

**ECON 4130**

Harald Goldstein, Nov. 2013

**Introduction to F-testing in linear regression models**

(Lecture note to lecture Friday 15.11.2013)

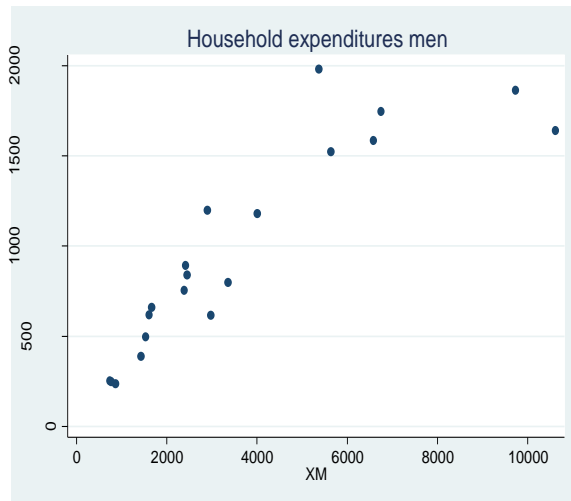
**1 Introduction**

- A F-test usually is a test where *several parameters* are involved at once in the null hypothesis in contrast to a T-test that concerns only *one parameter*.
- 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.
- The F-test can (e.g.) be used 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* in a model with normally distributed data is a refinement of a more general large sample Z-test.
- The F-test (as the T-test) can 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 38 (Hong Kong consumer data).** $Y_i$  = Consumption: 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.

#### The example:

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

Suppose  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  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, 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,  $D = \begin{cases} 1 & \text{if } X > 5000 \\ 0 & \text{if } X \leq 5000 \end{cases}$

Note that  $D$  is a function of  $X$ .

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

$$E(Y|x) = \beta_0 + \beta_1 x + \beta_2 d + \beta_3 dx = \begin{cases} (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x & \text{if } d = 1 \\ \beta_0 + \beta_1 x & \text{if } d = 0 \end{cases}$$

indicating a structural break if at least one of  $\beta_2, \beta_3$  is different from zero.

Considering the observed  $X$ 's as fixed, we may express the model simpler as

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

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

### Reduced model

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

which is the same as testing

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

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

The F-test has a simple recipe, but to understand this we need to define the F-distribution and 4 simple facts about the multiple regression model with *iid* and normally distributed error terms. First the F-distribution:

## 2 Introduction to the F-distribution

(see Rice, section 6.2)

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

$$F = \frac{Z_1/r_1}{Z_2/r_2} \quad \text{has a distribution called the F-distribution with } r_1 \text{ and } r_2 \text{ degrees of}$$

freedom (in short  $F \sim F(r_1, r_2)$ ).

### 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  $E(Z_j/r_j) = 1$ ).
- 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<sup>1</sup>, which summarizes various measures of variation relevant to the analysis. The basis for understanding this table is given in section 3 which you may skip at first if you just wish to know the recipe for performing the F-test in section 4 (in that case, just read Fact 1,2,3, and the Source table (8) in section 3, and come back to the explanation in section 3 later if needed).

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

#### First a summary of OLS

##### Model.

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

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

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

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

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

with respect to  $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ . 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,  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ , satisfying the  $k + 1$  so called “normal equations”:

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

We define the “predicted  $Y$ ’s” and residuals as respectively

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

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

<sup>1</sup> Other programs call this “Anova table”. Anova stands for “analysis of variance”.

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

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<sup>2</sup> that the mean of the  $Y$ 's must be equal to the mean of the predicted  $Y$ 's,

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

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

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:**  $SS_{tot} = SS_{pred} + SS_{res}$

**Note.** Often  $SS_{pred}$  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  $R^2 = SS_{pred} / SS_{tot}$  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,  $X_1, X_2, \dots, X_k$ , 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  $X_1, X_2, \dots, X_k$ ), 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.

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,  $\sigma^2$ , under various circumstances. Notice first 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 the lectures). Since a

---

<sup>2</sup> Whenever the regression function has a constant term,  $\beta_0$ , and only then.

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  $\sigma^2$ .

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, so only  $n-k-1$  can vary freely (hence the term "degree of freedom", being  $df_{res} = n-k-1$  for the residuals). Now the matrix OLS algebra (details omitted) gives us fact 2 showing that  $SS_{res}/\sigma^2$  is chi-square distributed with  $n-k-1$  degrees of freedom,

**Fact 2** 
$$\frac{SS_{res}}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n \hat{e}_i^2 \sim \chi_{n-k-1}^2 \Rightarrow E\left(\frac{SS_{res}}{\sigma^2}\right) = n-k-1 (= df_{res}) \Rightarrow E\left(\frac{SS_{res}}{df_{res}}\right) = \sigma^2$$

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

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

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

**Fact 3**  $SS_{res}$  and  $SS_{pred}$  are independent rv's.

Also  $SS_{pred}$  has similar properties as  $SS_{res}$ , namely that  $\frac{SS_{pred}}{\sigma^2} \sim \chi_k^2$  implying that

$MS_{pred} = SS_{pred}/k$  is an unbiased estimator of  $\sigma^2$ . But this is true only (!) if the hypothesis that all the regression coefficients (excluding the constant term) is zero (i.e.,

$H_0^{base} : \beta_1 = \beta_2 = \dots = \beta_k = 0$  ( $k$  restrictions), which is the same as saying that none of the explanatory variables have explanatory power). If  $H_0^{base}$  is not true, then the OLS algebra

(details omitted) shows that  $E(MS_{pred}) > \sigma^2$ . Thus, comparing  $MS_{pred}$  with  $MS_{res}$  gives

information on  $H_0^{base}$  - leading to a F-test. The test statistic is  $F = MS_{pred}/MS_{res}$  which, under

$H_0^{base}$ , is F-distributed with  $df_{pred} = k$  and  $df_{res} = n-k-1$  degrees of freedom, and we reject

$H_0^{base}$  if  $F$  is sufficiently large. This test is always reported to the right of the Source table and

is taken as a green light for doing the regression if it leads to rejection of  $H_0^{base}$ . If the test

does not reject  $H_0^{base}$ , it is interpreted as that there is too little information (too small  $n$ ) in the data to find any effects among the  $X$ 's on the response,  $Y$ .

All the information in facts 1,2,3 is summarized in the Source table<sup>3</sup> constructed as follows,

(8) **The Source table**

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

- According to this, the estimate of the error variance,  $\sigma^2$ , is 27 977.484. The square root of this (167.26) is the estimate of  $\sigma$  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  $SS_{pred} / SS_{tot}$  and shows that 92.82% of the variation in the data of  $Y_i$  is explained by the 3  $X$ 's in the model.
- Also the adjusted R-square 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  $SS_{res} = 0$  and R-squared = 1, and 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.

<sup>3</sup> This source table represent a regression model *with* a constant term ( $\beta_0$ ). If the regression function contains  $k$   $X$ 's only *without* a constant term, the source table is slightly different. Then  $SS_{tot} = \sum_i Y_i^2$  ( $= SS_{pred} + SS_{res}$ ),  $df_{res} = n - k$ ,  $df_{pred} = k$ , and  $df_{tot} = n$ . Otherwise, the same.

## 4 The recipe for F-testing of regression coefficients

The full Model is as in (1)

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

where the  $\{x_{ij}; i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, k\}$  are considered *fixed* numbers and represent  $n$  observations of  $k$  explanatory variables,  $X_1, X_2, \dots, X_k$  (see justification in the appendix of the lecture note on prediction). For the error terms we assume,  $e_1, e_2, \dots, e_n$  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) restrictions on  $\beta_0, \beta_1, \dots, \beta_k$ . 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 extra exercise in the seminar week 47 for an example) – and  $p+1$  regression parameters,  $\eta = (\eta_0, \eta_1, \dots, \eta_p)$ , (with  $\eta_0$  a constant term if present), where  $p < k$ .

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),  $df_{full} = n - k - 1$  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 4:

**Fact 4** The rv’s  $SS_{full}$  and  $SS_{red} - SS_{full}$  are independent, and, 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).

Hence,  $(SS_{red} - SS_{full}) / s$  is an unbiased estimator of  $\sigma^2$  if  $H_0$  is true, and, as can be proven, has expectation  $> \sigma^2$  if  $H_0$  is wrong. Since, in any case,  $SS_{full} / \sigma^2$  is chi-square with degree

of freedom  $df_{full}$ , we get our F test statistic 
$$F = \frac{(SS_{red} - SS_{full}) / s}{SS_{full} / df_{full}} = \frac{(SS_{red} - SS_{full}) / (\sigma^2 s)}{SS_{full} / (\sigma^2 df_{full})},$$

which, according to the construction in section 2, 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. 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,  $F = \frac{(SS_{red} - SS_{full}) / s}{SS_{full} / df_{full}}$ , 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,  $\alpha$ ).
- 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

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

#### Stata output full model

| Source   | SS         | df | MS         |                 |        |  |
|----------|------------|----|------------|-----------------|--------|--|
| Model    | 5784808.74 | 3  | 1928269.58 | Number of obs = | 20     |  |
| Residual | 447637.457 | 16 | 27977.341  | F( 3, 16) =     | 68.92  |  |
| Total    | 6232446.2  | 19 | 328023.484 | Prob > F =      | 0.0000 |  |
|          |            |    |            | R-squared =     | 0.9282 |  |
|          |            |    |            | Adj R-squared = | 0.9147 |  |
|          |            |    |            | 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 ( $H_0$ )

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

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

**Stata output reduced model**

| Source   | SS         | df | MS         |                 |        |  |
|----------|------------|----|------------|-----------------|--------|--|
| Model    | 4787630.87 | 1  | 4787630.87 | Number of obs = | 20     |  |
| Residual | 1444815.33 | 18 | 80267.5185 | F( 1, 18) =     | 59.65  |  |
| Total    | 6232446.2  | 19 | 328023.484 | Prob > F =      | 0.0000 |  |
|          |            |    |            | R-squared =     | 0.7682 |  |
|          |            |    |            | Adj R-squared = | 0.7553 |  |
|          |            |    |            | Root MSE =      | 283.32 |  |

| M1    | Coef.    | Std. Err. | t    | P> t  | [95% Conf. Interval] |          |
|-------|----------|-----------|------|-------|----------------------|----------|
| D     | 1067.667 | 138.2437  | 7.72 | 0.000 | 777.2275             | 1358.106 |
| _cons | 656      | 75.71918  | 8.66 | 0.000 | 496.9199             | 815.0801 |

The relevant quantities are

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

$$SS_{red} = 1\,444\,815.33 \quad df_{red} = 18$$

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

$$F = \frac{(SS_{red} - SS_{full}) / s}{SS_{full} / df_{full}} = \frac{(1\,444\,815.33 - 447\,637.457) / 2}{447\,637.457 / 16} = 17.82$$

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

P-value (using F.Dist in Excel):  $P_{H_0}(F > F_{obs}) = P_{H_0}(F > 17.82) = 8.49 \cdot 10^{-5} = 0.000$

so the evidence for a structural break as defined 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,  $\sigma^2$ , 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  $\sigma_0^2, \sigma_1^2$  be the error term variances for the  $d=0$  group and  $d=1$  group respectively.

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

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,  $F = \frac{MS_0}{MS_1} = \frac{SS_0 / df_0}{SS_1 / df_1}$ , 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  $F = \frac{\sigma_1^2}{\sigma_0^2} \cdot \frac{SS_0 / (\sigma_0^2 df_0)}{SS_1 / (\sigma_1^2 df_1)} = \frac{\sigma_1^2}{\sigma_0^2} \cdot V$ , 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  $\alpha$ , 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:  $= 2 \cdot (\text{the smallest of } P_{H_0}(F < F_{obs}) \text{ and } P_{H_0}(F > F_{obs}))$ .

### Stata output for the example

Group D = 0

| Source   | SS         | df | MS         | Number of obs = | 14     |
|----------|------------|----|------------|-----------------|--------|
| Model    | 997175.494 | 1  | 997175.494 | F( 1, 12) =     | 40.56  |
| Residual | 295016.506 | 12 | 24584.7088 | Prob > F =      | 0.0000 |
| Total    | 1292192    | 13 | 99399.3846 | R-squared =     | 0.7717 |
|          |            |    |            | Adj R-squared = | 0.7527 |
|          |            |    |            | 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       |
|----------|------------|----|------------|-----------------|---------|
| Model    | 2.38293417 | 1  | 2.38293417 | F( 1, 4) =      | 0.00    |
| Residual | 152620.95  | 4  | 38155.2376 | Prob > F =      | 0.9941  |
| Total    | 152623.333 | 5  | 30524.6667 | R-squared =     | 0.0000  |
|          |            |    |            | Adj R-squared = | -0.2500 |
|          |            |    |            | 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:  $F = \frac{MS_0}{MS_1} \sim F(12, 4)$  under  $H_0$ .

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

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

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

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

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

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

Conclusion: Don't reject  $H_0$ .

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

## 6 Appendix - The F-test as a likelihood ratio test (optional reading)

Consider the model in (10)

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

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

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

Since  $h(x) = e^{-x}$  is a decreasing function, then, whatever the value of  $\sigma$ , the maximum of  $L$  over  $\beta$  is obtained by minimizing  $Q(\beta)$ , i.e., when  $\beta$  is equal to the OLS  $\hat{\beta}$ . Hence the mle  $\hat{\beta}$  is equal to the OLS estimator. We then find the mle of  $\sigma^2$  by maximizing

$$\ln L(\hat{\beta}, \sigma) = -\frac{n}{2} \ln(2\pi) - n \ln \sigma - \frac{1}{2\sigma^2} Q(\hat{\beta}) \text{ with respect to } \sigma.$$

$$\frac{\partial}{\partial \sigma} \ln L(\hat{\beta}, \sigma) = -\frac{n}{\sigma} + \frac{2}{2\sigma^3} Q(\hat{\beta}) = 0 \text{ gives the mle } \hat{\sigma}^2 = \frac{1}{n} Q(\hat{\beta}) = \frac{1}{n} SS_{res}.$$

Substituting this in the likelihood, we get the maximum value

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

Now let  $\Omega$  denote the parameter set,  $(\beta, \sigma)$ , under the full model (12), and  $\omega$  the parameter set,  $(\eta, \sigma)$ , under the reduced model in section 4. Let  $L_\Omega$  and  $L_\omega$  be the maximum likelihoods over  $\Omega$  and  $\omega$  respectively. The likelihood ratio (LR) then becomes

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

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.