ECON 4130
Harald Goldstein, revised Nov. 2019

Introduction to F-testing in linear regression models

(Lecture note to lecture Monday 18.11.2019)

1 Introduction

Some general points:

- An F-test usually is a test in regression models where *several parameters* are involved jointly in the null hypothesis in contrast to a T-test that concerns only *one parameter*.
- The recipe (see page 5) for constructing a F-test in a regression model is based on the source table (also called the anova table) from *two* regression runs, one under the full model, and one under the reduced model (H_0).

- The F-test can often be considered a refinement of the more general likelihood ratio test (LR) considered as a large sample chi-square test (see, e.g., the appendix below, optional reading).

- The F-test can (e.g.) be used as an exact test in the special case that the error term in a regression model is normally distributed. This is in the same way as the T-test for *a single parameter,* being an exact test in a model with normally distributed data, can be considered a refinement of a more general large sample Z-test (usually achieved by an application of Slutsky's lemma).
- The F-test (as the T-test) can, since they are exact tests, be used also for small data sets in contrast to the large sample chi-square tests (and large sample Z-tests), but require additional assumptions of normally distributed data (or error terms).
- Note also that, if the null-hypothesis consists of only *one parameter*, then the F and T test statistics satisfy $F = T^2$ exactly, so that a two-sided T-test with *d* degrees of freedom is equivalent to a F-test with 1 and *d* degrees of freedom.

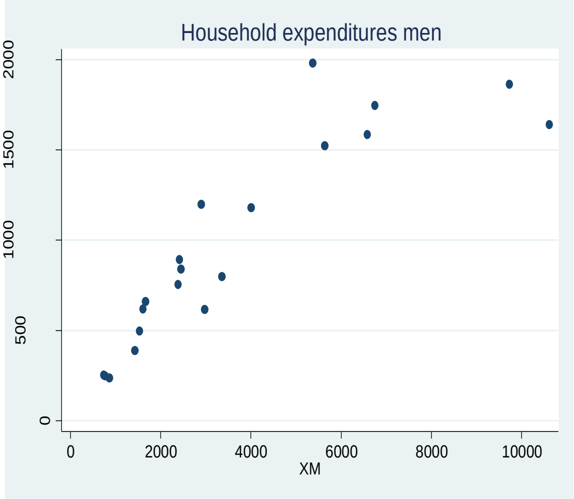
Example from no-seminar exercise week 41 (Hong Kong consumer data).

$Y_i =$ Consumption (men): housing, including fuel and light.

$X_i =$ Income (i.e., we use total expenditure as a proxy).

$i = 1, 2, \dots, n$ where $n = 20$ consumers.

	Lower inc. (< 5000)		Higher inc. (> 5000)	
	Y =cons.	X=inc.	Y=cons.	X=inc.
1	497	1532	1585	6582
2	839	2448	1641	10615
3	798	3358	1981	5371
4	892	2416	1746	6748
5	755	2385	1865	9731
6	388	1429	1524	5637
7	617	2972		
8	248	773		
9	1180	4004		
10	619	1606		
11	253	738		
12	661	1659		
13	238	864		
14	1199	2899		



Testing of structural break as an example of F-testing

This is a typical F-test type of problem in a regression model.

Full model (including the possibility of a structural break between lower and higher incomes)

Suppose $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ are *iid* pairs as $(X,Y) \sim f(x,y) = f(y\,|\,x)f_X(x)$ (where $f(x,y)$ denotes the joint population pdf of (X,Y).

As discussed before (see appendix 1 in the lecture note on prediction and the *iid* models), when all parameters of interest are contained in the conditional pdf $f(y\,|\,x)$, we do not need to say anything about the marginal pdf $f_X(x)$, and we can consider all X_i as fixed equal to their observed values, x_i.

Let D be a dummy for higher income, $D = \begin{cases} 1 & \text{if } X > 5000 \\ 0 & \text{if } X \le 5000 \end{cases}$

Note that D is a function of X.

For using the F-test we need to postulate[1] *a normal and homoscedastic* pdf for $f(y\,|\,x)$, i.e., $(Y\,|\,X=x) \sim N\left(E(Y\,|\,x),\ \sigma^2\right)$, where

---

[1] The F-test is an exact test under normality of the error term. If this assumption is doubtful, then one may use a likelihood ratio test (LR), based on asymptotical normality of estimators, which provides an approximate test for moderately large n. The LR test uses the χ^2_{df} distribution for providing the critical value, where the degrees of freedom (df) is equal to the df of the numerator of the corresponding F-test – i.e., equal to the number of independent restrictions tested.

$$E(Y \mid x) = \beta_0 + \beta_1 x + \beta_2 d + \beta_3 dx = \begin{cases} (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x & \text{if } d = 1, \text{ i.e., for } x > 5000 \\ \beta_0 + \beta_1 x & \text{if } d = 0, \text{ i.e., for } x \le 5000 \end{cases}$$

indicating a structural break if at least one of β_2, β_3 is different from zero.

> **Note.** It is common to separate between two types of variables in regression models like this, *explanatory* and *regressor* variables. The present model has one explanatory variable, *x*, and three regressor variables, $x_1 = x,\ x_2 = d,\ x_3 = x \cdot d$.

Considering the observed *X*'s as fixed, we may express the corresponding model for the data as

(1) $Y_i = \beta_0 + \beta_1 x_i + \beta_2 d_i + \beta_3 d_i x_i + e_i$ where $e_1, e_2, \ldots, e_n \sim iid$ with $e_i \sim N(0, \sigma^2)$.

(or, in terms of the regressor variables, $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$)

We want to test the null hypothesis of no structural break as expressed by the

Reduced model

(2) $Y_i = \beta_0 + \beta_1 x_i + e_i$ where $e_1, e_2, \ldots, e_n \sim iid$ with $e_i \sim N(0, \sigma^2)$.

which is the same as testing

$H_0 : \beta_2 = 0$ and $\beta_3 = 0$ against $H_1 :$ At least one of $\beta_2, \beta_3 \ne 0$ (i.e., the full model).

We see that H_0 here contains two restrictions on the betas – so an F-test is proper here.

The F-test has a simple recipe, but to understand this we need to define the F-distribution and 5 simple facts about the multiple (homoscedastic) regression model with *iid* and normally distributed error terms. First the F-distribution:

2 Introduction to the F-distribution

(see Rice, section 6.2, optional reading)

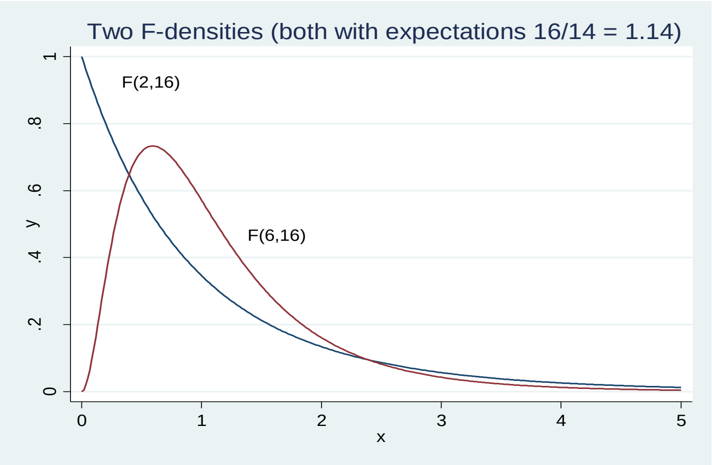
Definition. If Z_1, Z_2 are independent and chi-square distributed with r_1, r_2 degrees of freedom (df) respectively (in short $Z_j \sim \chi^2_{r_j},\ j = 1, 2$), then

$F = \dfrac{Z_1/r_1}{Z_2/r_2}$ has a distribution called the F-distribution with r_1 and r_2 degrees of freedom (in short $F \sim F(r_1, r_2)$).

[Pdf (**optional reading**):

$$f_F(x) = \frac{\Gamma\left(\frac{1}{2}(r_1 + r_2)\right)}{\Gamma\left(\frac{1}{2}r_1\right)\Gamma\left(\frac{1}{2}r_1\right)}\left(\frac{r_1}{r_2}\right)^{r_1/2} \cdot x^{\frac{r_1}{2}-1}\left[1 + (r_1/r_2)x\right]^{-\frac{r_1+r_2}{2}} \quad \text{for } x > 0$$

($f_F(x) = 0$ for $x \le 0$) Expectation: $\dfrac{r_2}{r_2 - 2}$ for $r_2 > 2$]



Two F-densities (both with expectations 16/14 = 1.14)

Notes

- The F-distribution is a one-topped non-symmetric distribution on the positive axis concentrated around 1 (note that, since $E(Z_j) = df = r_j$, then $E\left(Z_j/r_j\right) = 1$).
- If $F \sim F(r_1, r_2)$, then $1/F \sim F(r_2, r_1)$ (follows directly from definition).
- Table 5 in the back of Rice gives only upper percentiles for various F-distributions. If you need lower percentiles, use the previous property (a lower percentile of F is an upper percentile of $1/F$).

The basic tool for performing a F-test is the “Source table” in a Stata-output[2], which summarizes various measures of variation relevant to the analysis.

---

[2] Other programs call this “Anova table”. Anova stands for “analysis of variance”.

Full model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 d_i + \beta_3 d_i x_i + e_i \quad \text{where} \quad e_1, e_2, \ldots, e_n \sim iid \quad \text{with } e_i \sim N(0, \sigma^2)$$

Stata output full model. **(OLS - command:** `regr Y D DX X` **)**

```
      Source |       SS       df       MS  (=SS/df)    Number of obs =      20
-------------+------------------------------           F(  3,    16) =   68.92
       Model |  5784808.74     3  1928269.58           Prob > F      =  0.0000
    Residual |  447637.457    16   27977.341           R-squared     =  0.9282
-------------+------------------------------           Adj R-squared =  0.9147
       Total |   6232446.2    19  328023.484           Root MSE      =  167.26

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   1639.755   283.2312     5.79   0.000     1039.331    2240.178
          DX |  -.2745789   .0572058    -4.80   0.000    -.3958499    -.153308
           X |   .2742643   .0459396     5.97   0.000     .1768768    .3716518
       _cons |   86.25502   105.3841     0.82   0.425    -137.1493    309.6594
------------------------------------------------------------------------------
```

Recipe for the F-test of the reduced model against the full model

- Run **two regressions**, one for the full model and one for the reduced.
- Pick out **the residual** sums of squares (i.e., $SS_{residual}$ that we call SS_{full} and SS_{red} respectively) from the two source tables.
- Pick out **the residual** degrees of freedom (i.e., $df_{residual}$ that we call df_{full} and df_{red} respectively) from the two source tables and calculate the number of restrictions to be tested, $s = df_{red} - df_{full}$.
- Calculate the F statistic, $F = \dfrac{(SS_{red} - SS_{full})/s}{SS_{full}/df_{full}}$, which, under H_0, is $F(s, df_{full})$ distributed. Then reject H_0 if F is larger than the upper $1-\alpha$ percentile in the $F(s, df_{full})$ distribution (corresponding to the level of significance, α).
- Or calculate the p-value, $P_{H_0}(F > F_{obs})$ (using e.g., the F.DIST function in Excel or one of the functions, “F” (the cdf) or “Ftail = 1-F”, in Stata).

[**Example:** The F-test reported (in red) is test for all the regression coefficients in front of explanatory variables, i.e., $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ against some $\beta_j\text{'s} \neq 0$. This is a standard F-test in all OLS-outputs. Non-rejection of this test indicates that there is no evidence in the data that the explanatory variables have any explanatory power at all– thus indicating that further analysis may be futile.]

The source tables of the two regression runs are all that we need for performing a F-test.

3 Some basic facts about the regression model and the source table

First a summary of OLS **(**See lecture note to Rice chap. 8, example 2 (p.10) for a matrix formulation):

Model.

(1) $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + e_i \quad i = 1, 2, \ldots, n$

where the $\{x_{ij}; i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, k\}$ are considered *fixed* numbers and represent n observations of k explanatory variables, $X_1, X_2, \ldots, X_k$ (see justification in the appendix of the lecture note on prediction). For the error terms we assume, $e_1, e_2, \ldots, e_n$ are iid and normally distributed, $e_i \sim N(0, \sigma^2)$.

The error terms (being non observable since the beta's are unknown) can be written

(2) $e_i = Y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik} = Y_i - E(Y_i)$

The OLS estimators (equal to the mle estimators in this model) are determined as minimizing

(3) $Q(\beta) = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik} \right)^2 = \sum_{i=1}^{n} e_i^2$

with respect to $\beta = (\beta_0, \beta_1, \ldots, \beta_k)$. The solution to this minimization problem (which is always unique unless there is an exact linear relationship in the data between some of the *X*-variables,) are the OLS estimators, $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$, satisfying the $k+1$ so called "normal equations":

(4) $\frac{\partial}{\partial \beta_j} Q(\hat{\beta}) = 0, \quad j = 0, 1, 2, \ldots, k$

We define the "predicted *Y*'s" and residuals as respectively

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$, and $\hat{e}_i = Y_i - \hat{Y}_i, \quad i = 1, 2, \ldots, n$

The normal equations (4) can be expressed in terms of the residuals as (defining, for convenience, a constant term variable, $x_{i0} = 1$),

(5) $\sum_{i=1}^{n} \hat{e}_i x_{ij} = 0$ for $j = 0, 1, 2, \ldots, k$

In particular, the first normal equation in (5) shows that $\sum_{i=1}^{n} \hat{e}_i = \sum_{i=1}^{n} \hat{e}_i x_{i0} = 0$, and, therefore[3] that the mean of the *Y*'s must be equal to the mean of the predicted *Y*'s,

(6) $\bar{Y} = \bar{\hat{Y}}$. (Notice $\bar{\hat{Y}} = \sum_i \hat{Y}_i \big/ n = \sum_i (Y_i - \hat{e}_i) \big/ n = \sum_i Y_i \big/ n = \bar{Y}$)

We now introduce the relevant sums of squares (SS's) which satisfy the same (fundamental) relationship (fact 1) as in the simple regression with one explanatory variable:
Define

- Total sum of squares, $SS_{tot} = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$
- Residual sum of squares, $SS_{res} = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = Q(\hat{\beta})$
- Model sum of squares, $SS_{\mathrm{model}} = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{\hat{Y}}\right)^2 \overset{(6)}{=} \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$

Writing $Y_i - \bar{Y} = Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y}$, squaring, and using a little bit of simple (matrix) OLS – algebra, we get the fundamental (and basis for the Source table)

Fact 1: $SS_{tot} = SS_{\mathrm{model}} + SS_{res}$

or $$\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{\hat{Y}}\right)^2 + \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

where $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$ (explained), and $\hat{e}_i = Y_i - \hat{Y}_i$ (unexplained), $i = 1, 2, \ldots, n$

- Often SS_{model} is interpreted as measuring the variation of the "explained part" ($\hat{Y}_i$) of the response Y_i , and SS_{res} as the variation of the "unexplained part" of Y_i. Introducing $R^2 = SS_{\mathrm{model}} \big/ SS_{tot}$ we get the so called "*coefficient of determination*" interpreted as the percentage (i.e., $100 \cdot R^2$) of the total variation of *Y* "explained" by the *k* regressors, $X_1, X_2, \ldots, X_k$, in the data.

- It can also be shown that, defining *R* as the sample correlation between, Y_i and $\hat{Y}_i$ (called the (sample) *multiple correlation* between *Y* and $X_1, X_2, \ldots, X_k$), then R^2 is exactly equal to the definition given. In the Stata output R^2 is reported to the right of the Source table. *R* being a correlation coefficient implies that $R^2 \leq 1$.

---

[3] Whenever the regression function has a constant term, β_0, and only then.

To do inference we also need to know **the distributional properties of the SS's**. First of all, they can be used to estimate the error variance, σ^2, under various circumstances. Notice first (see section 6 below) that $e_i \sim N(0,\sigma^2) \Rightarrow e_i/\sigma \sim N(0,1) \Rightarrow (e_i/\sigma)^2 \sim \chi_1^2$ (as shown in Rice, page 61, as an example, or in the supplementary lecture note on the mgf of the gamma distribution). Since a sum of independent chi-square variables is itself chi-square with degrees of freedom equal to the sum of degrees of freedom for each variable (recall also that the expected value of chi-square variable is equal to the degree of freedom), we have

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} e_i^2 \sim \chi_n^2 \Rightarrow E\left(\frac{1}{\sigma^2}\sum_{i=1}^{n} e_i^2\right) = n \Rightarrow E\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2\right) = \sigma^2$$

Hence, if we could observe the e_i 's, we could use $\frac{1}{n}\sum_{i=1}^{n} e_i^2$ as an unbiased estimator of σ^2.

The e_i 's being non observable, we use the residuals, $\hat{e}_i$'s, instead. The normal equations (5) show that the residuals must satisfy $k+1$ restrictions, $\sum_{i=1}^{n}\hat{e}_i x_{ij} = 0$ for $j = 0,1,2,\ldots,k$, so only $n-k-1$ residuals can vary freely. Hence the term "degree of freedom", being $df_{res} = n-k-1$ for the residuals.

Fact 2 If the regression function contains $k+1$ free parameters, $\beta = (\beta_0, \beta_1, \ldots, \beta_k)$, then $df_{res} = n-k-1 = n$ minus [no. of free parameters, $k+1$, in the regression function].

Now the matrix OLS algebra (details omitted) gives us fact 3 showing that SS_{res}/σ^2 is chi-square distributed with $n-k-1$ degrees of freedom,

Fact 3 $\dfrac{SS_{res}}{\sigma^2} = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}\hat{e}_i^2 \sim \chi_{n-k-1}^2 = \chi_{df_{res}}^2$

$$\Rightarrow E\left(\frac{SS_{res}}{\sigma^2}\right) = n-k-1 \ (= df_{res}) \Rightarrow E\left(\frac{SS_{res}}{df_{res}}\right) = \sigma^2$$

Hence, defining the mean sum of squared residuals as $\tilde{\sigma}^2 = MS_{res} = SS_{res}/df_{res} = SS_{res}/(n-k-1)$, we have obtained *an unbiased estimator* of σ^2,

(7) $\tilde{\sigma}^2 = MS_{res} = SS_{res}/df_{res} = Q(\hat{\beta})/df_{res}$

(Note in contrast that the mle estimator is $\hat{\sigma}^2 = SS_{res}/n$ (shown in the appendix).)

Fact 4 (i) SS_{res} and SS_{model} are independent rv's.

(ii) If all $\beta_1, \beta_2, \ldots, \beta_k$ are 0, then $SS_{\text{model}} / \sigma^2 \sim \chi^2_k \quad \Rightarrow \quad E\left(SS_{\text{model}} / k\right) = \sigma^2$

Otherwise, if some $\beta_j \neq 0$, $E\left(SS_{\text{model}} / k\right) > \sigma^2$

All the information in facts 1,2,…,5 is summarized in the Source table[4] constructed as follows,

(8) **The Source table**

Source	SS	df	MS=SS/df
Model	SS_{model}	$df_{\text{model}} = k$	MS_{model}
Residual	SS_{res}	$df_{res} = n - k - 1$	MS_{res}
Total	$SS_{tot} = \sum_i (Y_i - \bar{Y})^2$	$n - 1$	MS_{tot}

The Source table for the full model (1) in the example - together with the diagnostic information to the right - became

(9) **The Source table for the full model (1)**

```
      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  3,    16) =   68.92
       Model |  5784808.74     3  1928269.58           Prob > F      =  0.0000
    Residual |  447637.457    16   27977.341           R-squared     =  0.9282
-------------+------------------------------           Adj R-squared =  0.9147
       Total |   6232446.2    19  328023.484           Root MSE      =  167.26
```

- According to this, the estimate of the error variance, σ^2, is 27 977.484. The square root of this (167.26) is the estimate of σ and is given as Root MSE to the right.
- The F-test for the H_0^{base} (consisting of 3 restrictions) is at the right and has a p-value 0.0000, indicating that the (3) explanatory variables have explanatory power, so it makes sense to continue the analysis.
- R-squared is simply $SS_{\text{model}} / SS_{tot}$ and shows that 92.82% of the variation in the data of Y_i is explained by the 3 variables in the model[5] (all determined by our single *X*).

---

[4] This source table represent a regression model *with* a constant term (β_0). If the regression function contains *k X's* only *without* a constant term, the source table is slightly different. Then $SS_{tot} = \sum_i Y_i^2 \; \left(= SS_{pred} + SS_{res}\right)$, $df_{res} = n - k,$ $df_{pred} = k,$ and $df_{tot} = n$. Otherwise, the same.

[5] I.e., in this case all 3 variables in the regression function (usually called "regressor" variables) are actually determined by a single *X*. This is ok, however, as long as the three resulting (regressor) variables are not exactly linearly dependent. If they had been exactly linearly dependent, the model becomes non identifiable and OLS brakes down (actually, there will then be infinitely many equivalent solutions to the minimization of *Q*, i.e., what Frisch called "multicollinearity").

- Also the adjusted R-square[6] is a diagnostic tool. If the difference between the two R-squares is substantial, this is a sign that too many explanatory variables have been included in the model in relation to the number of observations (*n*). (In the extreme case, for example, that we include $n-1$ *X*'s in the model, we get all $Y_i = \hat{Y}_i \quad \Rightarrow \quad \text{all } \hat{e}_i = 0 \quad \Rightarrow \quad SS_{res} = 0$ and, therefore, $R^2 = 1$. In this case the regression analysis collapses completely, i.e., there is no information at all in the data for such a model.) In the present example there is no danger of such a possibility since both values are quite close.

4 The recipe for F-testing of regression coefficients

The full Model is as in (1)

(10) $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + e_i \quad i = 1, 2, \ldots, n$

where the $\{x_{ij}; i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, k\}$ are considered *fixed* numbers and represent *n* observations of *k* explanatory variables, $X_1, X_2, \ldots, X_k$ (see justification in the appendix of the lecture note on prediction). For the error terms we assume, $e_1, e_2, \ldots, e_n$ are iid and normally distributed, $e_i \sim N(0, \sigma^2)$.

The reduced Model

We want to test a null hypothesis consisting of *s* (linear and independent) restrictions[7] on $\beta_0, \beta_1, \ldots, \beta_k$. When the restrictions are linear, the model under H_0 can be expressed as a regression model (called the "reduced model") with *p* regressor variables – some of which may be different from the *X*'s (see the exercise (II) in the no-seminar week 48 for an example) – and $p+1$ regression parameters, $\eta' = (\eta_0, \eta_1, \ldots, \eta_p)$, (with η_0 a constant term if present), where $p < k$.

> [**For example**: Suppose the full model is $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$, and we want to test $H_0 : \beta_1 = \beta_2$ (call the common value $\bar{\beta}$, say). Then the reduced model becomes, $Y = \beta_0 + \bar{\beta} X_1 + \bar{\beta} X_2 + \beta_3 X_3 + e = \beta_0 + \bar{\beta}(X_1 + X_2) + \beta_3 X_3 + e$. Then $\eta' = (\beta_0, \bar{\beta}, \beta_3) = (\eta_0, \eta_1, \eta_2)$, and *p* = 2 and $s = 1$.

---

[6] **For the curious ones:** We have $R^2 = \frac{SS_{\text{model}}}{SS_{tot}} = 1 - \frac{SS_{res}}{SS_{tot}}$.

The formula for R^2_{adj} is, $R^2_{adj} \overset{DEF}{=} 1 - \frac{SS_{res} / df_{res}}{SS_{tot} / df_{tot}} = 1 - \frac{n-1}{n-k-1}(1 - R^2)$

[7] A linear restriction on *r* parameters, $\theta_1, \theta_2, \ldots, \theta_r$, has the general form, $a_0 + a_1\theta_1 + a_2\theta_2 + \cdots + a_r\theta_r = 0$, where $a_0, a_1, a_2, \ldots, a_r$ are known constants. Several restrictions are *independent* if no one of them can be derived by the others.

The full analysis is OLS –regression of Y on X_1, X_2, X_3 (with $df_{res} = df_{full} = n-4$). The reduced analysis is achieved by OLS of Y on two variables, $(X_1 + X_2)$ and X_3 (with $df_{res} = df_{red} = n-3$)]

Let SS_{full}, SS_{red} denote the *residual* sum of squares (SS_{res}) for the full model and the reduced model respectively and the corresponding degrees of freedom (in the case that a constant occurs in both the full and the reduced model – otherwise, see footnote 3), $df_{full} = n\text{-}k\text{-}1$ and $df_{red} = n-p-1$. The likelihood ratio principle tells us (see the appendix) that we should compare SS_{full} and SS_{red} to test the reduced model against the full model. This is exactly what the F-test does.

The matrix OLS algebra (details omitted) gives us what we need for the F-test in fact 5:

Fact 5 (i) The rv's SS_{full} and $SS_{red} - SS_{full}$ **are independent**.

(ii) If H_0 (the reduced model) **is true**, then $(SS_{red} - SS_{full})/\sigma^2$ is chi-square distributed with degree of freedom (equal to the expected value) equal to $s = df_{red} - df_{full}$ (valid in general with or without constant terms in the two models).

(iii) If H_0 **is false,** then $(SS_{red} - SS_{full})/\sigma^2$ tends to get larger values than what is likely in the χ_s^2 distribution

Hence, $(SS_{red} - SS_{full})/s$ is an unbiased estimator of σ^2 **if** H_0 **is true**, and, as can be proven, has expectation $> \sigma^2$ **if** H_0 **is wrong**. Since, in any case, SS_{full}/σ^2 is chi-square with degree of freedom df_{full} , and, hence, $\tilde{\sigma}^2 = SS_{full}/df_{full}$ unbiased (and consistent),we get our F test statistic

$$F = \frac{(SS_{red} - SS_{full})/s}{SS_{full}/df_{full}} = \frac{(SS_{red} - SS_{full})/(\sigma^2 s)}{SS_{full}/(\sigma^2 df_{full})} = \frac{Z_1/s}{Z_2/df_{full}} ,$$

where Z_1, Z_2 are independent and, under H_0, chi-square with s and df_{full} degrees of freedom respectively.

Then, according to the construction in section 2, F is F-distributed with $s = df_{red} - df_{full}$ and df_{full} degrees of freedom if H_0 is true. If H_0 is wrong, the F tends to get larger, so we reject H_0 if F is sufficiently large.

Note also that $F = \dfrac{(SS_{red} - SS_{full})/s}{\tilde{\sigma}^2}$, where $\tilde{\sigma}^2$ is an unbiased and consistent estimator of σ^2, no matter if H_0 is true or false.

In other words, the recipe of the F-test is as follows:

(11) Recipe for the F-test of the reduced model against the full model

- Run two regressions, one for the full model and one for the reduced.
- Pick out the residual sums of squares (SS_{full} and SS_{red}) from the two source tables.
- Pick out the residual degrees of freedom (df_{full} and df_{red}) from the two source tables and calculate the number of restrictions to be tested, $s = df_{red} - df_{full}$.
- Calculate the F statistic, $F = \dfrac{(SS_{red} - SS_{full})/s}{SS_{full}/df_{full}}$, and reject H_0 if F is larger than the upper $1-\alpha$ percentile in the $F(s, df_{full})$ distribution (corresponding to the level of significance, α).
- Or calculate the p-value, $P_{H_0}(F > F_{obs})$ (using e.g., the F.DIST function in Excel or a similar function in Stata).

Example of testing structural break described in the introduction.

Full model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 d_i + \beta_3 d_i x_i + e_i \quad \text{where} \quad e_1, e_2, \ldots, e_n \sim iid \quad \text{with } e_i \sim N(0, \sigma^2)$$

Stata output full model

```
. regr Y d dx x

      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  3,    16) =   68.92
       Model |  5784808.74     3  1928269.58           Prob > F      =  0.0000
    Residual |  447637.457    16   27977.341           R-squared     =  0.9282
-------------+------------------------------           Adj R-squared =  0.9147
       Total |   6232446.2    19  328023.484           Root MSE      =  167.26

------------------------------------------------------------------------------
          M1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   1639.755   283.2312     5.79   0.000     1039.331    2240.178
          DX |  -.2745789   .0572058    -4.80   0.000    -.3958499    -.153308
          XM |   .2742643   .0459396     5.97   0.000     .1768768    .3716518
       _cons |   86.25502   105.3841     0.82   0.425    -137.1493    309.6594
------------------------------------------------------------------------------
```

Reduced model (H_0)

$$Y_i = \beta_0 + \beta_1 x_i + e_i \quad \text{where} \quad e_1, e_2, \ldots, e_n \sim iid \quad \text{with } e_i \sim N(0, \sigma^2)$$

$$\Leftrightarrow \quad H_0 : \beta_2 = \beta_3 = 0$$

Stata output reduced model

```
. regr Y x

      Source |       SS           df       MS      Number of obs   =        20
-------------+----------------------------------   F(1, 18)        =     62.89
       Model |  4845492.5          1   4845492.5   Prob > F        =    0.0000
    Residual |  1386953.7         18  77052.9834   R-squared       =    0.7775
-------------+----------------------------------   Adj R-squared   =    0.7651
       Total |  6232446.2         19  328023.484   Root MSE        =    277.58

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .1748752   .0220523     7.93   0.000      .128545    .2212053
       _cons |   331.2991   102.3147     3.24   0.005      116.344    546.2543
------------------------------------------------------------------------------
```

The relevant quantities are

$$SS_{full} = 447\,637.457 \qquad df_{full} = 16$$

$$SS_{red} = 1\,386\,953.7 \qquad df_{red} = 18$$

No. of restrictions under H_0**:** $s = df_{red} - df_{full} = 2$

$$F = \frac{(SS_{red} - SS_{full})/s}{SS_{full}/df_{full}} = \frac{(1\,386\,953.7 - 447\,637.457)/2}{447\,637.457/16} = 16.787\ldots$$

$F \sim F(2,16)$ under H_0**.**

P-value (using F.Dist in Excel): $P_{H_0}(F > F_{obs}) = P_{H_0}(F > 16.787) = 0.0001177$, i.e., 0.000, so the evidence for a structural break as defined at 5000 is strong, i.e., the reduced model is rejected.

5. Specification test of same variance in the two income groups

The F-test in section 4 assumes constant error variance, σ^2*, in both groups*. If this assumption is wrong, the F-test in section 4 is invalidated. It is therefore natural to ask if there is any evidence in the data for doubting the constant variance assumption. For this purpose we can use another F test which often can be used to compare the variances in two independent groups.

Let σ_0^2, σ_1^2 be the error term variances for the $d=0$ group and $d=1$ group respectively.

We want to test $H_0: \sigma_0^2 = \sigma_1^2 \quad \Leftrightarrow \quad \dfrac{\sigma_0^2}{\sigma_1^2} = 1$ against $H_1: \dfrac{\sigma_0^2}{\sigma_1^2} \neq 1$

The F test is well suited for this:

- Run two regressions, one for each group.
- Pick out the two MS_{res}, called MS_0 and MS_1 respectively, from the two runs and form the F statistic, $F = \dfrac{MS_0}{MS_1} = \dfrac{SS_0/df_0}{SS_1/df_1}$, where df_0, df_1 are the residual degrees of freedom in the two groups. Note that MS_0 and MS_1 must be independent since they come from two independent groups.
- Since $F = \dfrac{\sigma_1^2}{\sigma_0^2} \cdot \dfrac{SS_0/(\sigma_0^2 df_0)}{SS_1/(\sigma_1^2 df_1)} = \dfrac{\sigma_1^2}{\sigma_0^2} \cdot V$, where $V \sim F(df_0, df_1)$, it follows that $F \sim F(df_0, df_1)$ if H_0 is true.
- The problem is two-sided, so we reject H_0 if $F < c_1$ or $F > c_2$, where the critical values, c_1, c_2 for level of significance α, are determined by $P_{H_0}(F < c_1) = \alpha/2$ and $P_{H_0}(F > c_2) = \alpha/2$.
- Or calculate the p-value: $= 2 \cdot \left(\text{the smallest of } \mathrm{P}_{H_0}(F < F_{obs}) \text{ and } \mathrm{P}_{H_0}(F > F_{obs})\right)$.

Stata output for the example

```
Group D = 0

      Source |       SS       df       MS              Number of obs =      14
-------------+------------------------------           F(  1,    12) =   40.56
       Model |  997175.494     1  997175.494           Prob > F      =  0.0000
    Residual |  295016.506    12  24584.7088           R-squared     =  0.7717
-------------+------------------------------           Adj R-squared =  0.7527
       Total |     1292192    13  99399.3846           Root MSE      =   156.8

------------------------------------------------------------------------------
          M1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          XM |   .2742643   .0430642     6.37   0.000     .1804356    .3680931
       _cons |   86.25502   98.78806     0.87   0.400    -128.9857    301.4957
------------------------------------------------------------------------------

----------------------------------------------------------------------------------------------
----
Group D = 1

      Source |       SS       df       MS              Number of obs =       6
-------------+------------------------------           F(  1,     4) =    0.00
       Model |  2.38293417     1  2.38293417           Prob > F      =  0.9941
    Residual |   152620.95     4  38155.2376           R-squared     =  0.0000
```

```
-------------+------------------------------           Adj R-squared = -0.2500
       Total |  152623.333     5  30524.6667           Root MSE      =  195.33

------------------------------------------------------------------------------
          M1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          XM |  -.0003146   .0398097    -0.01   0.994     -.110844    .1102148
       _cons |    1726.01   307.0134     5.62   0.005     873.6039    2578.415
```

Test: $F = \dfrac{MS_0}{MS_1} \sim F(12,4)$ under H_0. ($\Rightarrow$ $1/F \sim F(4,12)$ under H_0)

The critical values at the 5% level from table 5 back in Rice :

$$P_{H_0}(F > c_2) = 0.025 \quad \Leftrightarrow \quad P_{H_0}(F \le c_2) = 0.975 \quad \Rightarrow c_2 = 8.75$$

$$P_{H_0}(F < c_1) = 0.025 \quad \Leftrightarrow \quad P_{H_0}\left(\frac{1}{F} > \frac{1}{c_1}\right) = 0.025 \quad \Leftrightarrow \quad P_{H_0}\left(\frac{1}{F} \le \frac{1}{c_1}\right) = 0.975$$

$$\Rightarrow \quad \frac{1}{c_1} = 4.12 \quad \Rightarrow c_1 = \frac{1}{4.12} = 0.24$$

so we reject H_0 if $F < 0.24$ or $F > 8.75$.

Observed: $F_{obs} = \dfrac{MS_0}{MS_1} = \dfrac{24584.7088}{38155.2376} = 0.64$

Conclusion: Don't reject H_0.

In other words: Our (full) model in section 4 passed the specification test, which increases its credibility.

6 Some useful facts about chi-square- and T-distributions

(i) $\chi_r^2 = \Gamma(\tfrac{r}{2}, \tfrac{1}{2})$ distributions.

(ii) $Z \sim \chi_r^2 \quad \Rightarrow \quad E(Z) = r, \quad \text{var}(Z) = 2r$

(iii) $X \sim N(0,1) \quad \Rightarrow \quad Z = X^2 \sim \chi_1^2$ (see, e.g., Rice page 61 (example C))

(iv) $Z_1, Z_2, \ldots, Z_k$ independent and

$$Z_j \sim \chi_{r_j}^2, \qquad \Rightarrow \quad Z = \sum_{j=1}^{k} Z_j \sim \chi_r^2, \text{ where } r = r_1 + r_2 + \cdots + r_k$$

(v) Construction of *T*:

If X, Z are independent, $X \sim N(0,1)$, and $Z \sim \chi_r^2$, then $T = \dfrac{X}{\sqrt{Z/r}} \sim t_r$ (i.e., t-distributed with r degrees of freedom (see Rice Chap. 6 (optional reading)).

(vi) From (iii) and section 2 above, we conclude that, if $T \sim t_r$, then $F = T^2 \sim F(1,r)$.

(vii) Testing an individual coefficient, $H_0 : \beta_j = 0$ against $H_1 : \beta_j \ne 0$, we would use a t-test with $r = n - k - 1$ degrees of freedom and test-statistic $T = \dfrac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \sim t_{n-k-1}$

under H_0. This test is equivalent with an $F(1,\ n-k-1)$ - test, since
$F = T^2 \sim F(1,\ n-k-1)$ under H_0.

7 Appendix – The F-test as a likelihood ratio test (optional reading)

Consider the model in (10)

(12) $Y_i = E(Y_i) + e_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + e_i \quad i = 1,2,\ldots,n$, where $e_1, e_2, \ldots, e_n$ are iid and $e_i \sim N(0,\sigma^2)$. This implies that $Y_1, Y_2, \ldots, Y_n$ are independent and
$Y_i \sim N(E(Y_i),\sigma^2)$ for $i = 1,2,\ldots,n$.

The likelihood is (writing $\beta = (\beta_0, \beta_1, \ldots, \beta_k)$)

$$L(\beta,\sigma) = f(y_1, y_2, \ldots, y_n; \beta, \sigma) = \frac{1}{(2\pi)^{n/2}}\frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_i (y_i - E(Y_i))^2} = \frac{1}{(2\pi)^{n/2}}\frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}Q(\beta)}$$

Since $h(x) = e^{-x}$ is a decreasing function, then, whatever the value of σ, the maximum of L over β is obtained by minimizing $Q(\beta)$, i.e., when β is equal to the OLS $\hat\beta$. Hence, the mle $\hat\beta$ is equal to the OLS estimator. We then find the mle of σ^2 by maximizing
$\ln L(\hat\beta,\sigma) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}Q(\hat\beta)$ with respect to σ.
$\frac{\partial}{\partial\sigma}\ln L(\hat\beta,\sigma) = -\frac{n}{\sigma} + \frac{2}{2\sigma^3}Q(\hat\beta) = 0$ gives the mle $\hat\sigma^2 = \frac{1}{n}Q(\hat\beta) = \frac{1}{n}SS_{res}$.

Substituting this in the likelihood, we get the maximum value

(13) $$L(\hat\beta,\hat\sigma) = \frac{1}{(2\pi)^{n/2}}\frac{1}{\hat\sigma^n} e^{-\frac{n}{2Q(\hat\beta)}Q(\hat\beta)} = \frac{1}{(2\pi)^{n/2}}\frac{1}{\left(Q(\hat\beta)/n\right)^{\frac{n}{2}}} e^{-\frac{n}{2}} = \frac{1}{(2\pi)^{n/2}}\frac{n^{\frac{n}{2}}}{\left(Q(\hat\beta)\right)^{\frac{n}{2}}} e^{-\frac{n}{2}}$$

Now let Ω denote the parameter set, (β,σ), under the full model (12), and ω the parameter set, (η,σ), under the reduced model in section 4. Let L_Ω and L_ω be the maximum likelihoods over Ω and ω respectively. The likelihood ratio (LR) then becomes

$$\Lambda = \frac{L_\omega}{L_\Omega} = \frac{\dfrac{1}{(2\pi)^{n/2}} \dfrac{n^{\frac{n}{2}}}{\left(Q(\hat{\eta})\right)^{\frac{n}{2}}} e^{-\frac{n}{2}}}{\dfrac{1}{(2\pi)^{n/2}} \dfrac{n^{\frac{n}{2}}}{\left(Q(\hat{\beta})\right)^{\frac{n}{2}}} e^{-\frac{n}{2}}} = \left(\frac{Q(\hat{\beta})}{Q(\hat{\eta})} \right)^{\frac{n}{2}} = \left(\frac{SS_{full}}{SS_{red}} \right)^{\frac{n}{2}}$$

The LR test tells us to reject the reduced model (H_0) if $W = -2\ln\Lambda = n\ln\left(\frac{SS_{red}}{SS_{full}}\right)$ is sufficiently large, which is the same as saying that H_0 should be rejected if $\frac{SS_{red}}{SS_{full}}$ is sufficiently large (since the ln-function is increasing), or if $\frac{SS_{red}}{SS_{full}} - 1 = \frac{SS_{red} - SS_{full}}{SS_{full}}$ is sufficiently large. This is equivalent to rejecting H_0 if the F statistic, $F = \frac{n-k-1}{s} \cdot \frac{SS_{red} - SS_{full}}{SS_{full}}$ is sufficiently large. The distribution of *F* is known exactly (as a F-distribution) under H_0 - no matter sample size - in contrast to the general LR test which is only approximately a Chi-square test (with degree of freedom *s*) for large samples.