Forecast Assessment
Outline

1. Why Forecast?

2. Forecasting basics

3. Estimating Parameters for forecasting

4. Forecast Assessment
   (a) Comparing Forecasts/Forecasters
   (b) Comparing Forecasts/Models
1. Why Forecast?

(a) You want to know about the future (SPF, Financial Markets)

(b) You want to evaluate a model

Why use forecasting methods to evaluate models?

(i) Less prone to data mining, in-sample overfitting (indeed, in-sample overfitting leads to out-of-sample “underfitting”)

(ii) Instability

(iii) In-sample methods too difficult

Recently developed forecast assessment methods have focused on both of these (and it is useful to keep goals in mind as we discuss methods)
2. Forecasting Basics

$Y_{t+h}$: variable to be forecast

$X_t$: vector of variables used to make forecast (typically would include current and lags of $Y_t$ and other variables).

$f_{t+h/t}$: Forecast of $Y_{t+h}$ made a time $t$.

$e_{t+h} : Y_{t+h} - f_{t+h/t}$

$L(e)$: Loss associated with the error

“Risk” associated with $f: E(L(e))$. 

(a) Minimum MSE forecasts: Loss quadratic
\[ L(e) = a + be^2 \] and Risk is Mean Squared Error (MSE). Goal is to find mind minimum MSE (MMSE) forecast.

Key Result: MMSE forecast is the regression function:
\[ f_{t+h/t} = E(Y_{t+h}|X_t) \]

Some properties of MMSE forecasts:

(1) \[ E(e_{t+h} | X_t) = 0, \text{ so that } E(e_{t+h}X_t) = 0 \]

(2) If \( X_t \) includes current and past values of \( Y_t \), then it implicitly includes current and lagged values of \( e_t \). Thus \( E(e_{t+h}e_t) = 0 \), so that \( e_{t+h} \) follows an MA(\( h-1 \)) process.

(3) \( Y_{t+h} = f_{t+h/t} + e_{t+h}, \) where the two rhs terms are uncorrelated. Thus,
\[ \sigma_Y^2 = \sigma_f^2 + \sigma_e^2, \text{ and } \sigma_Y^2 \geq \sigma_f^2. \]
(b) More than one forecast:

Let $f^1$ and $f^2$ denote two forecasts of $Y$. Let $f^3 = \beta_0 + \beta_1 f^1 + \beta_2 f^2$ denote a third forecast.

Question: How should $f^1$ and $f^2$ be “combined” to form $f^3$ (what values of $\beta$ should be chosen)?

Answer: MMSE forecasts are regressions, so $\beta$'s are given by the (population) values from the linear regression of $Y_{t+h}$ onto $f^1_{t+h/t}$ and $f^2_{t+h/t}$. (Bates and Granger (1969), Granger and Ramanathan (1984)).

The extension to $n \geq 2$ forecasts is obvious.

References: see Timmermann (2006)
Problem: As a practical matter you must use estimates of the $\beta$’s from a sample regression.

Forecast combining “puzzle”: When the number of forecasts to be combined ($n$) is even moderately large, forecasts constructed using estimated $\beta$’s don’t perform very well. Better to use ad hoc averages like sample means, medians, trimmed means, “consensus” forecasts and so forth. (Large literature surveyed in Timmermann.) JHS will say more about this in Lecture 12.
(c) Other Loss Functions:
Granger (1969) and Christoffersen and Diebold (1997).

If \( \text{Risk}(f) = E(L(Y-f)) \), and \( Y_{t+h} \mid X_t \sim N(\mu_{t+h/t}, \sigma^2_{t+h/t}) \), then the optimal forecast is \( f_{t+h/t} = \mu_{t+h/t} + \alpha(\sigma^2_{t+h/t}) \).

Proof: Write \( Y_{t+h} = \mu_{t+h/t} + e_{t+h} \) and \( f_{t+h/t} = \mu_{t+h/t} + \alpha_{t+h/t} \). Then
\[ E(L(Y-f)) = E(L(e-\alpha)) \], and the probability density of \( L(e-\alpha) \) depends only on \( \sigma^2_{t+h/t} \).

Implication: In condition homoskedastic Gaussian world, optimal forecasts are MMSE forecasts \( (\mu_{t+h/t}) + \text{constant} \).
Elliott, Komunjer and Timmermann (2005):

\[ L(Y-f) = [\alpha + (1-2\alpha) \times 1(Y-f < 0)] |Y-f|^p \]

Non-quadratic, non-symmetric, …

(They study properties of optimal linear forecasters \( f_{t+h}/t = \theta'X_t \), with an aim to robustifying tests for forecast efficiency.)
Estimating Parameters for Use in Forecasting Models:

(1) Should you use MLE or other estimators?

Example: AR(1) model: \( y_t = \phi y_{t-1} + \varepsilon_t \)

Goal: forecast \( y_{t+2} \)

Optimal forecast: \( f_{t+2/t} = \beta y_t \), where \( \beta = \phi^2 \)

Two ways to forecast \( y_{t+2} \)

“Iterated”: Estimate \( \phi \) from \( y_t = \phi y_{t-1} + \varepsilon_t \), and use \( \hat{f}_{t+2/t}^{\text{iterated}} = \hat{\phi}^2 y_t \)

“Direct”: Estimate \( \beta \) from \( y_t = \beta y_{t-2} + u_t \), and use \( \hat{f}_{t+2/t}^{\text{direct}} = \hat{\beta} y_t \)
Pros and Cons:

\( \hat{\phi} \) is MLE, and is “efficient” under AR(1), but what if model is misspecified?

\( \hat{\beta} \) has larger variance than \( \hat{\phi} \) under correct specification, but is robust to misspecification (for the class of forecasts under consideration).


170 Monthly U.S. Macro Series, 1959-2002
Pseudo-out-of-sample forecasts (POOS)
AR and Bivariate VAR for \( h = 3, 6, 12, 24 \)
Relative pseudo-out-of-sample MSE = \[ \frac{\sum (e_{t+h}^{\text{direct}})^2}{\sum (e_{t+h}^{\text{iterated}})^2} \]

<table>
<thead>
<tr>
<th>Lag Length</th>
<th>Horizon</th>
<th>3</th>
<th>6</th>
<th>12</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(4)</td>
<td></td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
<td>1.05</td>
</tr>
<tr>
<td>AR(12)</td>
<td></td>
<td>1.01</td>
<td>1.01</td>
<td>1.03</td>
<td>1.10</td>
</tr>
<tr>
<td>AR(BIC)</td>
<td></td>
<td>0.98</td>
<td>0.97</td>
<td>0.99</td>
<td>1.05</td>
</tr>
<tr>
<td>AR(AIC)</td>
<td></td>
<td>1.00</td>
<td>1.01</td>
<td>1.02</td>
<td>1.09</td>
</tr>
</tbody>
</table>

Across Lag-Length Methods: AIC Iterated seemed to work best.
Estimating Parameters for Use in Forecasting Models:

(2) Should you use “Real-Time Data”?

\[ y_{t+h} = \beta'x_t + u_{t+h} \]

Data Revisions in \( y \) and \( x \). Issues to think about:

(a) \( y \): what the goal of forecasting (first release, final release, … )

(b) \( x \): real-time forecasting using real-time data, so it seems that \( x^{\text{initial}} \) should be used in regressions to estimate \( \beta \).

Key question: is projection of \( y \) on \( x \) the same as projection of \( y \) on \( x^{\text{initial}} \)?

\[ x = x^{\text{initial}} + x^{\text{revision}} \] … is revision “News” or ‘Noise” in the Mankiw, Runkle, and Shapiro (1984) dichotomy.
Forecast Assessment – Part 1 (Evaluating Forecasts/Forecasters … Not Models)

(a) Mincer-Zarnowitz (1969) Regressions: \( Y_{t+h} = \alpha + \beta f_{t+h/t} + \gamma W_t + u_{t+h} \)

If \( f_{t+h/t} \) is MMSE forecast then \( \alpha = 0, \beta = 1, \gamma = 0. \) Test all or subset.

Inference Issues:

(i) \( h > 1, u_{t+h} \sim \text{MA}(h-1) \) under \( H_0. \) Use HAC SEs.

(ii) If \( Y_t \) is persistent (say \( I(1) \)), then so is \( f_t \), so “unit-root” regression inference problems. Easy fix: \( (Y_{t+h} - Y_t) = \alpha + \beta(f_{t+h/t} - Y_t) + \gamma W_t + u_{t+h}. \)

(iii) \( h \) large, \( u \) is very persistent. HAC works poorly. (Richardson and Stock (1989)).
(b) Combining or “Encompassing” Regressions

2 forecasts $f^1$ and $f^2$.

Forecast combining regression: $Y_{t+h} = \beta_0 + \beta_1 f^1_{t+h/t} + \beta_2 f^2_{t+h/t} + u_{t+h}$

If $f^1$ is MMSE forecast, then $\beta_0 = \beta_2 = 0$ and $\beta_1 = 1$.

Inference Issues: − Same as MZ regressions. Also, you might want to make alternative more parsimonious imposing $\beta_0 = 0$ and $\beta_1 + \beta_2 = 1$. Imposing this constraint yields

$$Y_{t+h} - f^1_{t+h/t} = \beta_2(f^2_{t+h/t} - f^1_{t+h/t}) + u_{t+h}.$$
(c) Loss-Function Tests

2 forecasts $f^1$ and $f^2$ with forecast errors $e^1$ and $e^2$.

Question: Is $E(L(e^1)) = E(L(e^2))$ (or < or >)?

Testing using quadratic loss. Realized forecast errors $e^1_t$ and $e^2_t$.

Question: $E[(e^1_t)^2] = E[(e^2_t)^2]$ or $E[(e^1_t)^2 - (e^2_t)^2] = 0$?

Sample moments of differences: $\overline{d} = \frac{1}{T} \sum_{t=1}^{T} d_t$, where $d_t = (e^1_t)^2 - (e^2_t)^2$

Is $\overline{d}$ statistically significantly different from zero.

Inference Issues: (i) $h > 1$, serial correlation. Use HAC se’s.

Refs: Diebold and Mariano (1995), see West (2006) for history
(d) Loss function tests with many competing forecasts:

\( f^1 \) is a benchmark model (say, random walk)

\( f^k, k = 1, \ldots, n \) are \( n \) competing models

Question: Do any of the \( n \) competing models dominate the benchmark?

\[
\bar{d}_k = \frac{1}{T} \sum_{t=1}^{T} d^k_t, \text{ where } d^k_t = \{ (e^1_t)^2 - (e^k_t)^2 \}, \text{ so that } \bar{d}_k \text{ is the sample MSE improvement for model } k \text{ over benchmark prediction.}
\]

Test Statistic: \( RC = \max_k \bar{d}_k \) (RC for “Reality Check”)

White (2000) derives limiting distribution of RC under null that benchmark is optimal, so that critical values can be computed.

(See Hansen (2005) for a refinement.)
Density Forecasts:

Chart 2  CPI inflation projection based on market interest rate expectations

Percentage increase in prices on a year earlier

Lecture 10 - 18, July 21, 2008
Evaluating Density Forecasts

Ref: Diebold, Gunther, Tay (1998)

Key Insight: Suppose $Y$ has CDF $F$. Then $U = F(Y) \sim U[0,1]$. (Recall random number generators often use $Y = F^{-1}(U)$.)

Thus, if $F_{t+h/t}$ is the conditional CDF of $Y_{t+h/t}$ then $U_{t+h} = F_{t+h/t}(Y_{t+h})$ should be $U[0,1]$ and $U_{t+h}$ should be independent of data dated $t$ and earlier.

Interval Forecasts: “Provide a 95% ‘confidence interval’” for GDP growth in the final quarter of 2008. (ref: Christoffersen (1998)).
Forecast Assessment – Part 2 (Evaluating Models using POOS)

Key References: West (1996, 2006)

Key Statistical Difference Between this and Part 1: Explicitly account for sampling variability in estimated model parameters.

Setup: Model 1 has forecasts \( f^1(\theta_1) \) and Model 2 has forecasts \( f^2(\theta_2) \). The forecasts that are evaluated are based on estimated models \( f^1(\hat{\theta}_1) \) and \( f^2(\hat{\theta}_2) \).
Pseudo-Out-Of Sampling (POOS) Forecasting Strategy:


Strategy:
(i) Estimate $\theta$ using observations 1:$R$, $\hat{\theta}_R$
(ii) Forecast $Y_{R+h}$ using data through $R$ and $\hat{\theta}_R$.

then (recursive POOS)
(iii) Estimate $\theta$ using observations 1:$R+1$, $\hat{\theta}_{R+1}$
(iv) Forecast $Y_{R+1+h}$ using data through $R+1$ and $\hat{\theta}_{R+1}$.

or (rolling POOS)
(iii) Estimate $\theta$ using observations 2:$R+1$, $\hat{\theta}_{R+1}$
(iv) Forecast $Y_{R+1+h}$ using data through $R+1$ and $\hat{\theta}_{R+1}$.

…
Note (and this is important):

The Question of interest is:

(i) \( E[L(Y - f^1(\theta_1))] = E[L(Y - f^2(\theta_2))] \) ( or < or >)

NOT:

(ii) \( E[L(Y - f^1(\hat{\theta}_1))] = E[L(Y - f^2(\hat{\theta}_2))] \) ( or < or >) (but I will return to this)

Many interesting cases in which

\[ E[L(Y - f^1(\hat{\theta}_1))] < E[L(Y - f^2(\hat{\theta}_2))] \]

but \( E[L(Y - f^1(\theta_1))] > E[L(Y - f^2(\theta_2))] \)

(Example: Random Walk in Exchange Rates … Engel and West (2005), Clark and West (2006), Rossi (2005).)

Focus here is on risk comparisons. Related issues arise in combining tests.
A Complication: Nested vs. Non-nested models

Nested Models:

Model 1: $y_{t+1} = x_t'\beta + \varepsilon_{t+1}$ is a special case of

Model 2: $y_{t+1} = x_t'\beta + z_t'\gamma + e_{t+1}$

Non-Nested Models

Model 1: $y_{t+1} = x_t'\beta + \varepsilon_{t+1}$ is not a special case of

Model 2: $y_{t+1} = z_t'\gamma + e_{t+1}$
Non-Nested Models (West (1996)):

Model 1: \( y_{t+1} = x_t \beta + \epsilon_{t+1} \) (scalar \( x \), \( h = 1 \)-step-ahead for convenience)

True Forecast: \( x_t \beta \)

Estimated Forecast: \( x_t \hat{\beta}_t \)

True Forecast error: \( \epsilon_{t+1} \)

Estimated error: \( Y_{t+1} - x_t \hat{\beta}_t = \hat{\epsilon}_{t+1} = \epsilon_{t+1} + x_t(\hat{\beta}_t - \beta) \)

Model 2: \( y_{t+1} = z_t \gamma + e_{t+1} \)

True Forecast: \( z_t \gamma \)

Estimated Forecast: \( z_t \hat{\gamma}_t \)

True Forecast error: \( e_{t+1} \)

Estimated error: \( Y_{t+1} - z_t \hat{\gamma}_t = \hat{e}_{t+1} = e_{t+1} + z_t(\hat{\gamma}_t - \gamma) \)
Recall that Loss Function Test looks at averages of $d_t = (\varepsilon_t^2 - e_t^2)$ over the prediction period. Here, we must use $\hat{d}_t = (\hat{\varepsilon}_t^2 - \hat{e}_t^2)$. How are the sample averages of $\hat{d}_t$ and $d_t$ related?
A Calculation:

\[ \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} (\hat{\varepsilon}_t^2 - \hat{\varepsilon}_t^2) = \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} (\varepsilon_t^2 - e_t^2) \]

\[ + \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} (\hat{\beta}_{t-1} - \beta)^2 x_{t-1}^2 + 2 \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} (\hat{\beta}_{t-1} - \beta)x_{t-1}\varepsilon_t \]

\[ + \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} (\hat{\gamma}_{t-1} - \gamma)^2 z_{t-1}^2 + 2 \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} (\hat{\gamma}_{t-1} - \gamma)z_{t-1}\varepsilon_t \]

\[ = \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} (\varepsilon_t^2 - e_t^2) + o_p(1) \]

where (from West (1996)) the final inequality holds when the \( E(\varepsilon_t x_{t-1}) = E(e_t z_{t-1}) = 0 \) (and some technical assumptions are satisfied).

Bottom Line: \( \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} (\hat{\varepsilon}_t^2 - \hat{\varepsilon}_t^2) \approx \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} (\varepsilon_t^2 - e_t^2) \) (so sampling error if \( \hat{\beta} \) and \( \hat{\gamma} \) do not matter.)

Identity (from last page):

\[
\sum_{t=R+1}^{T} (\hat{\varepsilon}_t^2 - \hat{\varepsilon}_t^2) = \sum_{t=R+1}^{T} (\varepsilon_t^2 - e_t^2) \\
+ \sum_{t=R+1}^{T} (\hat{\beta}_{t-1} - \beta)^2 x_{t-1}^2 + 2 \sum_{t=R+1}^{T} (\hat{\beta}_{t-1} - \beta)x_{t-1}\varepsilon_t \\
+ \sum_{t=R+1}^{T} (\hat{\gamma}_{t-1} - \gamma)^2 z_{t-1}^2 + 2 \sum_{t=R+1}^{T} (\hat{\gamma}_{t-1} - \gamma)z_{t-1}e_t
\]

Non-Nested Model: \( \sum_{t=R+1}^{T} (\varepsilon_t^2 - e_t^2) \sim O_p(P^{1/2}) \) dominates rhs

Nested Models: \( y_{t+1} = x_t'\beta + \varepsilon_{t+1} \) and \( y_{t+1} = x_t'\beta + z_t'\gamma + e_{t+1} \)

under equal loss, models are the same, so that \( \varepsilon_t = e_t \). First term vanishes.
This is harder … Papers by McCracken and Clark and McCracken attack this by studying behavior of terms on rhs.

Key results: Limits are (somewhat complicated/messy) functions of normals.

Implication: Can use parametric bootstrap (Gaussian errors) to approximate limiting distribution, compute critical values and so forth.
An Important Special Case: $f^1$ is a random walk forecast, $f^2$ is nested within the random walk. (Clark and West (2006)).

Let $y_t$ denote the first difference of a variable of interest (e.g., an exchange rate). Under the null, $y_t$ is mds, while under the alternative it can be predicted by $x_t$.

Model 1: $y_{t+1} = \varepsilon_{t+1}$
Forecast 1: $\hat{f}^1_{t+1/t} = 0$
Errors (under H$_0$): $\hat{e}_{t+1}^1 = \varepsilon_{t+1}$

Model 2: $y_{t+1} = x_t' \beta + \varepsilon_{t+1}$
Forecast 2: $\hat{f}^2_{t+1/t} = x_t' \hat{\beta}_t$
Errors (under H$_0$): $\hat{e}_{t+1}^2 = \varepsilon_{t+1} - x_t' \hat{\beta}_t$

MSPE (under H$_0$):
\[
\frac{1}{P} \sum_{t=R+1}^{P} \varepsilon_{t+1}^2 + \frac{1}{P} \sum_{t=r+1}^{P} (x_t' \hat{\beta}_t)^2 - 2 \frac{1}{P} \sum_{t=r+1}^{P} \varepsilon_{t+1} x_t' \hat{\beta}_t
\]

Expectation:
\[
\sigma^2 + E\left( \frac{1}{P} \sum_{t=r+1}^{P} (x_t' \hat{\beta}_t)^2 \right)
\]
Thus, under RW null

\[
\text{E(MSE for RW)} = \text{E(MSE for alternative)} - E\left(\frac{1}{P} \sum_{t=r+1}^{P} (x_t' \hat{\beta}_t)^2\right)
\]

RW does not suffer from overfitting \((x_t' \hat{\beta}_t)\) and should forecast **better** than alternative … and … the “overfitting” term can be estimated:

\[
E\left(\frac{1}{P} \sum_{t=r+1}^{P} (x_t' \hat{\beta}_t)^2\right) \approx \frac{1}{P} \sum_{t=r+1}^{P} (x_t' \hat{\beta}_t)^2 = \frac{1}{P} \sum_{t=r+1}^{P} (f_{tit-1}^2)^2
\]

Clark-West test: standardized version of

\[
\hat{\sigma}_1^2 - \left(\hat{\sigma}_2^2 - \frac{1}{P} \sum_{t=r+1}^{P} (f_{tit-1}^2)^2\right),
\]

where \(\hat{\sigma}_1^2\) and \(\hat{\sigma}_2^2\) are the POOS mean square prediction errors for \(f^1\) (the RW) and \(f^2\).
What is the distribution of this difference?

\[
\sqrt{P} \left( \hat{\sigma}_1^2 - \left( \hat{\sigma}_2^2 - \frac{1}{P} \sum_{t=r+1}^{P} (f_{t/t-1}^2) \right) \right) = \frac{1}{\sqrt{P}} \sum_{t=r+1}^{P} \varepsilon_t f_{t/t-1}^2
\]

And \( \varepsilon_t f_{t/t-1}^2 \) is a mds under the null. Maybe a normal limit?

Caution: (example) \( x_t = 1, f_{t/t-1}^2 = \hat{\beta}_{t-1} = \frac{1}{t-1} \sum_{i=1}^{t-1} y_i = \frac{1}{t-1} \sum_{i=1}^{t-1} \varepsilon_i \), and

\[
\varepsilon_t f_{t/t-1}^2 \overset{H_0}{=} \frac{1}{t-1} \varepsilon_t \sum_{i=1}^{t-1} \varepsilon_i.
\]

Remember unit root AR Model, numerator of \( \hat{\rho} - \rho \) is \( \sum_{t=1}^{T} \varepsilon_t \sum_{i=1}^{t-1} \varepsilon_i \).

CW: Using “Rolling” estimate of \( \beta \) based on \( R \) observations to limit this dependence. (Where \( R \) is fixed and not too large – Giacomini and White (2001), discussed below.)
CW Empirical Example:

Exchange Rates ($’s) Monthly: Canada, Japan, Switzerland, UK

POOS Period 1990-2003 ($P = 166$), use $R = 120$.

$x = (1, 1$-month interest differential)$
<table>
<thead>
<tr>
<th>Country</th>
<th>Prediction sample</th>
<th>(3) $\sigma_1^2$</th>
<th>(4) $\sigma_2^2$</th>
<th>(5) adj.</th>
<th>(6) $\sigma_2^2$-adj.</th>
<th>(7) MSPE-adjusted $\hat{\sigma}_1^2 - (\hat{\sigma}_2^2 - \text{adj.})$</th>
<th>(8) MSPE-normal $\hat{\sigma}_1^2 - \hat{\sigma}_2^2$</th>
<th>(9) CCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>1990:1–2003:10</td>
<td>2.36</td>
<td>2.32</td>
<td>0.09</td>
<td>2.22</td>
<td>0.13</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.08)</td>
<td>1.78**</td>
<td>0.54††</td>
</tr>
<tr>
<td>Japan</td>
<td>1990:1–2003:10</td>
<td>11.32</td>
<td>11.55</td>
<td>0.75</td>
<td>10.80</td>
<td>0.53</td>
<td>-0.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.43)</td>
<td>1.24</td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>1985:1–2003:10</td>
<td>12.27</td>
<td>12.33</td>
<td>0.96</td>
<td>11.37</td>
<td>0.90</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.48)</td>
<td>1.88**</td>
<td></td>
</tr>
<tr>
<td>U.K.</td>
<td>1985:1–2003:10</td>
<td>9.73</td>
<td>10.16</td>
<td>0.44</td>
<td>9.72</td>
<td>0.01</td>
<td>-0.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.33)</td>
<td>0.03</td>
<td>-1.27</td>
</tr>
</tbody>
</table>
1. Forecast Assessment – Part 1 (Evaluating Forecasts/Forecasters)

\[ E[L(Y - f^1)] = E[L(Y - f^2)] \ ( \text{or} < \text{or} >) \]

2. Forecast Assessment – Part 2 (Evaluating Models using POOS forecasting)

\[ E[L(Y - f^1(\theta_1))] = E[L(Y - f^2(\theta_2))] \ ( \text{or} < \text{or} >) \]

3. Forecasting Assessment – Part 3 (Evaluating Forecasting Models using POOS forecasting)

\[ E[L(Y - \hat{f}^1(\hat{\theta}_1))] = E[L(Y - \hat{f}^2(\hat{\theta}_2))] \ ( \text{or} < \text{or} >) \]

Giacomini and White (2006)
Giacomini and White (2006):

<table>
<thead>
<tr>
<th>Data</th>
<th>Forecasting Procedure</th>
<th>Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>{x_i, y_i}_{i=1}^t</td>
<td>(\hat{\theta}_1) and whatever</td>
<td>(f_{t+1/t}^1)</td>
</tr>
<tr>
<td>{x_i, y_i}_{i=1}^t</td>
<td>(\hat{\theta}_2) and whatever</td>
<td>(f_{t+1/t}^2)</td>
</tr>
</tbody>
</table>

Question: \(E\left[ L(f_{t+1/t}^1) - L(f_{t+1/t}^2) \mid g_t \right] = 0 \) (or \(<\text{ or }>\))

(Using time \(t\) information, can I predict which forecast will have smaller loss?)

GMM Problem … \(E\left[ \left( L(f_{t+1/t}^1) - L(f_{t+1/t}^2) \right) q(g_t) \right] = 0\)
Moment: \( E \left[ \left( L(f_{t+1/t}^1) - L(f_{t+1/t}^2) \right) q(g_t) \right] = 0 \)

Forecasting Model 1: \( \{x_i, y_i\}_{i=1}^t \rightarrow \hat{\theta}_1 \) and whatever \( \rightarrow f_{t+1/t}^1 \)

Forecasting Model 2: \( \{x_i, y_i\}_{i=1}^t \rightarrow \hat{\theta}_2 \) and whatever \( \rightarrow f_{t+1/t}^2 \)

Inference Issues:

(1) Stationarity: using rolling samples \( \{x_i, y_i\}_{i=t-R}^t \)

(2) \( h \)-step-ahead prediction: HAC (allowing for \( h \) dependence)

(3) Unconditional comparisons (\( q(g_t) = 1 \)): Much more dependence (at least \( R \)) … method is (arguably) less useful. (Note: This is a different question than the question that motivated their analysis.)