Lecture 7

What we know now…

T tests $\hat{t} = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$

and confidence intervals

$$\Pr\left[\hat{\beta}_1 - t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1)\right] = 0.95$$

are useful ways to check whether the OLS estimates are compatible with our (economic) priors about how the world works

But they are not the only tests that we can use to judge the performance of the model and our economic priors

In what follows look at a set if different tests (F tests) that can be used to test different sorts of hypotheses and apply them to the issues of forecasting and parameter stability

F test of Goodness of Fit of the Model

We know that the R^2 gives an idea of the strength of association between dependent and explanatory variable

F test of Goodness of Fit of the Model

We know that the R^2 gives an idea of the strength of association between dependent and explanatory variable

Because of statistical variation unlikely that R^2 would ever fall to zero in absence of any (true) correlation between y and X variables

F test of Goodness of Fit of the Model

We know that the R^2 gives an idea of the strength of association between dependent and explanatory variable

Because of statistical variation unlikely that R^2 would ever fall to zero in absence of any (true) correlation between y and X variables

So how do we know that the value for the R^2 reflects a true relationship rather than the result of chance?

F test of Goodness of Fit of the Model

We know that the R^2 gives an idea of the strength of association between dependent and explanatory variable

Because of statistical variation unlikely that R^2 would ever fall to zero in absence of any (true) correlation between y and X variables

So how do we know that the value for the R^2 reflects a true relationship rather than the result of chance?

We know that

$$\mathrm{Var}(Y) = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(\hat{u})$$

F test of Goodness of Fit of the Model

We know that the R^2 gives an idea of the strength of association between dependent and explanatory variable

Because of statistical variation unlikely that R^2 would ever fall to zero in absence of any (true) correlation between y and X variables

So how do we know that the value for the R^2 reflects a true relationship rather than the result of chance?

We know that

$$\mathrm{Var}(Y) = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(\hat{u})$$

and that this can be also written as

$$\sum(Y - \bar{Y})^2 = \sum(\hat{Y} - \bar{Y})^2 + \sum \hat{u}^2$$

$$TSS = ESS + RSS$$

Can show (don't need to learn proof) that can use the value

$$F = \frac{ESS / k - 1}{RSS / N - k} \sim F[k-1, N-k]$$

to test whether the R^2 is statistically significantly different from zero – *allowing for the random variation that use of a sample entails*

Can show (don't need to learn proof) that can use the value

$$F = \frac{ESS / k - 1}{RSS / N - k} \qquad \sim F[k-1, N-k]$$

to test whether the R^2 is statistically significantly different from zero *allowing for the random variation that use of a sample entails*

Intuitively if the model is a good fit then the ESS will be high relative to the RSS.

Can show (don't need to learn proof) that can use the value

$$F = \frac{ESS / k - 1}{RSS / N - k} \quad \sim F[k-1, N-k]$$

to test whether the R^2 is statistically significantly different from zero *allowing for the random variation that use of a sample entails*

Intuitively if the model is a good fit then the ESS will be high relative to the RSS.

Just how high depends (like t values) on the sample size N and the number of right hand side coefficients k (hence the corrections above)

Can show (don't need to learn proof) that can use the value

$$F = \frac{ESS / k - 1}{RSS / N - k} \sim F[k-1, N-k]$$

to test whether the R^2 is statistically significantly different from zero *allowing for the random variation that use of a sample entails*

Intuitively if the model is a good fit then the ESS will be high relative to the RSS.

Just how high depends (like t values) on the sample size N and the number of right hand side coefficients k (hence the corrections above)

The F statistic

(named after Ronald Fisher 1890-1962 – the same person who developed the t test and the same person who wrote a book attributing the decline of the British empire to the failure of its upper classes to breed)

is the ratio of 2 chi-squared distributions divided by their respective degrees of freedom

The F statistic

has its own set of degrees of freedom

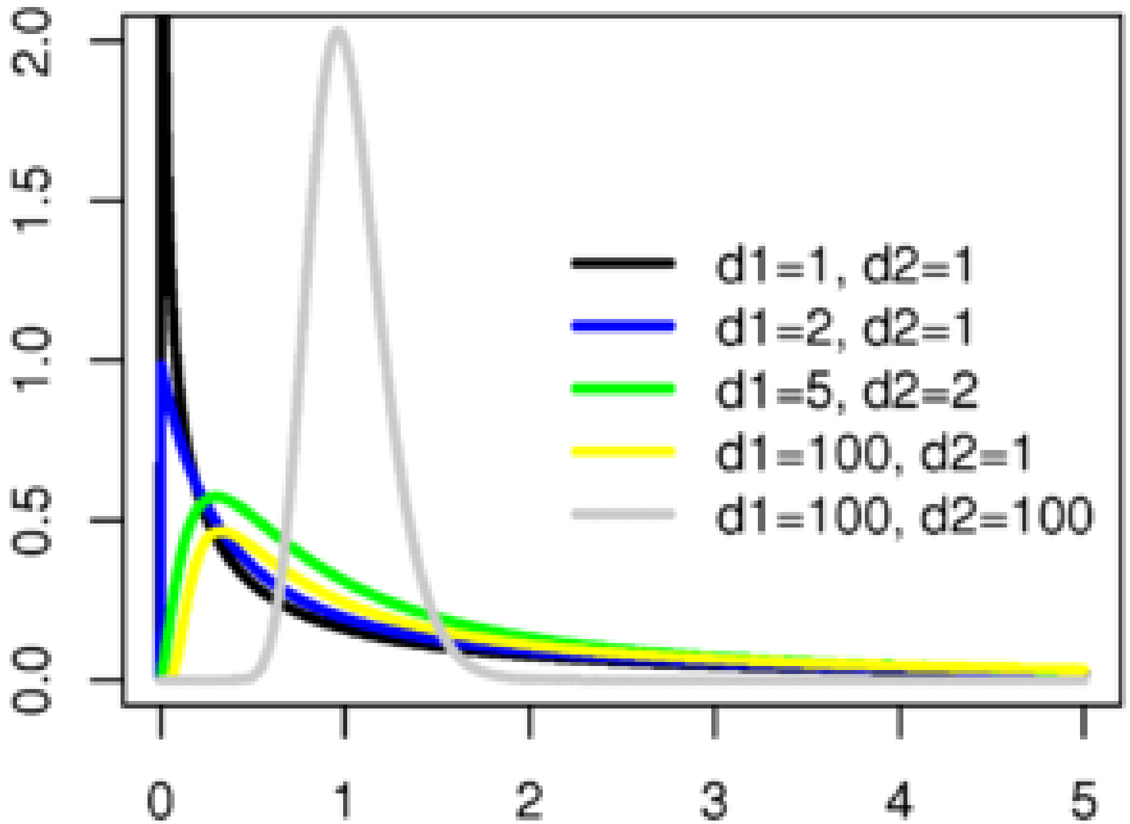
$F[k-1, N-k]$

The F statistic

has its own set of degrees of freedom

$F[k-1, N-k]$

k-1 said to be the numerator
N-k said to be the denominator (for obvious reasons)



(In stata
rndf 1000 5 55

Now, unlike the t distribution, the F distribution because it is effectively the t distribution “squared”) is bounded from below at zero.

Now, unlike the t distribution, the F distribution (because it is effectively the t distribution "squared") is bounded from below at zero.

This means that any hypothesis testing is based on a "1-tailed test" since the rejection region now only lies at the far right-end tail of the distribution

Now, unlike the t distribution, the F distribution because it is effectively the t distribution "squared") is bounded from below at zero.

This means that any hypothesis testing is based on a "1-tailed test" since the rejection region now only lies at the far right-end tail of the distribution

The basic rule is that

$$\text{if } \hat{F} > F^{(k-1,N-k)}_{\alpha\_critical}$$

Now, unlike the t distribution, the F because it is effectively the t distribution “squared”) is bounded from below at zero.

This means that any hypothesis testing is based on a “1-tailed test” since the rejection region now only lies at the far right-end tail of the distribution

The basic rule is that

$$\text{if } \hat{F} > F^{(k-1, N-k)}_{\alpha\_critical}$$

then can be 1-a% confident that result is unlikely to have arisen by chance (the ESS is high wrt the RSS)

Since the F-statistic can also be manipulated to give

$$F = \frac{ESS / k - 1}{RSS / N - k} \qquad = \frac{(ESS / TSS) / k - 1}{(RSS / TSS) / N - k}$$

Since the F-statistic can also be manipulated to give

$$F = \frac{ESS / k - 1}{RSS / N - k} \qquad = \frac{(ESS / TSS) / k - 1}{(RSS / TSS) / N - k}$$

then $F = \dfrac{R^2 / k - 1}{(1 - R^2) / N - k}$

which is an equivalent way of doing the test

Since the F-statistic can also be manipulated to give

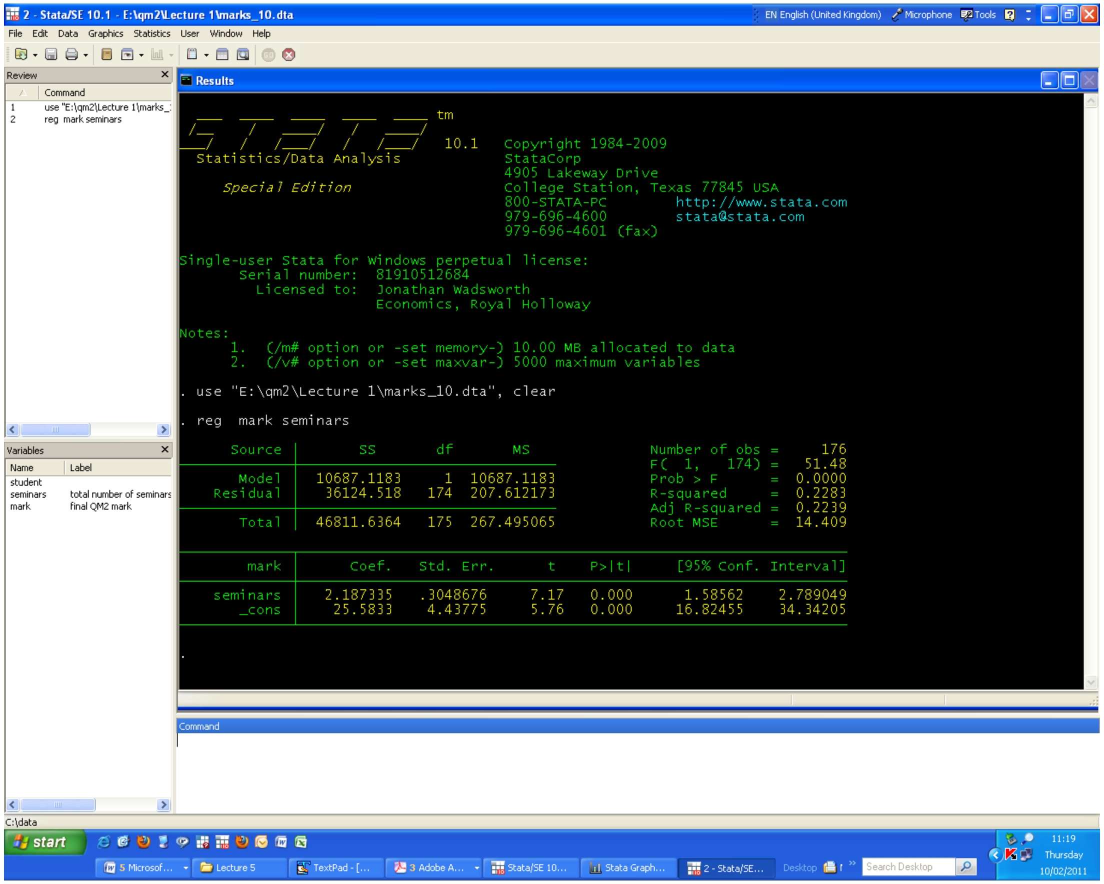
$$F = \frac{ESS / k-1}{RSS / N-k} \qquad = \frac{(ESS/TSS)/k-1}{(RSS/TSS)/N-k}$$

then $F = \dfrac{R^2 / k-1}{(1-R^2)/N-k}$

which is an equivalent way of doing the test

and still the rule is reject the null hypothesis if

$$\hat{F} > F_{\alpha\_critical}^{(k-1, N-k)}$$



The F test in this example is 51.48

The p value is less than 0.05 which is a quick way of realising that this F value is statistically significant from zero

Which means the chances of this R^2 arising by chance are small (<5%)

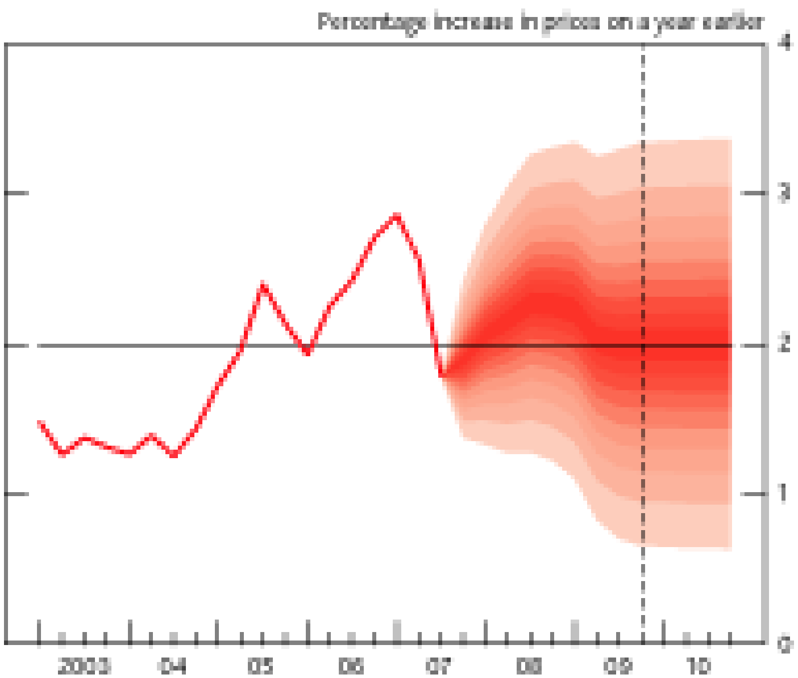
Which means the model has some explanatory power

Forecasting Using OLS

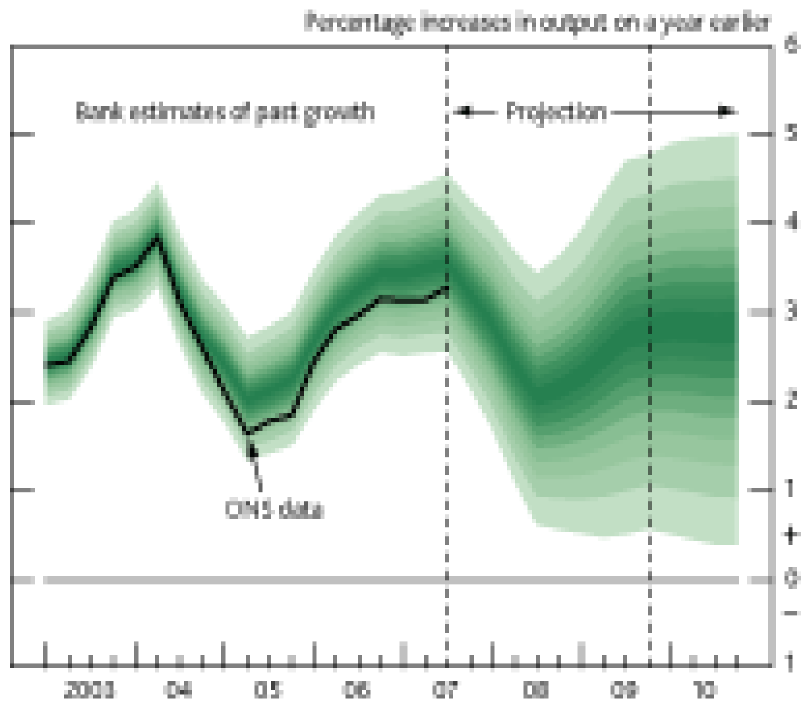
One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

Bank of England forecast of level of inflation (CPI) & GDP growth in November 2007

November 2007 CPI Fan Chart



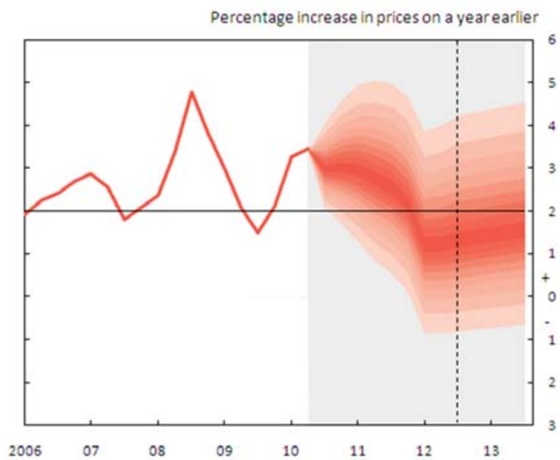
November 2007 GDP Fan Chart



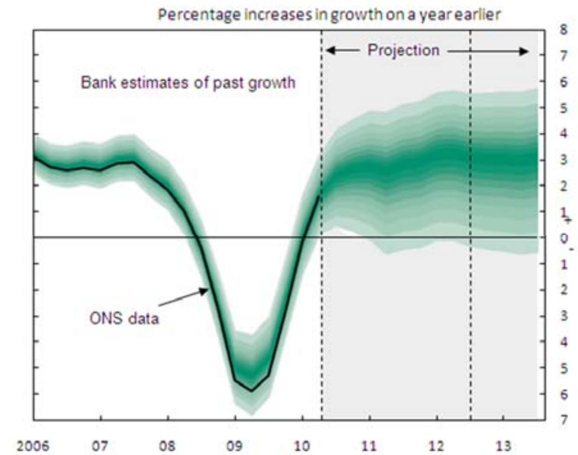
Source Bank of England Inflation report

Actual Outturn

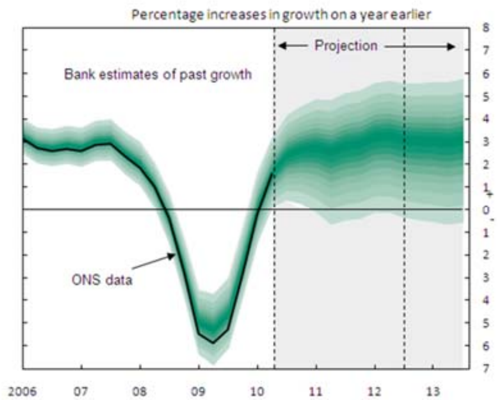
CPI inflation projection based on market interest rate expectations and £200 billion asset purchases

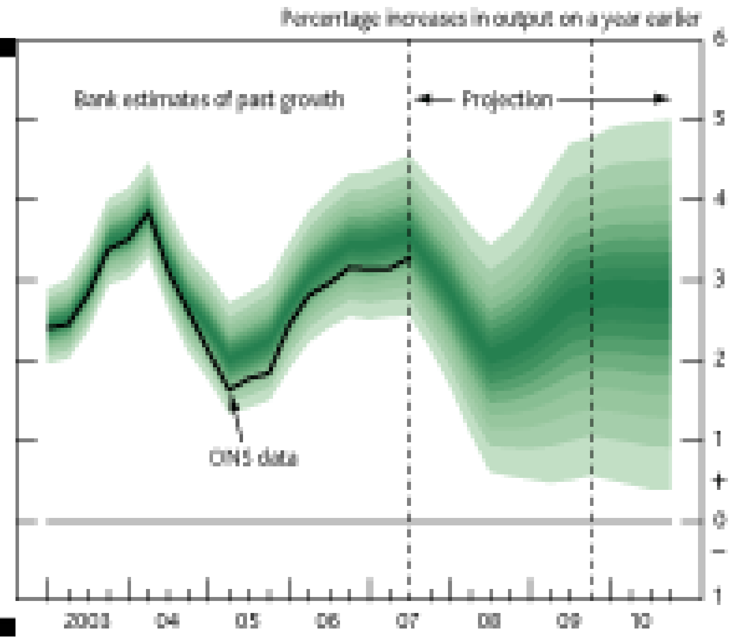


GDP projection based on market interest rate expectations and £200 billion asset purchases



GDP projection based on market interest rate expectations and £200 billion asset purchases





Forecasting Using OLS

One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

Already seen that given a set of estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ and a set of observations on the variable X can use this to predict a value for y

Forecasting Using OLS

One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

Already seen that given a set of estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ and a set of observations on the variable X can use this to predict a value for y

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

Forecasting Using OLS

One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

Already seen that given a set of estimates $\hat{\beta}_0$ and $\hat{\beta}_1$

and a set of observations on the variable X can use this to predict a value for y

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

and can do this for any X value even if the X value did not belong to the original data set

Forecasting Using OLS

One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

Already seen that given a set of estimates $\hat{\beta}_0$ and $\hat{\beta}_1$

and a set of observations on the variable X can use this to predict a value for y

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

and can do this for any X value even if the X value did not belong to the original data set

So it is good practice when estimating equations to leave out some observations (eg end few time periods or a random sub-sample of a cross-section)

If the model is a good one it should predict well out-of-sample.

How do we know how good the prediction is ?

How do we know how good the prediction is ?

- use the idea of a **forecast error**

the closer is the estimate $\hat{y}$ to its actual value, the smaller will be the forecast error

How do we know how good the prediction is ?

- use the idea of a **forecast error**

the closer is the estimate $\hat{y}$ to its actual value, the smaller will be the forecast error which is just the difference between the prediction (the forecast) and the actual (out-of-sample) value

How do we know how good the prediction is ?

- use the idea of a **forecast error**

the closer is the estimate $\hat{y}$ to its actual value, the smaller will be the forecast error which is just the difference between the prediction (the forecast) and the actual (out-of-sample) value

$\hat{u}_0 = \hat{y}_o - y_o$ (o is "out of sample")

How do we know how good the prediction is ?

- use the idea of a **forecast error**

the closer is the estimate $\hat{y}$ to its actual value, the smaller will be the forecast error which is just the difference between the prediction (the forecast) and the actual (out-of-sample) value

$\hat{u}_0 = \hat{y}_o - y_o$ (o is "out of sample")

and if we use OLS to generate the estimates to make the forecasts then can show that the forecast error will on average be zero and have the smallest variance of any other (linear unbiased) technique.

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

using what we know about the expected value of OLS estimates,

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

using what we know about the expected value of OLS estimates,

$E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

using what we know about the expected value of OLS estimates,

$E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$

$$E(\hat{u}_0) = \beta_0 + \beta_1 X_o - \beta_0 - \beta_1 X_o - E(u_0)$$

$$= 0 - E(u_0)$$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

using what we know about the expected value of OLS estimates,

$E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$

$$E(\hat{u}_0) = \beta_0 + \beta_1 X_o - \beta_0 - \beta_1 X_o - E(u_0)$$

$$= 0 - E(u_0)$$

and since Gauss-Markov assumption is that for any observation the expected value of the *true* residual is zero, $E(u_0)=0$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

using what we know about the expected value of OLS estimates,

$E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$

$$E(\hat{u}_0) = \beta_0 + \beta_1 X_o - \beta_0 - \beta_1 X_o - E(u_0)$$

$$= 0 - E(u_0)$$

and since Gauss-Markov assumption is that for any observation the expected value of the *true* residual is zero, $E(u_0)=0$

$E(\hat{u}_0) = 0$ the mean value of any OLS forecast error is zero

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2\left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right]$$

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2 \left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2 \left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2 \left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's
b) the larger the sample size

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2\left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's
b) the larger the sample size
(not expect a good forecast if only have little information)

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2\left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's
b) Larger sample size
(not expect a good forecast if only have little information)
c) Larger variance of in-sample X's

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2 \left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's
c) Larger sample size
(not expect a good forecast if only have little information)
c) Larger variance of in-sample X's
(wider variance means more likely already encountered a big range of possible X values to make a forecast with)

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2 \left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's
b) Larger sample size
(not expect a good forecast if only have little information)

c) Larger variance of in-sample X's
(wider variance means more likely already encountered a big range of possible X values to make a forecast with)

d) Better fitting model (implies smaller RSS and smaller s^2. The better the model is at predicting in-sample, the more accurate will be the prediction out-of-sample)

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Just as

$$\frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$$ follows a t distribution

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Just as

$$\frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$$ follows a t distribution

using $\hat{u}_0$ rather than $\hat{\beta}_1$ then

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Just as

$$\frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$$ follows a t distribution

using $\hat{u}_0$ rather than $\hat{\beta}_1$ then

$$\frac{\hat{u}_0 - \bar{u}_0}{s.e.(\hat{u}_0)}$$ also follows a t distribution

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Just as

$$\frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$$ follows a t distribution

using $\hat{u}_0$ rather than $\hat{\beta}_1$ then

$$\frac{\hat{u}_0 - \bar{u}_0}{s.e.(\hat{u}_0)}$$ also follows a t distribution

and since $\bar{u}_0 = E(u_0) = 0$

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Just as

$$\frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$$ follows a t distribution

using $\hat{u}_0$ rather than $\hat{\beta}_1$ then

$$\frac{\hat{u}_0 - \bar{u}_0}{s.e.(\hat{u}_0)}$$ also follows a t distribution

and since $\bar{u}_0 = E(u_0) = 0$

$$\frac{\hat{u}_0}{s.e.(\hat{u}_0)}$$ also follows a t distribution

Given $\frac{\hat{u}_0}{s.e.(\hat{u}_0)}$

Sub. in $\hat{u}_0 = y_o - \hat{y}_o$

Given $$\frac{\hat{u}_0}{s.e.(\hat{u}_0)}$$

Sub. in $\hat{u}_0 = y_o - \hat{y}_o$

becomes $$\frac{y_o - \hat{y}_o}{s.e.(\hat{u}_0)}$$

Given $\dfrac{\hat{u}_0}{s.e.(\hat{u}_0)}$

Sub. in $\hat{u}_0 = y_o - \hat{y}_o$

becomes $\dfrac{y_o - \hat{y}_o}{s.e.(\hat{u}_0)}$

Since this follows a t distribution we know that 95% of observations will lie in the region

$$\Pr\left[-t_{N-k}^{.05/2} \leq \frac{y_o - \hat{y}_o}{s.e.(\hat{u}_0)} \leq +t_{N-k}^{.05/2}\right] = 0.95$$

Given $\dfrac{\hat{u}_0}{s.e.(\hat{u}_0)}$

Sub. in $\hat{u}_0 = y_o - \hat{y}_o$

becomes $\dfrac{y_o - \hat{y}_o}{s.e.(\hat{u}_0)}$

Since this follows a t distribution we know that 95% of observations will lie in the region

$$\Pr\left[-t_{N-k}^{.05/2} \le \frac{y_o - \hat{y}_o}{s.e.(\hat{u}_0)} \le +t_{N-k}^{.05/2} \right] = 0.95$$

and rearranging terms gives

$$\Pr[\hat{Y}_o - t_{N-k}^{\alpha/2} s.e.(\hat{u}_o) \le Y_o \le \hat{Y}_o + t_{N-k}^{\alpha/2} s.e.(\hat{u}_o)] = 0.95$$

Given

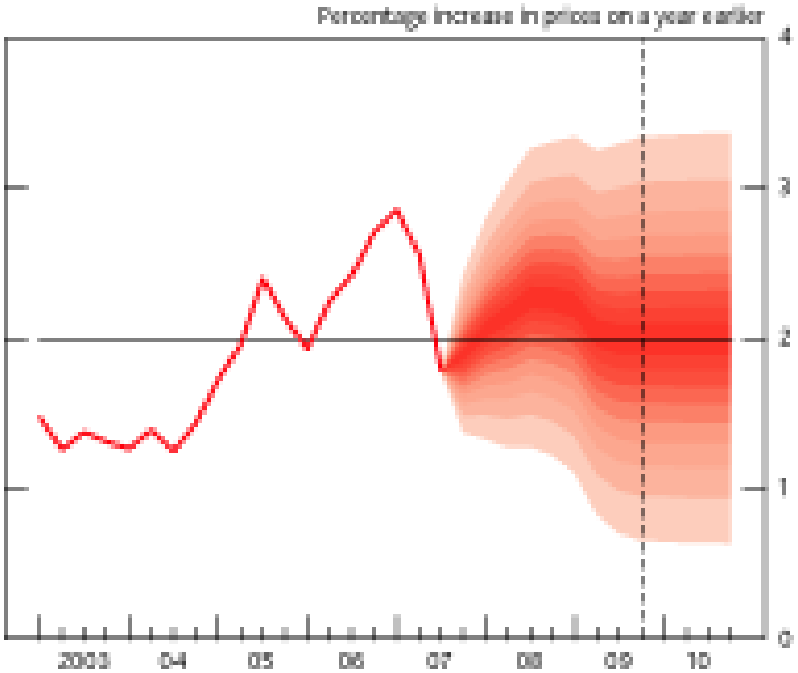
$$\Pr[\hat{Y}_o - t^{\alpha/2}_{N-k}\, s.e.(\hat{u}_o) \le Y_o \le \hat{Y}_o + t^{\alpha/2}_{N-k}\, s.e.(\hat{u}_o)] = 0.95$$

$$= \hat{Y}_o \pm t_{\alpha/2} SE(\hat{u}_o)$$

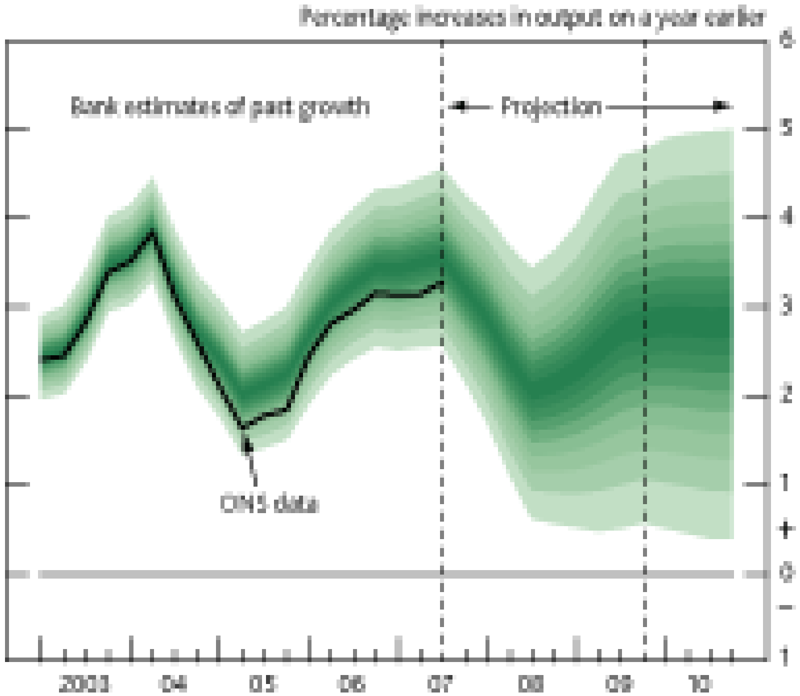
then we can be 95% confident that the true value will lie within this range

If it does not then the model does not forecast very well

November 2007 CPI Fan Chart



November 2007 GDP Fan Chart



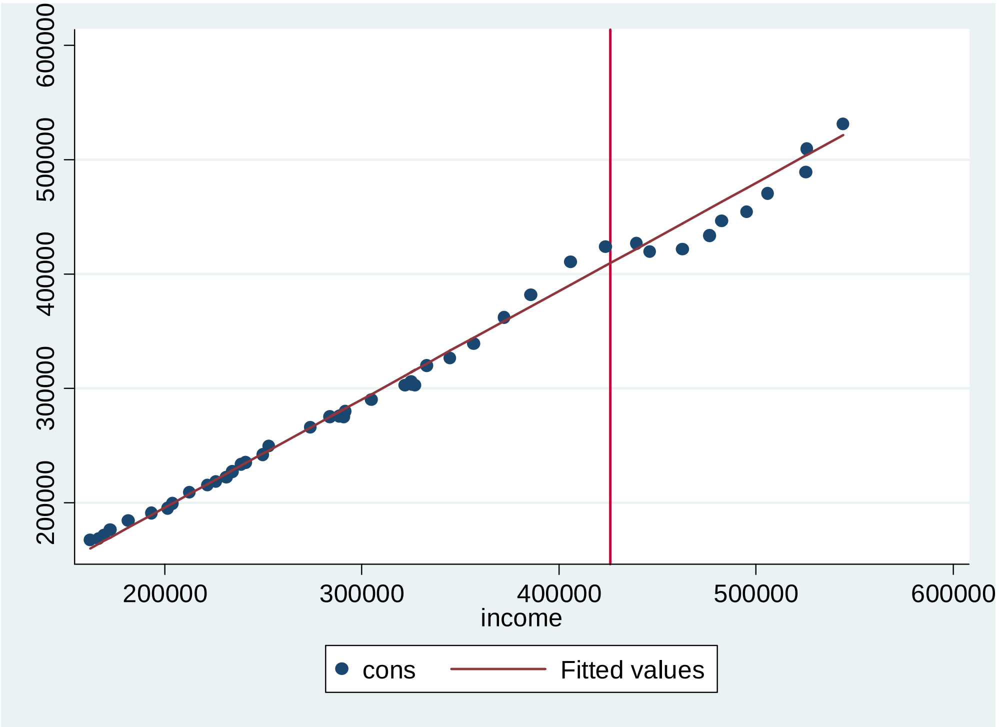
Source Bank of England Inflation report

Example

```
 reg cons income if year<90   /* use 1st 35 obs and save last 10 for forecasts*/

  Source |       SS       df       MS                  Number of obs =      35
---------+------------------------------               F(  1,    33) = 3190.74
   Model |  1.5750e+11     1  1.5750e+11               Prob > F      =  0.0000
Residual |  1.6289e+09    33  49361749.6               R-squared     =  0.9898
---------+------------------------------               Adj R-squared =  0.9895
   Total |  1.5913e+11    34  4.6803e+09               Root MSE      =  7025.8

------------------------------------------------------------------------------
    cons |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  income |   .9467359   .0167604     56.487   0.000       .9126367    .9808351
   _cons |   6366.214   4704.141      1.353   0.185      -3204.433    15936.86

predict chat            /* gets fitted (predicted) values from regression */

predict forcse, stdf    /* gets standard error around forecast value */

/* graph actual and fitted values, draw line through OLS predictions */
```

two (scatter cons income) (line chat income if year<90, xline(426000))

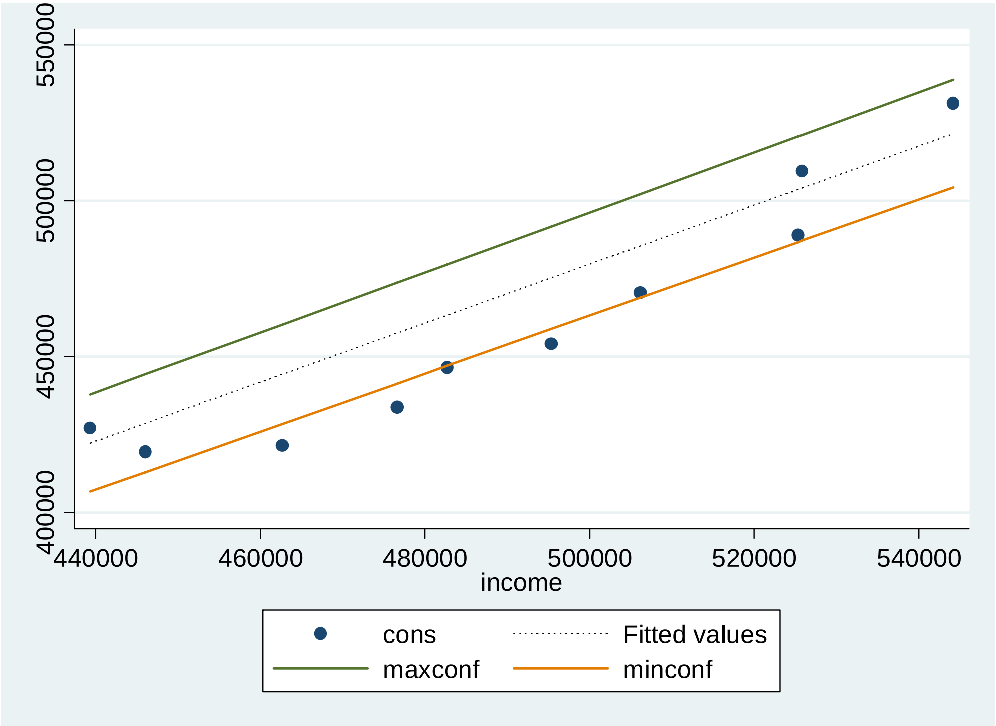
```
twoway (scatter cons income) (line chat income, xline(426000))
```



Vertical line denotes boundary between in and out of sample observations

/* now calculate confidence intervals using cons $=\hat{cons} \pm t_{\partial/2}^{N-k} * SE(\hat{u})$

```
 g minconf= chat - (2.04*forcse)
 g maxconf= chat + (2.04*forcse)

/* graph predicted consumption values and confidence intervals */
twoway (scatter cons income if year>89) (line chat income if year>89, lstyle(dot))  (line maxconf income if year>89)
> (line minconf income if year>89)
```



So for most of the out of sample observations the actual value lies *outside* the 95% confidence interval. Hence the predictions of this particular model are not that good.

Chow Test of Forecast Stability

If a model forecasts well out of sample then we would expect ***all*** the out-of-sample residuals to be close to zero.

Chow Test of Forecast Stability

If a model forecasts well out of sample then we would expect all the out-of-sample residuals to be close to zero.

It can be shown that the joint test of all the out-of-sample-residuals being close to zero is given by:

$$F = \frac{RSS_{in+out} - RSS_{in} / N_o}{RSS_{in} / N - k} \sim F[N_o, N - k]$$

where N_o is the number of out-of-sample observations

N is the number of in-sample observations

k is the number of RHS coefficients

Chow Test of Forecast Stability

If a model forecasts well out of sample then we would expect all the out-of-sample residuals to be close to zero.

It can be shown that the joint test of all the out-of-sample-residuals being close to zero is given by:

$$F = \frac{RSS_{in+out} - RSS_{in} / N_o}{RSS_{in} / N - k} \sim F[N_o, N-k]$$

where N_o is the number of out-of-sample observations
N is the number of in-sample observations
k is the number of RHS coefficients

Intuitively if the model fits well the RSS from the combined regression should be close to that from the in-sample regression.

A "large" difference suggest the RSS are different and so model does not forecast well)

Given a null hypothesis that the model is stable out of sample (predicts well) then if

$$\hat{F} > F_{critical}{}^{\alpha}[N_o, N-k]$$

reject null of model stability out-of-sample

Example:
Estimate of consumption function model including out-of-sample observations is

```
reg cons income if year<90

      Source |       SS       df       MS              Number of obs =      35
-------------+------------------------------           F(  1,    33) = 3190.74
       Model |  1.5750e+11     1  1.5750e+11           Prob > F      =  0.0000
    Residual |  1.6289e+09    33  49361749.6           R-squared     =  0.9898
-------------+------------------------------           Adj R-squared =  0.9895
       Total |  1.5913e+11    34  4.6803e+09           Root MSE      =  7025.8

------------------------------------------------------------------------------
        cons |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .9467359   .0167604    56.49   0.000     .9126367    .9808351
       _cons |   6366.214   4704.141     1.35   0.185    -3204.433    15936.86
------------------------------------------------------------------------------

. reg cons income

      Source |       SS       df       MS              Number of obs =      45
-------------+------------------------------           F(  1,    43) = 5969.79
       Model |  4.7072e+11     1  4.7072e+11           Prob > F      =  0.0000
    Residual |  3.3905e+09    43  78849774.6           R-squared     =  0.9928
-------------+------------------------------           Adj R-squared =  0.9927
       Total |  4.7411e+11    44  1.0775e+10           Root MSE      =  8879.7

------------------------------------------------------------------------------
        cons |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .9172948   .0118722    77.26   0.000     .8933523    .9412372
       _cons |   13496.16   4025.456     3.35   0.002      5378.05    21614.26
------------------------------------------------------------------------------
```

```
Comparing RSS from this with that above can calculate the F value
```

$$F = \frac{RSS_{in+out} - RSS_{in} / N_o}{RSS_{in} / N-k} \sim F[N_o, N-k]$$

```
di  ((3.3905-1.6289)/10)/(1.6289/(35-2))
```

F= $$\frac{3.3905 - 1.6289/10}{1.6289/35-2} \sim F[10,35-2]$$

= 3.57 ~ F[10,33]

From tables $F_{critical}$ $^{.05}$[10,33] = 2.10

So $\hat{F} > F_{critical}$ and therefore **reject** null that model predicts well out of sample.