FORECAST (ACCURACY) EVALUATION

1. MINIMUM MSPE (MEAN SQUARE PREDICTION ERROR)

$$\text{MSPE} = \frac{1}{H}\sum_{i=1}^{H} \hat{e}_{t+i}^2(1),$$

where $\hat{e}_t(1) = X_{t+1} - \hat{X}_t(1)$ is the one-step ahead forecast error at time t.

2. USUAL *F* STATISTIC

Suppose that the following three assumptions exist:

1) The forecast errors have zero mean and normally distributed ;
2) The forecast errors are serially uncorrelated;
3) The forecast errors are contemporaneously uncorrelated with each other.

Under the above assumptions and additional assumption that model 1 and 2 have same forecast power (H_0), we have

$$F = \frac{\sum_{i=1}^{H} e_{1i}^2}{\sum_{j=1}^{H} e_{2i}^2} \sim F(H, H),$$

where e_{ki} stands for the one-step ahead forecast error of model $k, k = 1,2$ at time $t + i$.

Remark: The value of F will be unity if forecast errors from two models are identical. Thus, a very large value of F implies that the forecast errors from the first model will substantially larger than those from the second.

3. GRANGER-NEWBOLD TEST

Granger and Newbold (1976) showed how to overcome the problem of contemporaneously correlated forecast errors. Specifically, consider the following transformation

$$x_i = e_{1i} + e_{2i} \text{ and } z_i = e_{1i} - e_{2i}, i = 1, \dots, H$$

Given the first two assumptions are valid and under the null hypothesis that $H_0: var(e_{1i}) = var(e_{2i})$ (i.e., equal forecast accuracy), x_i and z_i should be uncorrelated

$$cov(x, z) = E(xz) = E(e_1^2 - e_2^2) = 0.$$

Let r_{xz} denote the sample correlation coefficient between $\{x_i\}$ and $\{z_i\}$, where

$$r_{xz} = \frac{\sum(x_i - \bar{x})(z_i - \bar{z})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(z_i - \bar{z})^2}}.$$

Granger and Newbold (1976) showed that if assumption 1 and 2 hold,

$$\frac{r_{xz}}{\sqrt{\frac{(1 - r_{xz}^2)}{H-1}}} \sim t_{H-1}.$$

Thus, if r_{xz} is statistically different from zero, we reject the null hypothesis. Specifically, if $r_{xz} > 0$, model 1 has a larger MPSE (so less accuracy); and if $r_{xz} < 0$, model 2 has a larger MPSE.

~~4.~~ THE DIEBOLD-MARIANO TEST

Consider the following transformation between two forecast errors

$$d_i = g(e_{1i}) - g(e_{2i}),$$

where $g(\cdot)$ is a loss function. For example, $g(y) = y^2$ form MPSE. The mean loss may be defined as

$$\bar{d} = \frac{1}{H}\sum_{i=1}^{H}\{g(e_{1i}) - g(e_{2i})\}.$$

Under the null hypothesis of equal forecast accuracy, the value of $\bar{d}$ should be zero. By CLT and some technical conditions, we have

$$\frac{\bar{d}}{\sqrt{var(\bar{d})}} \sim t_{H-1}.$$

where $\text{var}(\bar{d}) = (H-1)^{-1} \cdot \gamma(0)$ if $\{d_i\}$ are serially uncorrelated and $\text{var}(\bar{d}) = (H-1)^{-1} \cdot \gamma(0) \cdot \left(1 + 2\sum_{j=1}^{q}\rho(j)\right)$ if $\{d_i\}$ are serially correlated. Note that both Granger-Newbold and Diebold-Mariano tests do not work for nested models.

~~5.~~ CLARK AND WEST TEST

Clark and West (2007) propose a forecast accuracy test for nested models. Let $f_{ki} = X_{t+i}^{k}(1)$ denote the one-step ahead forecast at time $t + i$ from model $k, k = 1{,}2$. Assume that model 1 is nested within model 2 and consider

$$H_0\colon \sigma_{e1}^2 = \sigma_{e2}^2 \text{ vs. } H_a\colon \sigma_{e1}^2 < \sigma_{e2}^2$$

$$Z_i = e_{1i}^2 - \{e_{2i}^2 - (f_{1i} - f_{2i})^2\}$$

We can regress $\{Z_i\}$ on a constant and reject the null hypothesis if the value of t statistic is greater than 1.645 at 5% significant level.

6. <u>PRACTICE QUESTIONS</u>

Consider and ARMA(1,1) model

$$(1 - 0.5B)(X_t - 4) = (1 + 0.5B)a_t, \quad a_t \sim NID(0,1).$$

Its one-step forecast at origin $t = 99$ is $\hat{X}_{99}(1) = 2.09$, and

$\{X_{99}, X_{100}, X_{101}, X_{102}, X_{103}, X_{104}, X_{105}\} = \{2.11, 1.39, 2.57, 4.11, 6.28, 4.89, 5.94\}$.

We shall refer the above $ARMA(1,1)$ model as Model A. Use the above information to answer the following question.

a) [6%] Calculate the l step ahead forecast $\hat{X}_{100}(l)$ for $l = 1,2,3$.

Answer: [Marking scheme: 3 points for the general forecast formula, eqn. (1) and (2); and one point each for $\hat{X}_{100}(l), l = 1, 2, 3$]

The difference equation form the above $ARMA(1,1)$ process at time $n + l$is given by

$$X_{n+l} = 4 + 0.5(X_{n+l-1} - \mu) + a_{n+l} + 0.5a_{n+l-1}.$$

Using the conditional expectation given filtration at time n, we have

$$\hat{X}_n(1) = 4 + 0.5(X_n - 4) + 0.5\hat{a}_n, \quad \hat{a}_n = X_n - \hat{X}_{n-1}(1), \quad (1)$$

and

$$\hat{X}_n(l) = 4 + 0.5\left(\hat{X}_n(l-1) - 4\right), l \geq 2, \qquad (2)$$

Use the above results and let $n = 100$. We have

$$\hat{a}_{100} = X_{100} - \hat{X}_{99}(1) = 1.39 - 2.09 = -0.7,$$

i. $\hat{X}_{100}(1) = 4 + 0.5(1.39 - 4) - 0.5 \cdot 0.7 = 2.345,$
ii. $\hat{X}_{100}(2) = 4 + 0.5(2.345 - 4) = 3.1725,$
iii. $\hat{X}_{100}(3) = 4 + 0.5(3.175 - 4) = 3.58625.$

b) [4%] Calculate the forecast error variance for $l = 1,2,3$.

Answer: [Marking shceme: one point for eqn. (3) and one point each for $r(e_{100}(l)), l = 1, 2, 3$]

Since $\phi = 0.5 < 1$ we can calculate the ψ weights as follows:

$$(1 - \phi B)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1 - \theta B.$$

Equating the coefficients of B^j on both sides give

$$\psi_j = 0.5^{j-1}(0.5 + 0.5) = 0.5^{j-1}, \;\; j \geq 1. \qquad (3)$$

Since the forecast error variance is given by $var(e_n(l)) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2$, we have

i. $var(e_{100}(1)) = 1$
ii. $var(e_{100}(2)) = 1 + \sum_{j=1}^{2-1} [0.5^{j-1}]^2 = 1 + 0.5^0 = 2$
iii. $var(e_{100}(3)) = 1 + \sum_{j=1}^{3-1} [0.5^{j-1}]^2 = 1 + 0.5^0 + 0.5^2 = 2.25$

c) [5%] Describe the Granger-Newbold test for compare forecast accuracy and its assumptions.

Answer: [Marking scheme: 2 points for assumption and 3 points for describe the test]

Granger and Newbold (1976) considers the following transformation

$$x_i = e_{1i} + e_{2i} \text{ and } z_i = e_{1i} + e_{2i}, i = 1, \ldots, H,$$

where e_{ki} stands for the one-step ahead forecast error of model k, $k = 1,2$ at time $t + i$.

Granger and Newbold (1976) assumes that

i. The forecast errors have zero mean and normally distributed ;
ii. The forecast errors are serially uncorrelated;

Under the above two assumptions and under the assumption of equal forecast accuracy (H_0), x_i and z_i should be uncorrelated ($\because \rho_{xz} = E(xz) = E(e_1^2 - e_2^2) = 0$).

We can therefore use the sample correlation coefficient between $\{x_i\}$ and $\{z_i\}$, denoted as r_{xz}, to evaluate the accuracy between model 1 and 2. In particular, Granger and Newbold (1976) showed that

$$\frac{r_{xz}}{\sqrt{\frac{(1 - r_{xz}^2)}{H - 1}}} \sim t_{H-1}$$

if assumption 1 and 2 hold.

Thus, if r_{xz} is statistically different from zero, we reject the null hypothesis. Specifically, if $r_{xz} > 0$, model 1 has a larger MPSE (so less accuracy); and if $r_{xz} < 0$, model 2 has a larger MPSE.

d) [15%] Consider a non-nested competitive model B with the following one step ahead forecast errors $\{e_{100}(1), e_{101}(1), e_{102}(1), e_{103}(1)\ , e_{104}(1)\} = \{0.3, 0.9, 2, -1.5, 1.8\}$. Use the Granger-Newbold test to evaluate the forecast accuracy between model A and B. (Hint: Calculate for $e_{100}(1), e_{101}(1), e_{102}(1), e_{103}(1)$ and $e_{104}(1)$ for Model A and use 2.13 as the 95% quantile of a Student t distribution with 4 degrees of freedom)

Answer: [Marking scheme: see square-brackets below]

Recall that the general one step ahead forecast formula is given by

$$\hat{X}_n(1) = 4 + 0.5(X_n - 4) + 0.5\hat{a}_n, \quad \hat{a}_n = X_n - \hat{X}_{n-1}(1)$$

and in question a), we have $\hat{X}_{100}(1) = 2.345$ and $\hat{a}_{100} = X_{100} - \hat{X}_{99}(1) = -0.7$.

1) Use all these information, we can calculate the one-step ahead forecast for different origins $t = 101, 102, 103$, and 104. Specifically, we have [2 point each for $\hat{a}_i(1),\ i = 101,102,103,104$]
   i. $\hat{X}_{101}(1) = 4 + 0.5(X_{101} - 4) + 0.5\hat{a}_{101}, \quad \hat{a}_{101} = X_{101} - \hat{X}_{100}(1) = 0.7125$
   ii. $\hat{X}_{102}(1) = 4 + 0.5(X_{102} - 4) + 0.5\hat{a}_{102}, \quad \hat{a}_{102} = X_{102} - \hat{X}_{101}(1) = 1.86875$
   iii. $\hat{X}_{103}(1) = 4 + 0.5(X_{103} - 4) + 0.5\hat{a}_{103}, \quad \hat{a}_{103} = X_{103} - \hat{X}_{102}(1) = -1.184375$
   iv. $\hat{X}_{104}(1) = 4 + 0.5(X_{104} - 4) + 0.5\hat{a}_{101}, \quad \hat{a}_{105} = X_{101} - \hat{X}_{100}(1) = 2.0871875$
2) Let $x_i = \hat{a}_i + e_i(1)$ and $z_i = \hat{a}_i - e_i(1)$, $i = 100,101,102,103,104$. We can then apply the Granger-Newbold test defined in question c).
   i. The sample correlation $\hat{r}_{xz} = -0.336034$ [3 points] , and
   ii. The Granger-Newbold test is -1.427125. [3 points].
3) Since $-1.427125 > -2.13$ so we are not able to reject the null hypothesis (equal forecast error) at 10% confidence level. [1 point]

ARCH/GARCH MODEL

1. NOTATIONS OF ARCH PROCESSES

An autoregressive conditional heteroscedastic process of order p (denoted as $ARCH(p)$) is may be given by

$$X_t = \mu_t + \sigma_t Z_t,$$

$$\sigma_t^2 = w_0 + \sum_{i=1}^{p} w_i X_{t-i}^2,$$

where $Z_t \sim NID(0,1)$ and μ_t is sometimes referred to as the mean equation. Note that we require that $w_0 > 0, w_i \geq 0, i = 1, \ldots, p$ to ensure that $\sigma_t^2 > 0$. For simplicity and without loss of generality, we may assume $\mu_t = 0$ in some of subsequent analysis. To facilitate our subsequent discussion, we introduce the idea of filtration below. Assume that Z_t is a random variable at time t on a probability space (Ω, F, P) and assumed to be measurable with respect to a σ-algebra $F_t \subset F$. One may think F_t as the class of all events which are observable up to time t. Thus, it is natural to assume that

$$F_0 \subset F_1 \subset \cdots \subset F_T. \quad (1)$$

Having eqn. (1), we can now define *filtration*.

Definition: A family $(F_t)_{t=0,\ldots,T}$ of σ-algebras satisfying eqn. (1) is called *filtration*. In this case, $\left(\Omega, F, (F_t)_{t=0,\ldots,T}, P\right)$ is also called a filtered probability space.

1) ARCH(1) process is white noise:

Consider the following ARCH(1) process

$$X_t = \sigma_t \cdot Z_t,$$
$$\sigma_t^2 = w_0 + w_1 X_{t-1}^2.$$

We showed in class that that $E(X_t) = 0$, $\text{var}(X_t)$ is a constant, and $\gamma(l) = 0, \forall\, l \neq 0$. The sketchy proof is that

i. $E(X_t) = E[E(X_t|F_{t-1})] = E[E(\sigma_t Z_t|F_{t-1})] = E[\sigma_t E(Z_t|F_{t-1})] = 0$

ii. $\gamma_X(0) = E(X_t^2) = E(\sigma_t^2 Z_t^2) = E(\sigma_t^2) \cdot E(Z_t^2) = E(w_0 + w_1 X_{t-1}^2) = w_0 + w_1 E(X_{t-1}^2)$. Using the stationarity of X_t, we can then solve for $\gamma_X(0) = \frac{w_0}{1-w_1}$ provided that $w_1 \neq 1$.

iii. $\gamma(h) = E(X_t \cdot X_{t+h}) = E[E(X_t \cdot X_{t+h}|F_{t+h-1})] = E[X_t \cdot E(X_{t+h}|F_{t+h-1})]$
$= E[X_t \cdot E(\sigma_{t+h} Z_{t+h}|F_{t+h-1})] = E[X_t \sigma_{t+h} \cdot E(Z_{t+h}|F_{t+h-1})] = 0$, for $h > 0$ and $h \in$ Integers.

2) ARCH(1) process is fat-tailed:

We showed in class that $\mathrm{E}\{(\mathrm{X_t} - \mu_\mathrm{X})^4/\sigma_\mathrm{X}^4\} > 3$ so that an ARCH(1) process is fat-tailed. The sketchy proof is given as follows:

$E(X_t^4) = E[E(X_t^4|F_{t-1})] = E[E(Z_t^4|F_{t-1}) \cdot E(\sigma_t^4|F_{t-1})] = E[3 \cdot E(\sigma_t^4|F_{t-1})]$
$= 3E[(w_0 + w_1 X_{t-1}^2)^2] = 3E(w_0 + 2w_1 w_0 X_{t-1}^2 + w_1^2 X_{t-1}^4)$

Assuming that $E(X_{t+h}^4) = E(X_t^4) = m_4, \forall\, t, h \in$ Integers, we have

$$m_4 = \frac{3w_0^2(1 + w_1)}{[(1 - 3w_1^2)(1 - w_1)]}.$$

Under the condition that $0 \le w_1^2 \le 1/3$ for $m_4 > 0$, the kurtosis of X_t is given by

$$\frac{E(X_t^4)}{\left(E(X_t^2)\right)^2} = \frac{3(1 - w_1^2)}{1 - 3w_1^2} > 3.$$

2. ARCH MODEL BUILDING PROCEDURE

1) Specify a mean equation (μ_t) by testing for serial dependence and if necessary, building an econometric/statistical model (eg. ARMA) for time series data to remove any linear dependence;
2) Use the residuals of the mean equation to test ARCH effect;
3) Specify a volatility model if ARCH effects are statistically significant;
4) Perform a joint estimation of the mean equation and the volatility equation;
5) Check the fitted model carefully and redefine it if necessary.

3. TESTS FOR ARCH EFFECTS

1) Ljung-Box statistic on $\{a_t^2\}$ series of McLeod & Li (1983), where $a_t = X_t - \mu_t$ and see http://www.stats.uwo.ca/faculty/aim/vita/pdf/SquaredRACF.pdf for details

2) Lagrange multiplier test of Engle (1982)

- This test is equivalent to use the usual F statistic for testing $w_i = 0$, $i = 1, \dots, p$ in the linear regression
$$a_t^2 = w_0 + w_1 a_{t-1}^2 + \cdots + w_p a_{t-1}^2 + e_t, \quad t = p+1, \cdots, T$$
- $H_0: w_1 = w_2 = \cdots = w_p = 0$
- $SSR_0 = \sum_{p+1}^{T} (a_t^2 - \overline{w})^2$, where $\overline{w} = T^{-1}(\sum_1^T a_t^2)$
- $SSR_1 = \sum_{p+1}^{T} \hat{e}_t^2$, where $\hat{e}_t$ is the fitted residuals
- $F^0 = (SSR_0 - SSR_1) \cdot p^{-1} \,/\, SSR_1 \cdot (T - 2p - 1)^{-1} \sim \chi_p^2$

4. MODEL DIAGNOSTIC CHECKING

1) For a proper specified ARCH model, the standardized residuals $\tilde{a}_t = a_t/\sigma_t$ are a sequence of IID random variables;
2) Therefore one can check the adequacy of a fitted ARCH model by examining the series $\{\tilde{a}_t\}$. In particular, the Ljung-Box statistic of $\tilde{a}_t$ can be used to check the adequacy of the fitted mean equation and $\tilde{a}_t$;
3) Skewness, kurtosis and QQ-plot of standardized residuals $\{\tilde{a}_t\}$ can be used to check distributional assumption.

5. GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTIC (GARCH) MODEL

A generalized autoregressive conditional heteroscedastic model of order p and q (denoted as GARCH(p, q)) may be defined as follows:

$$X_t = \mu_t + \sigma_t Z_t,$$

$$\sigma_t^2 = w_0 + \sum_{i=1}^{p} w_i X_{t-i}^2 + \sum_{j=1}^{q} \eta_j \sigma_{t-j}^2,$$

where $Z_t \sim NID(0,1)$ and the conditions for $\sigma_t^2 > 0$ are $w_0 > 0, w_i \geq 0, \eta_j \geq 0, i = 1, \dots, p, j = 1, \dots, q$.