Chow forecast test:

Instead of using all the sample observations for estimation, the suggested procedure is to divide the data set

of N sample observations into N_1 observations to be used for estimation

and $N_2 = N - N_1$ for testing if the parameter vector is constant.

There are no hard and fast rules for determining the relative size of N_1 and N_2.

It is not uncommon to reserve $5, 10$, or 15 percent of the observations for testing.

Predictive accuracy test:

The test of predictive accuracy, widely referred to as Chow test, is as follows:

1. Estimate the OLS vector from the first N_1 observations, obtaining

$$\beta_1 = (X_1' X_1)^{-1} X_1' y_1$$

and the vector of the residuals $\widehat{e}_1 = y_1 - X_1\beta_1$ and the $RSS_1 = \widehat{e}_1'\widehat{e}_1$.

2. Fit the same regression to all $N = N_1 + N_2$ observations and obtain the (restricted) RSS$= \widehat{e}'\widehat{e}$.

3. Employ the F test statistic

$$F = \frac{\widehat{e}'\widehat{e} - \widehat{e}_1'\widehat{e}_1/N_2}{\widehat{e}_1'\widehat{e}_1/N_1 - k} \sim F(N_2, N_1 - k) \qquad (1)$$

4. Reject the null hypothesis of parameter constancy if F exceeds a preselected critical value.

THE CUSUM TEST

Recall that our OLS regression is

$$y_t = x_t' b + e_t,\ t = 1, \ldots, N,$$

where $x_t' = [1\ x_{2t} \cdots x_{kt}]$ is the row vector of the regressors at time t.

Alternatively, we have

$$y = Xb + e,$$

where

$$X = \begin{bmatrix} \cdots x_1' \cdots \\ \cdots x_2' \cdots \\ \vdots \\ \cdots x_N' \cdots \end{bmatrix}$$

The idea behind recursive estimation is very simple.

Fit the model to the first k observations

Next use the first $k + 1$ data points and compute the coefficient vector again.

Proceed in this way, adding one sample data point at a time, until the final coefficient vector is obtained, based on all N sample data points.

This process generates a sequence of vectors,

β_k, β_{k+1},. . .,β_N

where the subscript indicates the number of sample points used in the estimation.

In general

$$\beta_t = (X_t' X_t)^{-1} X_t' y_t$$

where $t = k, k+1, \ldots, N$.

X_t is the $t \times k$ matrix of the regressors for the first t sample points

and y_t is the $t \times 1$ vector of the first t observations on the dependent variable.

The standard errors of the various coefficients may be computed at each stage of the recursion,

except at the first step, since RSS is zero when $t = k$.

Graphs may be prepared showing the evolution of each coefficient, plus and minus two standard errors.

Visual inspection of the graphs may suggest parameter constancy, or its reverse.

ONE-STEP AHEAD PREDICTION ERRORS

By using all data up to and including period $t-1$, the one step ahead prediction of y_t is

$$x_t'\beta_{t-1} = \beta_{1,t-1} + \beta_{2,t-1}x_{2t} + \cdots + \beta_{k,t-1}x_{kt}$$

where $\beta_{t-1}' = [\beta_{1,t-1}\ \beta_{2,t-1}\cdots\beta_{k,t-1}]$.

The one-step ahead prediction error is thus

$$v_t = y_t - x_t'\beta_{t-1} \tag{2}$$

where $t = k+1, k+2, \ldots, N$.

The variance of v_t can be seen to be

$$\sigma^2[1 + \underbrace{x_t'}_{1\times k}\underbrace{(X_{t-1}'X_{t-1})^{-1}}_{k\times k}\underbrace{x_t}_{k\times 1}]. \tag{3}$$

The scaled recursive residuals are defined as

$$w_t = \frac{v_t}{\sqrt{[1 + \underbrace{x_t'}_{1\times k}\underbrace{(X_{t-1}'X_{t-1})^{-1}}_{k\times k}\underbrace{x_t}_{k\times 1}]}},$$

$t = k+1, k+2, \ldots, N.$

Under our assumptions $w_t \sim N(0, \sigma^2)$.

CUSUM test

The first test statistic, based on the scaled recursive residuals, is the CUSUM quantity

$$W_t = \sum_{j=k+1}^{t} w_j/s$$

where $s^2 =$RSS$_N/N-k$, with the RSS$_N$ being the residual sum of squares calculated from the full sample regression.

W_t is a cumulative sum, and it is plotted against t.

With constant parameters, $E(W_t) = 0$, but with non-constant parameters W_t will tend to diverge from the zero mean value line.

The significance of the departure from the zero line may be assessed by reference to a pair of straight lines that pass through the points

$$k \pm a\sqrt{N-k} \text{ and}$$
$$N \pm 3a\sqrt{N-k},$$

where α is a parameter depending on the significance level of a chosen for the test.

The correspondence for some conventional significance level is

$$\begin{aligned} 0.01;\, a &= 1.143 \\ 0.05;\, a &= 0.948 \\ 0.10;\, a &= 0.850. \end{aligned}$$

CUSUMQ test

The second test statistic is based on the cumulative sums of squared residuals, namely,

$$S_t = \frac{\sum_{j=k+1}^{t} w_j^2}{\sum_{j=k+1}^{N} w_j^2},\ t = k+1, \ldots, N$$

Under the null hypothesis the squared w's are independent $\chi^2(1)$ variables.

The numerator thus has an expected value of $t - k$,

and the denominator an expected value of $N - k$.

The mean value line, giving an approximate expected value of the test statistic under the null hypothesis, is

$$E(S_t) = \frac{t-k}{N-k},$$

which goes from 0 at $t = k$ to unity at $t = N$.

The significance of departures from the expected value line is assessed

by reference to a pair of lines drawn parallel to the $E(S_t)$ line

at a distance c_0 above and below.

Values of c_0 for various sample sizes and significance levels are tabulated in Appendix D8 in JD book.

Ramsey Reset Test

Ramsey has argued that various specification errors (omitted variables, incorrect functional form, correlation between X and e) give rise to a nonzero e vector.

Thus the null and alternative hypotheses are

$$\begin{aligned} H_0 &: e \sim N(0, \sigma^2 I), \\ H_1 &: e \sim N(\mu, \sigma^2 I). \end{aligned}$$

The test of H_0 is based on an augmented regression:

$$y = Xb + Za + e.$$

The test for the specification error is then $a = 0$.

Ramsey's suggestion is that Z should contain powers of the predicted values of the dependent variable:

$$Z = [\widehat{y}^{\wedge 2} \ \widehat{y}^{\wedge 3} \ \widehat{y}^{\wedge 4}],$$

where $\widehat{y} = X\beta$ and $^{\wedge}$ denotes element by element exponentiation:

$\widehat{y}^{\wedge i} = [\widehat{y}_1^i \ \widehat{y}_2^i \dots \widehat{y}_N^i]'$.

Chow test:

Split the data into two subperiods. Estimate the regression over the whole period

and then for the two sub-periods separately (3 regressions).

Obtain the RSS for each regression.

The restricted regression is now the regression for the whole period

while the unrestricted regression comes into two parts: one for each of the subsamples

The test statistic

$$\frac{\overbrace{RSS}^{\text{whole period}} - (\overbrace{RSS_1}^{\text{period 1}} + \overbrace{RSS_2}^{\text{period 2}})}{(RSS_1 + RSS_2)} \times \frac{N-2k}{k} \sim F(k, N-2k),$$

since the number of restrictions is equal to the number of coefficients that are estimated for each of the regression, i.e.k

The number of regressors in the unrestricted regression is $2k$

The test is one of how much the residual sum of squares for the whole sample, RSS (the restricted regression)

is bigger than the sum of the residual sum of squares for the two sub-samples, $RSS_1 + RSS_2$, unrestricted regression.

Dummy variables:

Note that it is also possible to use a dummy variables approach to calculate the Chow test.

The unrestricted regression will include dummy variables for the intercept and for all the slope coefficients.

Suppose that we have two regressors x_{2t} and x_{3t}. The unrestricted regression would be given by

$$y_t = b_1 + b_2 x_{2t} + b_3 x_{3t} + b_4 D_t + b_5 D_t x_{2t} + b_6 D_t x_{3t} + e_t,$$

where $D_t = 1$ for $t \in N_1$ $(t = 1, 2, \ldots, N_1)$ and zero otherwise.

In other words, D_t takes the value one for observations in the first sub-sample and zero for observations in the second sub-sample.

In particular, the estimated slope coefficient of x_{3t} for the first sub-period is $\beta_3 + \beta_6$

whereas for the second sub-sample is β_3.

The Chow test viewed in this way would then be the standard F-test of the joint restriction:

$H_0 : b_4 = b_5 = b_6$.

Example:

Consider the following regression for the standard CAPM b for the returns on a stock

$$r_{gt} = a + br_{Mt} + e_t,$$

where r_{gt} and r_{Mt} are excess returns on Glaxo shares and on a market portfolio.

Suppose that you are interested in estimating beta using monthly data from 1980 to 1991, to aid a stock selection decision.

Another researcher expresses concern that the October 1987 stock market crash fundamentally altered the risk-return relationship.

Test this conjecture using a Chow test.

The model for each sub-period is

$1981M1 - 1987M10$

$$\widehat{r}_{gt} = 0.24 + 1.2 r_{Mt},\ N = 82,\ RSS_1 = 0.03555$$

$1987M11 - 1992M12$

$$\widehat{r}_{gt} = 0.68 + 1.53 r_{Mt},\ N = 62,\ RSS_2 = 0.00336$$

$1981M1 - 1992M12$

$$\widehat{r}_{gt} = 0.39 + 1.37 r_{Mt},\ N = 144,\ RSS = 0.0434$$

The null hypothesis is $H_0 : a_1 = a_2$ and $b_1 = b_2$

where subscripts 1 and 2 denote the parameters for the first and second sub-samples, respectively.

The test statistic will be given by

$$\frac{\overbrace{0.0434}^{RSS} - (\overbrace{0.03555 + 0.00336}^{RSS_1+RSS_2})}{0.03555 + 0.00336} \times \frac{\overbrace{144 - 4}^{N-2k}}{2} = 7.698.$$

The test statistic should be compared with the 5%,

$F(2, 140) = 3.06.$

H_0 is rejected at the 5% level and hence it is concluded that the restriction that the coefficients are the same in the two periods cannot be employed.

(See Chris Brooks: "Introductory Econometrics for Finance", Example 4.4)

Predictive failure test:

A problem with Chow test is that it is necessary to have enough data to do the regression on both sub-samples,

i.e. $N_1 > k$ and $N_2 > k$.

This may not hold in the situation where the total number of observations available is small.

Even more likely is the situation where the researcher would like to examine the effect of splitting the sample at some point very close to the start or very close to the end of the sample.

An alternative formulation of a test for the stability is the predictive failure test, which requires estimation for the full sample and one of the sub-samples only.

The predictive failure test works by estimating the regression over a 'long' sub-period (i.e. most of the data and then using those coefficients estimates for predicting values of y for the other period.

The null hypothesis for this test is that the prediction errors for all the forecasted observations are zero.

(See the test above)

Example:

Suppose that the researcher decided to determine the stability of the estimated model for stock returns over the whole sample in the previous example by using a predictive failure test of the last two years of observations.

The following model would be estimated:

$1981M1 - 1992M12$

$$\widehat{r}_{gt} = 0.39 + 1.37 r_{Mt},\ N = 144,\ RSS = 0.0434$$

$1981M1 - 1990M12$

$$\widehat{r}_{gt} = 0.32 + 1.31 r_{Mt},\ N = 120,\ RSS_1 = 0.0420$$

The test statistic would be given by

$$\frac{\overbrace{0.0434}^{RSS} - (\overbrace{0.0420}^{RSS_1})}{0.0420} \times \frac{\overbrace{120-2}^{N_1-2}}{N_2} = 0.164.$$

Since $F(24, 118) = 1.66$ at the 5% the null hypothesis would not be rejected.

(See Chris Brooks: "Introductory Econometrics for Finance", Example 4.5)

Dummy variables:

An alternative way to test for predictive failure is using a regression containing dummy variables.

A separate dummy variable would be used for each observation that was in the prediction sample.

The unrestricted regression will then be the one that includes the dummy variables, which will be estimated using all N observations,

and will have $k + N_2$ regressors.

The restricted regression would then be the original regression containing the explanatory variables but none of the dummy variables.

In the context of our example suppose that the last three observations in the sample are used for a predictive failure test.

The unrestricted regression will include three dummy variables, one for each of the observations in N_2:

$$r_{gt} = a + br_{Mt} + \gamma_1 D1_t + \gamma_2 D2_t + \gamma_3 D3_t + e_t,$$

where $D1_t = 1$ for observation $N - 2$, $D2_t = 1$ for observation $N - 1$ and $D3_t = 1$ for observation N.

In this case $k = 2$ and $N_2 = 3$.

The null hypothesis for the predictive failure test in this regression is that the coefficients on all of the dummy variables are zero:

$H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0.$

Both approaches to conducting the predictive failure test described above are equivalent.

A major advantage of the dummy variables approach is that one can examine the significances of the coefficients on the individual dummy variables to see which part of the joint hypothesis is causing a rejection.

Hansen test:

A difficulty with the Chow test is the arbitrary nature of the partitioning of the data set. One such partitioning might reject the null hypothesis and another fail to reject.

This difficulty does not apply to the Hansen test, which is fits the linear equation to all N observations.

Recall that the equation with the k regressors is

$$\begin{aligned} y_t &= b_1 + b_2 x_{2t} + \cdots + b_k x_{kt} + e_t \text{ or} \\ & \quad [1\ x_{2t} \cdots x_{kt}] \\ y_t &= \overbrace{x_t'} \qquad \underbrace{b} \qquad + e_t. \\ & \qquad\qquad [b_1\ b_2 \cdots b_k]' \end{aligned}$$

Recall that the OLS fit gives the conditions

$$\sum_{t=1}^{N} x_{it}\widehat{e}_t = 0,\ i = 1, \ldots, k,$$

$$\sum_{t=1}^{N}(\widehat{e}_t^2 - \widehat{\sigma}^2) = 0 \text{ or } \widehat{\sigma}^2 = \frac{\sum_{t=1}^{N} \widehat{e}_t^2}{N}.$$

Next define

$$f_{it} = \begin{cases} x_{it}e_t & i = 1, \ldots, k \\ \widehat{e}_t^2 - \widehat{\sigma}^2 & i = k+1 \end{cases}$$

where $\sum_{t=1}^{N} f_{it} = 0,\ i = 1, \ldots, k+1.$

The Hansen test statistics are based on cumulative sums of the f_{it}, namely,

$$S_{it} = \sum_{j=1}^{t} f_{ij}.$$

He develops tests for the stability of each parameter individually and for the joint stability of all parameters.

The individual test statistics are

$$L_i = \frac{1}{NV_i} \sum_{t=1}^{N} S_{it}^2,\ i = 1, \ldots, k+1,$$

where $V_i = \sum_{j=1}^{N} f_{ij}^2$.

For the joint stability test let

$$\begin{aligned} f_t &= [f_{1t} \cdots f_{k+1t}]', \\ s_t &= [S_{1t} \cdots S_{k+1t}]'. \end{aligned}$$

The joint stability test is then

$$L_c = \frac{1}{N} \sum_{t=1}^{N} \underbrace{s_t'}_{1\times(k+1)} \underbrace{V}_{(k+1)\times(k+1)}^{-1} \underbrace{s_t}_{(k+1)\times 1}$$

where

$$\underbrace{V}_{(k+1)\times(k+1)} = \sum_{t=1}^{N} f_t f_t'.$$

Under the null hypothesis the cumulative sums will tend to be distributed around zero.

Thus 'large' values of the test statistics suggest rejection of the null.

The distribution theory is nonstandard, and only asymptotic critical values are available.

These are tabulated in Appendix D7 in JD book.

The 5% critical value for the individual coefficient test is 0.470.

Numerical Example:

We have three variables:

y_t: log of per capita real expenditure in gasoline and oil

x_{2t}: log of the real price index for gasoline and oil

x_{3t}: log or per capita real disposable personal income

The first oil shock hit in 1973.4, so we chose a sample period from 1959.1 to 1973.3 a periods for which it might seem reasonable to postulate parameter constancy.

The first 51 observations were used for estimation and the remaining 8 reserved for the Chow forecast test.

The simple specification

$$y = b_1 + b_2 x_{2t} + b_3 x_{3t} + e_t$$

was employed.

1. The specification passes the Chow test (see equation 1): the F statistic is only 0.18.

2. The one-step forecasts in the column "forecast" take the form:

$\widehat{y}_t = \underbrace{x_t'}_{(1\times 3)} \underbrace{\beta}_{(3\times 1)}$ where β is the coefficient vector estimated from all 51 observations.

The forecast standard error SE is given by

$$s\sqrt{1+\underbrace{x_t'}_{(1\times 3)}\underbrace{\left(\underbrace{X'}_{(3\times 51)}\underbrace{X}_{(51\times 3)}\right)^{-1}}_{(3\times 3)}\underbrace{x_t}_{(3\times 1)}}$$

where s is the standard error of the regression

and X is the the matrix of regressors for the 51 sample points

3. The column headed instab contains the Hansen statistics for testing the stability of individual coefficients:

The hypothesis of stability is rejected for all three coefficients, and, not surprisingly, the joint stability test decisively rejects the null of parameter constancy.

4. We have a very low value of the Durbin-Watson statistic. This indicates substantial autocorrelation in the disturbance term.

Recursive Residuals:

5a. Figure 4.3 shows the recursive residuals (see equation 2) along with two standard error bands (see equation 3)

A point on the graph lying outside the standard error bands is equivalent to a t statistic $[v_t/s.e(v_t)]$ being numerically greater than two and thus suggestive of parameter inconstancy.

There is one such point in 1966, and a number of similar points from 1968 to 1970.

5b. Figure 4.4 is an alternative way of showing the same information as that given by Figure 4.3.

The one-step Chow test for parameter constancy through the first j observations is based on

$$\begin{aligned} F_m &= \frac{RSS_j - RSS_{j-1}}{RSS_{j-1}/(j-k-1)} \sim F(1, j-k-1), \\ j &= m+1, \ldots, N \end{aligned}$$

where m is the number of observations used in the initial recursion: $m = 1, \ldots, N-1$.

Dividing the F statistic in the above expression by the 5% critical value from $F(1, j-k-1)$ gives the series plotted in Figure 4.4.

Any point lying above the horizontal line at 1 implies rejection of parameter constancy, whereas points below do not lead to rejection.

As in figure 4.3 there is one rejection in 1966 and a group of rejections in 1968 through 1970.

CUSUM tests:

6. The three panels in figure 4.5 show the recursively estimated coefficients, with two standard error bands.

As might be anticipated there are dramatic changes in the late 1960s, especially in the constant ($C1$), and the price elasticity, ($C2$).

In the first half of the sample, the price elasticity is not significantly different from zero, and its point estimate is positive.

Only when data from the 1970s are included does the price elasticity turn negative, and significantly so.

The income elasticity, ($C3$) is positive and reasonably stable.

7. The CUSUM tests reported in Figure 4.6 confirm the message of the previous findings.

8. Finally, the Ramsey RESET test, using just $\widehat{y}^2$, gives $F = 47.2$, which is a very strong indicator of specification error.

Normality test:

(See Chris Brooks: "Introductory Econometrics for Finance", Section 4.9)

On of the most commonly applied tests for normality is the Bera-Jarque (BJ) BJ test.

In a normally distributed random variable the entire distribution is characterized by the first two moments-the mean and the variance.

The standardized third and fourth moments of a distribution are known as its skewness and kurtosis.

Skewness measures the extent to which a distribution is not symmetric about its mean value

and kurtosis measures how fat the tails of the distribution are.

A normal distribution is not skewed and is defined to have a coefficient of kurtosis of 3.

In other words, the coefficient of excess kurtosis, that is the coefficient of kurtosis minus 3, is zero for the normal distribution.

A normal distribution is symmetric and said to be mesokurtic.

A skewed distribution will have one tail longer than the other.

A leptokurtic distribution is one which has fatter tails and is more peaked at the mean than a normally distributed random variable with the same mean and variance.

A platykurtic distribution will be less peaked in the mean, will have thinner tails, and more of the distribution in the shoulders than a normal.

In practice, a leptokurtic distribution is far more likely to characterize financial (and economic) time series.

JB test statistic:

The coefficients of skewness and kurtosis are

$$sk = \frac{E(e^3)}{(\sigma^2)^{\frac{3}{2}}},\ k = \frac{E(e^4)}{(\sigma^2)^2}.$$

Bera and Jarque (1981) test whether the coefficient of skewness and the coefficient of excess kurtosis are jointly zero

(the normality assumption).

The BJ test statistic is given by

$$JB = N[\frac{\widehat{sk}^2}{6} + \frac{(\widehat{k} - 3)^2}{24}],$$

where N is the sample size and $\widehat{sk}$, $\widehat{k}$ are the sample skweness and kurtosis.

The test statistic asymptotically follows a $\chi^2(2)$ under the null hypothesis that the distribution of the series is symmetric and mesokurtic.

The null hypothesis of normality would be rejected if the residuals from the model were either significantly skewed or leptokurtic

Non-normality:

For sample sizes that are sufficiently large, violation of the normality assumption is virtually inconsequential.

Appealing to the central limit theorem, the test statistics will asymptotically follow the appropriate distributions even in the absence of normality.

It is, of course, possible to employ an estimation method that does not assume normality, but another distribution

i.e. the t-distribution.

In economic or financial modelling it is quite often the case that one or two very extreme residuals cause a rejection of the normality assumption.

Such observations would appear in the tails of the distribution, and would therefore lead e^4, which enters in the definition of the kurtosis, to be very large.

Such observations that do not fit in with the pattern of the remainder of the data are known as outliers.

If this is the case, one way to improve the chances of error normality is to use dummy variables or some other method to effectively remove these observations.

Outliers:

In the time series context, suppose that a monthly model of asset returns from 1980-1990 had been estimated, and the residuals plotted, and that a particularly large outlier has been observed for October 1987.

A new variable could be defined as

$D87M10_t = 1$ during October 1987 and zero otherwise.

The dummy variable would then be used just like any other variable in the regression model, e.g.

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + b_3 D87M10_t + e_t$$

This type of dummy variable that takes the value one for only a single observation has an effect exactly equivalent to knocking out that observation from the sample altogether, by forcing the residual for that observation to zero.

The estimated coefficient on the dummy variable will be equal to the residual that the dummied observation would have taken if the dummy variable had not been included.

However,many econometricians would argue that dummy variables to remove outlying residuals can be used to artificially improve the characteristics of the model-in essence fudging the results.

Removing outlying observations will reduce standard errors, reduce the RSS, and therefore increase r^2, thus improving the apparent fit of the model of the data.

The removal of observations is also hard to reconcile with the notion in statistics that each data point represents a useful piece of information.

The other side of this argument is that outliers can have a serious effect on coefficient estimates, since by definition, OLS will receive a big penalty, in the form of an increased RSS, for points that are a long way from the fitted line. Consequently, OLS will try extra hard to minimize the distances of points that would have otherwise been a long way away from the line.

In figure (see transparencies) one point is a long way from the rest. If this point is included in the estimation sample, the fitted line will have a positive slope. If this observation were removed, the fitted line will have a negative and larger slope. OLS could not select this line if the outlier is included since the observation is a long way from the others and hence when the residual is squared, it would lead to a big increase in the RSS.

Note that outliers could be detected by plotting y against x only in the context of a bivariate regression. In the case where there are more explanatory variables, outliers are easiest identified by plotting the residuals over time.

So it can been seen that a trade-off potentially exists between the need to remove outlying observations that could have an undue impact on the OLS estimates and cause residual non-normality on the one hand, and the notion that each data point represents a useful piece of information on the other. the latter is coupled with the fact that removing observations at will could artificially improve the fit of the model.

A sensible way to proceed is by introducing dummy variables to the model only if there is both a statistical need to do so and a theoretical justification for their inclusion. Dummy variables may be justifiably used to remove observations corresponding to 'one-off' or extreme events that are considered highly unlikely to be repeated, and the information content of which is deemed of no relevance for the data as a whole. Examples may include stock market crashes, financial panics, government crises, and so on.