Applied Econometrics for Macro and Finance

Time Series Forecasting

Shiu-Sheng Chen

Department of Economics
National Taiwan University

Spring 2023

Part I

Time Series Forecasting

Using Regression Models for Forecasting

Forecasting and estimation of causal effects are quite different objectives.

- External validity is paramount: the model estimated using historical data must hold into the (near) future.
- Hence, stationarity is a key requirement for external validity of time series regression.
- Omitted variable bias may be less of a concern. Knowing that the number of people carrying umbrellas is a good predictor of whether or not it will rain in the afternoon is useful enough for me to forecast the weather, but I can't explain it (or I don't care about the explanation).

Using Regression Models for Forecasting

Forecasting and estimation of causal effects are quite different objectives.

- Statisticians do not worry about interpreting coefficients in forecasting models (but economists somewhat do).
- Why? Knowing the economic reasons of predictability helps understand the limitation of our predictive model.
  - For example, if people decide that they don't want to be exposed to the sun and start carrying umbrellas on rainy and sunny days, then the forecasting ability breaks down.

Time Series Forecasting Models

- We will focus on dynamic predictive model:

$$Y_{t+1} = x_t'\beta + \varepsilon_{t+1}$$

- If $x_t = (1, Y_t, Y_{t-1}, \ldots, Y_{t-p+1})$, we obtain an AR(p) predictive model:

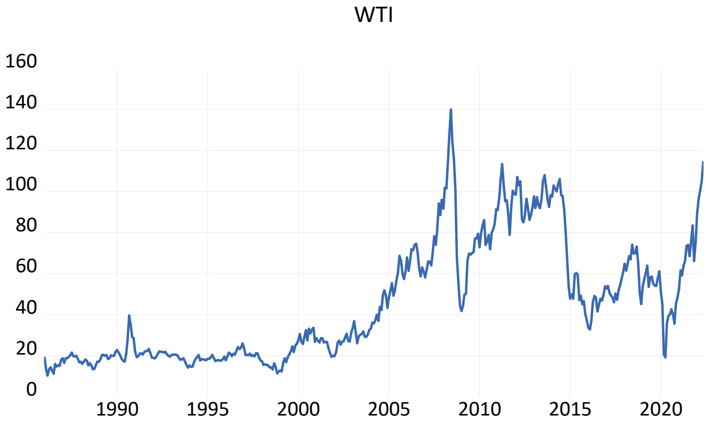
$$Y_{t+1} = \beta_0 + \beta_1 Y_t + \cdots + \beta_p Y_{t-p+1} + \varepsilon_{t+1}$$

- If $x_t = (1, Y_t, Y_{t-1}, \ldots, Y_{t-p+1}, Z_t, Z_{t-1}, \ldots, Z_{t-q+1})$, we obtain an ADL(p,q) predictive model:

$$\begin{aligned} Y_{t+1} = \beta_0 + \beta_1 Y_t + \cdots + \beta_p Y_{t-p+1} \\ + \delta_1 Z_t + \cdots + \delta_q Z_{t-p+1} + \varepsilon_{t+1} \end{aligned}$$

Example: Forecasting Oil Prices

- Monthly data from 1986:M1 to 2022:M5. The monthly change rate is calculated by $\Delta \text{WTI}_t = [\log(\text{WTI}_t) - \log(\text{WTI}_{t-1})] \times 100$



AR(1) Model of the Change in WTI Prices

EViews - [Equation: EQ1 Workfile: ENERGYMETRICS::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RWTI
Method: Least Squares
Date: 03/10/23 Time: 22:21
Sample (adjusted): 1986M03 2022M05
Included observations: 435 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.443209	0.459602	0.964334	0.3354
RWTI(-1)	0.134102	0.056768	2.362289	0.0186

R-squared	0.018437	Mean dependent var	0.495871
Adjusted R-squared	0.016170	S.D. dependent var	10.69094
S.E. of regression	10.60416	Akaike info criterion	7.564956
Sum squared resid	48690.04	Schwarz criterion	7.583693
Log likelihood	-1643.378	Hannan-Quinn criter.	7.572351
F-statistic	8.133072	Durbin-Watson stat	1.985057
Prob(F-statistic)	0.004555	Wald F-statistic	5.580410
Prob(Wald F-statistic)	0.018604		

- The lagged change in the WTI price is a useful predictor.

The AR(p) Model: Using Multiple Lags for Forecasting

- The p-th order autoregressive model AR(p) is

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + u_t$$

- The AR(p) model uses p lags of Y as regressors
    - The AR(1) model is a special case
- The coefficients do not have a causal interpretation
- To test the hypothesis that $Y_{t-2}, \ldots, Y_{t-p}$ do not further help forecast Y_t, beyond Y_{t-1}, use an F-test.
- Determine p using an information criterion (more on this later)

Example: AR(8) Model of the Change in WTI Prices

EViews - [Equation: EQ2 Workfile: ENERGYMETRICS::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RWTI
Method: Least Squares
Date: 03/10/23 Time: 22:36
Sample (adjusted): 1986M10 2022M05
Included observations: 428 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.550078	0.510791	1.076914	0.2821
RWTI(-1)	0.164341	0.060042	2.737089	0.0065
RWTI(-2)	-0.120041	0.088856	-1.350951	0.1774
RWTI(-3)	0.016575	0.044324	0.373956	0.7086
RWTI(-4)	-0.115064	0.055692	-2.066078	0.0394
RWTI(-5)	-0.069959	0.048646	-1.438124	0.1511
RWTI(-6)	0.029133	0.051062	0.570547	0.5686
RWTI(-7)	-0.049566	0.044839	-1.105409	0.2696
RWTI(-8)	-0.009966	0.042764	-0.233056	0.8158

R-squared	0.057108	Mean dependent var	0.479364
Adjusted R-squared	0.039105	S.D. dependent var	10.44968
S.E. of regression	10.24333	Akaike info criterion	7.511934
Sum squared resid	43963.88	Schwarz criterion	7.597289
Log likelihood	-1598.554	Hannan-Quinn criter.	7.545644
F-statistic	3.172167	Durbin-Watson stat	1.992678
Prob(F-statistic)	0.001672	Wald F-statistic	1.783567
Prob(Wald F-statistic)	0.078415		

- $\bar{R}^2$ increases by adding lags 2–8 (0.039 vs. 0.016).

Example: AR(8) Model of the Change in WTI Prices

F-test $(H_0 : \beta_2 = \beta_3 = \cdots = \beta_8 = 0)$

EViews - [Equation: EQ2 Workfile: ENERGYMETRICS::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Redundant Variables Test
Equation: EQ2
Redundant variables: RWTI(-2 TO -8)
Specification: RWTI C RWTI(-1 TO -8)
Null hypothesis: RWTI(-2 TO -8) are jointly insignificant

	Value	df	Probability
F-statistic	2.166884	(7, 419)	0.0361
Likelihood ratio	15.22013	7	0.0333

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	1591.533	7	227.3618
Restricted SSR	45555.41	426	106.9376
Unrestricted SSR	43963.88	419	104.9257

LR test summary:

	Value
Restricted LogL	-1606.164
Unrestricted LogL	-1598.554

Forecasts: Terminology and Notation

- Predicted values are in-sample (a.k.a. fitted values).
- Forecasts are out-of-sample in the future.
    - $\hat{Y}_{T+1|T}$ is the forecast of Y_{T+1} based on $Y_T, Y_{T-1}, \ldots$, using the estimated coefficients, which are estimated using data through period T.
    - For an AR(1):

$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T$$

    - The one-step ahead forecast error is

$$Y_{T+1} - \hat{Y}_{T+1|T}$$

Remarks

- We used ΔWTI, not WTI, in the AR's. Why?
  - Small sample bias
  - Possible nonstationarity problem
- Recommendation for forecasting Y_{T+1} when Y_t is highly persistent:
  - Estimate AR model using ΔY_t
  - Forecast $\widehat{\Delta Y}_{T+1|T}$
  - Recover $\widehat{Y}_{T+1|T}$ from $\widehat{\Delta Y}_{T+1|T}$

Example: Forecasting WTI Prices using an AR(1) Model

- AR(1) estimated using data from 1986:M1–2022:M5

$$\widehat{\Delta \text{WTI}}_t = 0.44 + 0.13 \times \Delta \text{WTI}_{t-1}$$

- Note that

$$\begin{aligned} \Delta \text{WTI}_{2022M5} &= (\log \text{WTI}_{2022M5} - \log \text{WTI}_{2022M4}) \times 100 \\ &= (\log(114.38) - \log(104.59)) \times 100 \\ &= 8.95 \end{aligned}$$

- The forecast of $\Delta \text{WTI}_{2022M6}$ is

$$\widehat{\Delta \text{WTI}}_{2022M6|2022M5} = 0.44 + 0.13 \times 8.95 = 1.6035$$

Example: Forecasting WTI Prices using an AR(1) Model

- Hence,

$$\log \widehat{\text{WTI}}_{2022M6|2022M5} = \log \text{WTI}_{2022M5} + \frac{\widehat{\Delta\text{WTI}}_{2022M6|2022M5}}{100}$$
$$= \log(114.38) + 0.016035 = 4.75556124$$

- We can further obtain

$$\widehat{\text{WTI}}_{2022M6|2022M5} = e^{4.75556124} = 116.228867$$

- Note that $\text{WTI}_{2022M6} = 107.76$, and the forecast error is

$$\text{WTI}_{2022M6} - \widehat{\text{WTI}}_{2022M6|2022M5} = 107.76 - 116.228867 = -8.468867$$

Pseudo Out-of-Sample Forecasts

- Sample: $t = 1, 2, \ldots, T$
  - In-sample: $t = 1, 2, \ldots, R$
  - Out-of-sample: $t = R + 1, R + 2, \ldots, T$, where $P = T - R$
- One-step-ahead forecast in an AR(1) model via fixed, recursive or rolling estimation

$$\hat{Y}_{t+1|t} = \hat{\alpha}_t + \hat{\beta}_t Y_t$$

  for $t = R, R + 1, \ldots, T - 1$.
- Picks up the model with the smallest MSFE:

$$\frac{1}{P} \sum_{t=R}^{T-1} \left(Y_{t+1} - \hat{Y}_{t+1|t} \right)^2$$

  where $Y_{t+1} - \hat{Y}_{t+1|t}$ is the out-of-sample (OOS) forecast error.

Autoregressive Distributed Lag (ADL) Model

- So far we have considered forecasting models that use only past values of Y
- It makes sense to add other variables (X) that might be useful predictors of Y, above and beyond the predictive value of lagged values of Y.
- We can consider the autoregressive distributed lag model with p lags of Y and q lags of X, i.e., the ADL(p, q) model:

$$
\begin{aligned}
Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \\
+ \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \delta_q X_{t-q} + u_t
\end{aligned}
$$

Example: WTI Prices and Global Real Economic Activity

- How to choose X?
  - Economic theory
  - Conventional wisdom
- For instance, according to Chen (2014, *Economic Inquiry*) and Chen (2016, *Canadian Journal of Economics*), the oil-sensitive stock price (NYSE ARCA OIL & GAS, XOI) change is able to forecast the oil price movement.
- In particular, ΔWTI_t is expected to be related to lagged values of the oil-sensitive stock price change (ΔXOI_t), with a positive coefficient.

An ADL Model for WTI Oil Prices

EViews - [Equation: EQ3 Workfile: ENERGYMETRICS::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RWTI
Method: Least Squares
Date: 03/10/23 Time: 22:36
Sample (adjusted): 1986M10 2022M05
Included observations: 428 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.443733	0.516968	0.858337	0.3912
RWTI(-1)	0.065486	0.069549	0.941586	0.3469
RWTI(-2)	-0.097673	0.077401	-1.261914	0.2077
RWTI(-3)	0.001633	0.047824	0.034149	0.9728
RWTI(-4)	-0.114913	0.053627	-2.142823	0.0327
RWTI(-5)	-0.072736	0.047224	-1.540218	0.1243
RWTI(-6)	0.030440	0.052519	0.579610	0.5625
RWTI(-7)	-0.071635	0.045582	-1.571541	0.1168
RWTI(-8)	-0.017650	0.044151	-0.399767	0.6895
RXOI(-1)	0.285708	0.151441	1.886601	0.0599

R-squared	0.077783	Mean dependent var	0.479364
Adjusted R-squared	0.057927	S.D. dependent var	10.44968
S.E. of regression	10.14251	Akaike info criterion	7.494435
Sum squared resid	42999.85	Schwarz criterion	7.589274
Log likelihood	-1593.809	Hannan-Quinn criter.	7.531891
F-statistic	3.917298	Durbin-Watson stat	2.014957
Prob(F-statistic)	0.000085	Wald F-statistic	1.702257
Prob(Wald F-statistic)	0.086330		

- $\bar{R}^2 = 0.058$ ($\bar{R}^2 = 0.039$ for AR(8) model)

Granger Causality Test

- The variable X affects a forecast for Y if lagged values of X have true non-zero coefficients in the distributed lag model
- That is, if one of the δ's are non-zero

$$
\begin{aligned}
Y_t = \beta_0 &+ \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \\
&+ \delta_1 x_{t-1} + \delta_2 x_{t-2} + \cdots + \delta_q x_{t-q} + u_t
\end{aligned}
$$

Clive W. J. Granger

- 1934–2009
- UCSD econometrician
- Winner of 2003 Nobel Prize
- Granger causality, Spurious regression, Cointegration

Non Causality Tests

- Given ADL($p.q$) model:

$$\begin{aligned} Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \\ + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \delta_q X_{t-q} + u_t \end{aligned}$$

- Null Hypothesis: $\delta_1 = \delta_2 = \cdots = \delta_q = 0$
  - X does not Granger cause Y
- Joint F test

Non Causality Tests

Pairwise Granger Causality Tests
Date: 03/12/23 Time: 09:33
Sample: 1986M01 2022M05
Lags: 8

Null Hypothesis:	Obs	F-Statistic	Prob.
RXOI does not Granger Cause RWTI	428	2.20111	0.0265
RWTI does not Granger Cause RXOI		1.56902	0.1321

- In our example, the F-statistic is able to reject the null hypothesis
  - The oil-sensitive stock price change Granger causes changes in WTI oil prices

Caveat

When we say that X Granger causes Y,

- It does not mean causality in a mechanical sense
- Only that X predictively causes Y
  - Common factors and lead-lag relationship: Oil price vs. Oil-sensitive stock price
  - True causality could actually be the reverse

Other Candidates as Leading Indicators

- Potential candidates of X?
  - Price of Oil Futures Contracts
  - U.S. Dollar Exchange Rate
  - Crack Spreads (裂解價差, 煉油公司利潤)
  - Changes in Prices of Industrial Raw Materials (other than crude oil)
  - Expert Survey Forecasts

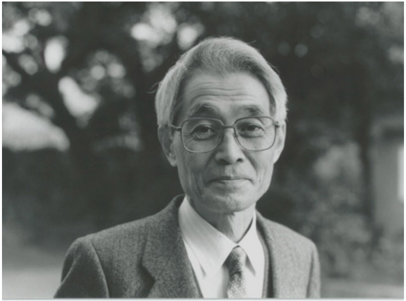
Model Selection

- The dynamic distributed lag model has p lags of Y and q lags of X, a total of $1 + p + q$ estimated coefficients
- Models (p and q) can be selected by calculating and minimizing the Akaike information criterion (AIC) or Bayesian information criterion (BIC)

$$\text{AIC} = T \log\left(\frac{SSR}{T}\right) + 2(p + q + 1)$$

$$\text{BIC} = T \log\left(\frac{SSR}{T}\right) + (p + q + 1)\log(T)$$

Hirotugu Akaike (赤池 弘次)



- 1927–2009
- Japanese statistician
- In the early 1970s he formulated an information criterion for model identification which has become known as the Akaike information criterion.

Thomas Bayes

- 1701–1761
- He is credited with inventing Bayes Theorem
- Let $\mathcal{M}_1$ be model 1, $\mathcal{M}_2$ be model 2, and D denote data

$$P(\mathcal{M}_1|D) = \frac{P(D|\mathcal{M}_1)}{P(D|\mathcal{M}_1)P(\mathcal{M}_1) + P(D|\mathcal{M}_2)P(\mathcal{M}_2)}$$

Bayesian information criterion

- It has been shown that

$$P(\mathcal{M}_1|D) = \frac{P(D|\mathcal{M}_1)}{P(D|\mathcal{M}_1)P(\mathcal{M}_1) + P(D|\mathcal{M}_2)P(\mathcal{M}_2)}$$
$$\propto \exp\left(-\frac{T}{2} \cdot \text{BIC}\right)$$

where

$$\text{BIC} = T \log\left(\frac{SSR}{T}\right) + (p + q + 1)\log(T)$$

$$SSR = \sum_{t=1}^{T}(Y_t - \hat{Y}_t)^2$$

- The Bayes method is to select the model with the highest posterior probability, i.e., the model with the smallest value of BIC

Trade-off in AIC/BIC

$$\text{AIC} = T\log\left(\frac{SSR}{T}\right) + 2(p+q+1)$$

$$\text{BIC} = T\log\left(\frac{SSR}{T}\right) + (p+q+1)\log(T)$$

- When we compare models, the larger model (the model with more lags p and q) will have
  - Smaller SSR
  - Larger p and q
- The AIC/BIC trades these off
  - The first term is decreasing in p and q
  - The second term is increasing in p and q

BIC vs. AIC

$$\text{AIC} = T \log\left(\frac{SSR}{T}\right) + 2(p+q+1)$$

$$\text{BIC} = T \log\left(\frac{SSR}{T}\right) + (p+q+1)\log(T)$$

- Both criteria make similar trade-offs
  - The AIC penalty is 2
  - The BIC penalty is $\log(T) > 2$ (if $T > 7$)
- For example, if $T = 240$, $\log(T) = 5.5$ is much larger than 2
- Unlike BIC, the AIC is designed to find models with low forecast risk

BIC vs. AIC

- In general, AIC typically selects a larger model than BIC
- Mechanically, it is because BIC puts a larger penalty on the dimension of the model ($\log(T)$ vs. 2)
- Conceptually, it is because
  - BIC assumes that there is a true finite model, and is trying to find the true model
  - AIC assumes all models are approximations, and is trying to find the model which makes the best forecast.
    - Extra lags are included if they help to forecast
- Finally, a sophisticated selection method is to compute the pseudo out-of-sample forecasts and forecast errors, and pick the model with the smallest out-of-sample forecast variance

Part II

Forecasting Evaluation

Measures of Forecast Accuracy

- Let $\hat{Y}_{t,t+h}$ be the h-step-ahead forecast of Y_{t+h}, and $\hat{e}_{t,t+h}$ be corresponding h-step-ahead forecast error

$$\hat{e}_{t,t+h} = Y_{t+h} - \hat{Y}_{t,t+h}$$

- The crucial object in measuring forecast accuracy is the loss function, $L(Y_{t+h}, \hat{Y}_{t,t+h})$, often restricted to $L(\hat{e}_{t,t+h})$, which is the loss, cost, or disutility associated with various pairs of forecasts and realizations.

Important and Popular Accuracy Measures

- Under quadratic loss, we have the mean squared error (MSE)

$$\text{MSE} = E(\hat{e}^2_{t,t+h})$$

- The sample counterpart is

$$\widehat{\text{MSE}} = \frac{1}{T}\sum_{t=1}^{T} \hat{e}^2_{t,t+h}$$

Important and Popular Accuracy Measures

- We sometimes take square roots to preserve units, yielding the root mean squared error

$$\text{RMSE} = \sqrt{E(\hat{e}^2_{t,t+h})}$$

- The sample counterpart is

$$\widehat{\text{RMSE}} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\hat{e}^2_{t,t+h}}$$

- Suppose that the forecast errors are measured in dollars. Then the mean squared error, which is built up from squared errors, is measured in dollars squared. Taking square roots brings the units back to dollars.

Important and Popular Accuracy Measures

- Under absolute loss, we have the mean absolute error (MAE)

$$\text{MAE} = E\left|\hat{e}_{t,t+h}\right|$$

- The sample counterpart is

$$\widehat{\text{MAE}} = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{e}_{t,t+h}\right|$$

- When using MAE we don't have to take square roots to preserve units.

Benchmark Comparisons

It is sometimes of interest to compare forecast performance to that of an allegedly-naive benchmark.

- Theil's U-Statistic

$$U = 1 - \frac{\sum_{t=1}^{T} \hat{e}_{t-1,t}^2}{\sum_{t=1}^{T} (Y_t - Y_{t-1})^2}$$

where the benchmark is a no change forecast.

Statistical Assessment of Accuracy Rankings

- Once we've decided on a loss function, we would like to know further whether one forecast is more accurate than another.
- In hypothesis testing terms, we might want to test the equal accuracy hypothesis,

$$E[L(\hat{e}^a_{t,t+h})] = E[L(\hat{e}^b_{t,t+h})]$$

  against the alternative hypothesis that one or the other is better.
- For example, consider inflation forecasting.
  - One might obtain survey-based forecasts from the Survey of Professional Forecasters (S), $\{\pi^S_t\}^T_{t=1}$.
  - One might also obtain market-based forecasts from inflation-indexed bonds (B), $\{\pi^B_t\}^T_{t=1}$.

Statistical Assessment of Accuracy Rankings

- Suppose that loss is quadratic and that during $t = 1, \ldots, T$, the sample mean-squared errors are

$$\widehat{\text{MSE}}(\pi_t^S) = 1.80, \quad \widehat{\text{MSE}}(\pi_t^B) = 1.92$$

  Evidently S wins, and one is tempted to conclude that S provides better inflation forecasts than does B.

- The forecasting literature is filled with such horse races, with associated declarations of superiority based on outcomes.

Statistical Assessment of Accuracy Rankings

- Obviously, however, the fact that $\widehat{\text{MSE}}(\pi_t^S) < \widehat{\text{MSE}}(\pi_t^B)$ in a particular sample realization does not mean that S is necessarily truly better than B in population.
- That is, even if in population $\text{MSE}(\pi_t^S) = \text{MSE}(\pi_t^B)$, in any particular sample realization $t = 1, \ldots, T$, one or the other of S and B must "win" so the question arises in any particular sample as to whether S is truly superior or merely lucky.
- The Diebold-Mariano test answers that question, allowing one to assess the significance of apparent predictive superiority.

The Diebold-Mariano Perspective

- Denote the loss associated with forecast error $\hat{e}_t$ by $L(\hat{e}_t)$.
  - For example, quadratic loss would be $L(\hat{e}_t) = \hat{e}_t^2$.
- The loss differential between forecasts 1 and 2 is then

$$d_t = L(\hat{e}_{1t}) - L(\hat{e}_{2t})$$

- The key hypothesis of equal predictive accuracy (i.e., equal expected loss) is

$$E[L(\hat{e}_{1t})] = E[L(\hat{e}_{2t})]$$

  which implies

$$E(d_t) = 0$$

The Diebold-Mariano Perspective

- To test the null hypothesis, Diebold and Mariano (1995) propose the Diebold-Mariano (DM) test statistic

$$DM = \frac{\bar{d}}{\hat{\sigma}_{\bar{d}}}$$

where $\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$, and $\hat{\sigma}_{\bar{d}}$ is a consistent estimate of the standard deviation of $\bar{d}$.

- It is shown that

$$DM \xrightarrow{d} N(0,1)$$

- In practice, simply regress the loss differential series d_t on an intercept using Newey-West HAC standard errors (DM HAC regression).

The Diebold-Mariano Perspective

In some instances that standard normal distribution fail to provide good approximation for DM statistic.

- Parameter estimation uncertainty. See West (1996, Econometrica).
    - Valid for non-nested models.
    - Not valid for nested models.
- Tests of comparing a pair of nested models.
    (1) Clark-West Statistic (See Clark and West, 2007, J. Econometrics)
    (2) Bootstrap your own critical values.
    (3) Better solution: (1) + (2)

The Clark-Weat Statistic

- Consider the following nested models:
  - Model 1: $Y_{t+k} = \alpha + e_{1t+k}$
  - Model 2: $Y_{t+k} = \alpha + \beta x_t + e_{2t+k}$
- The forecasts are
  - Model 1: $\hat{Y}_{1t,t+k} = \hat{\alpha}$
  - Model 2: $\hat{Y}_{2t,t+k} = \hat{\alpha} + \hat{\beta} x_t$
- Define forecasting errors

$$\hat{e}_{1t+k} = Y_{t+k} - \hat{Y}_{1t,t+k}$$
$$\hat{e}_{2t+k} = Y_{t+k} - \hat{Y}_{2t,t+k}$$

The Clark-Weat Statistic

- Clark and West (2007) MSPE-adj statistic is computed as:

$$\text{CW} = \frac{\sqrt{P}\bar{f}}{\sqrt{\hat{V}}},$$

where

$$\bar{f} = P^{-1}\sum_{t}\hat{f}_{t+k},$$

$$\hat{f}_{t+k} = (\hat{e}_{1t+k})^2 - \left[(\hat{e}_{2t+k})^2 - (\hat{Y}_{1t,t+k} - \hat{Y}_{2t,t+k})^2\right],$$

and $\hat{V}$ is the sample variance of $(\hat{f}_{t+k} - \bar{f})$.

- The Clark–West test is an approximately normal test for equal predictive accuracy in nested models.
- Test for equal MSPE by regressing $\hat{f}_{t+k}$ on a constant and using the resulting t-statistic with HAC standard error.