \mid \mathcal{F}_t \right] \\
= q_h^* \forall t
\]

we thus have that \( I_{t+h,t}^* \) is independent of all \( Z_t \in \mathcal{F}_t \). \( \blacksquare \)
PT’s Proposition 3, implications

- PT’s Prop 3 shows that forecast optimality can be tested for a wide range of loss functions and DGP via a simple test on the indicator variable:

\[ I_{t+h|t} \equiv 1 \{ Y_{t+h} \leq \hat{Y}_{t+h|t} \} \]

- For example, using a standard regression

\[ I_{t+h|t} = \alpha + \beta' Z_t + u_{t+h}, \text{ for any } Z_t \in F_t \]

\[ H_0 : \beta = 0 \]

\[ H_a : \beta \neq 0 \]

or alternatively via a logit/probit regression.

- This test can be implemented using only the sequence of realisations and forecasts; no auxiliary estimates of the conditional variance are needed.
Summary

- This lecture described the three main types of economic forecasts
  - Point forecasts, Interval forecasts, Density forecasts
- We reviewed the formal definition of an optimal point forecast and its main ingredients
  - Choice of loss function, information set, horizon, etc.
- We derived the optimal forecast for a few well-known loss functions, and used that to obtain testable implications of forecast optimality
- We also considered optimal forecasts under general known, and some unknown, loss functions.
Additional references


