Advanced Econometrics

Professor: Julie Shi

Timekeeper: Rui Zhou

Fall 2023

Contents

3 Hypothesis Testing **2**
3.1 Wald Test 2
3.2 Fit-Based Test 6

3 Hypothesis Testing

The general linear hypothesis is a *set* of J restrictions on the linear regression model $y = X\beta + \varepsilon$, which can be written as:

$$
\begin{aligned}
& R\beta = q \\
\iff & \begin{cases} r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1K}\beta_K = q_1 \\ r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2K}\beta_K = q_2 \\ \vdots \\ r_{J1}\beta_1 + r_{J2}\beta_2 + \cdots + r_{JK}\beta_K = q_J \end{cases}
\end{aligned}
$$

The matrix R has K columns to be conformable with β, J rows for a total of J restrictions. R must be *full row rank* to be meaningful, so J must be less than or equal to K. However, the case of $J = K$ must also be ruled out. If the K coefficients satisfy $J = K$ restrictions, then R is square and non-singular, so $\beta = R^{-1}q$. There would be no estimation or inference problem. The restrictions $R\beta = q$ imposes J restrictions on K otherwise free parameters. Hence, with the restrictions imposed, only $K - J$ free parameters are remained.

The hypothesis implied by the restrictions is written

$$
\begin{aligned}
H_0 :& R\beta - q = \mathbf{0} \\
H_1 :& R\beta - q \neq \mathbf{0}
\end{aligned}
$$

We will consider two approaches to testing the hypothesis, Wald tests and fit based tests. The hypothesis characterizes the population. If the hypothesis is correct, then the sample statistics should mimic that description. The tests will proceed as follows:

- Wald tests: The hypothesis states that $R\beta - q$ equals $\mathbf{0}$. The least squares estimator, b, is an unbiased and consistent estimator of β. If the hypothesis is correct, then the sample discrepancy, $R\beta - q$ should be close to zero.
- Fit based tests: We obtain the best possible fit—highest R^2—by using least squares without imposing the restrictions. We will show later that the sum of squares will never decrease when we impose the restrictions.

To develop the test statistics in this section, we will assume normally distributed disturbances. With this assumption, we will be able to obtain the exact distributions of the test statistics.

3.1 Wald Test

The Wald test is the most commonly used procedure. It is often called a "significance test." The operating principle of the procedure is to fit the regression without the restrictions, and then assess whether the results appear, within sampling variability, to agree with the hypothesis.

3.1.1 Single Coefficient

The Wald distance of a coefficient estimate from a hypothesized value is the linear distance, measured in standard deviation units. Thus, the distance of b_k from β_k^0 would be

$$W_k = \frac{b_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}}$$

The following procedure is the same with constructing confidence interval, thus omitted.

3.1.2 Multiple Coefficients

Given the least squares estimator b, our interest centers on the *discrepancy vector* $Rb - q := m$. It is unlikely that m will be exactly $\mathbf{0}$. The statistical question is whether the deviation of m from $\mathbf{0}$ can be attributed to sampling error or whether it is significant.

Since b is normally distributed and m is a linear function of b, m is also normally distributed. If the null hypothesis is true, then $R\beta - q = \mathbf{0}$ and

$$\begin{aligned} \mathrm{E}\left[m|X\right] &= R\mathrm{E}\left[b|X\right] - q = R\beta - q = \mathbf{0} \\ \mathrm{Var}\left[m|X\right] &= \mathrm{Var}\left[Rb - q|X\right] = \mathrm{Var}\left[Rb|X\right] = R\mathrm{Var}\left[b|X\right]R' = \sigma^2 R(X'X)^{-1}R' \end{aligned}$$

We can base a test of H_0 on the *Wald criterion.* Conditioned on X, we find:

$$\begin{aligned} W &= m' \left\{\mathrm{Var}\left[m|X\right]\right\}^{-1} m \\ &= (Rb - q)' \left[\sigma^2 R(X'X)^{-1}R'\right]^{-1} (Rb - q) \\ &= \frac{(Rb - q)' \left[R(X'X)^{-1}R'\right]^{-1} (Rb - q)}{\sigma^2} \sim \chi^2\left[J\right] \end{aligned}$$

The statistic W has a chi-squared distribution with J degrees of freedom if the hypothesis is correct. Intuitively, the larger m is—that is, the worse the failure of least squares to satisfy the restrictions—the larger the chi-squared statistic. Therefore, a large chi-squared value will weigh against the hypothesis.
However, the chi-squared statistic of W is not usable because of the unknown σ^2. By using s^2 instead of σ^2 and dividing the result by J, we obtain a usable F statistic with J and $n - K$ degrees of freedom.

$$\begin{aligned} F &= \frac{W}{J} \cdot \frac{\sigma^2}{s^2} \\ &= \left(\frac{(Rb - q)' \left[R(X'X)^{-1}R'\right]^{-1} (Rb - q)}{\sigma^2}\right) \cdot \frac{1}{J} \cdot \frac{\sigma^2}{s^2} \cdot \frac{n-K}{n-K} \\ &= \frac{(Rb - q)' \left[\sigma^2 R(X'X)^{-1}R'\right]^{-1} (Rb - q)/J}{\left[(n-K)\, s^2/\sigma^2\right]/(n-K)} \end{aligned}$$

If the null hypothesis is true, that is, $R\beta = q$, then

$$Rb - q = Rb - R\beta = R(b-\beta) = R(X'X)^{-1}X'\varepsilon$$

Let $C = R(X'X)^{-1}R'$, then we have

$$\frac{Rb-q}{\sigma} = R(X'X)^{-1}X'\left(\frac{\varepsilon}{\sigma}\right) := D\left(\frac{\varepsilon}{\sigma}\right)$$

Then the numerator of F can be simplified as

$$\begin{aligned}
\frac{(Rb-q)'\left[\sigma^2 R(X'X)^{-1}R'\right]^{-1}(Rb-q)}{J} &= \frac{\left(\frac{Rb-q}{\sigma}\right)'\left\{R(X'X)^{-1}R'\right\}^{-1}\left(\frac{Rb-q}{\sigma}\right)}{J} \\
&= \frac{\left(\frac{\varepsilon}{\sigma}\right)' D'\left\{R(X'X)^{-1}R'\right\}^{-1} D\left(\frac{\varepsilon}{\sigma}\right)}{J} \\
&= \frac{\left(\frac{\varepsilon}{\sigma}\right)' D'C^{-1}D\left(\frac{\varepsilon}{\sigma}\right)}{J} \\
&:= \frac{\left(\frac{\varepsilon}{\sigma}\right)' T\left(\frac{\varepsilon}{\sigma}\right)}{J}
\end{aligned}$$

where $T = D'C^{-1}D$. The numerator is $\frac{W}{J}$ and is distributed as $\frac{1}{J}$ times a $\chi^2[J]$. Also, as we have done in constructing t statistic for testing significance of a single coefficient, the denominator of F what we are familiar with

$$\frac{\frac{(n-K)s^2}{\sigma^2}}{n-K} = \frac{\frac{\varepsilon' M\varepsilon}{\sigma^2}}{n-K} = \frac{1}{n-K}\cdot\left(\frac{\varepsilon}{\sigma}\right)' M\left(\frac{\varepsilon}{\sigma}\right) \sim \frac{1}{n-K}\cdot\chi^2[n-K]$$

Therefore, the F statistic is the ratio of two chi-squared variables each divided by its degrees of freedom. Since $M\left(\frac{\varepsilon}{\sigma}\right)$ and $T\left(\frac{\varepsilon}{\sigma}\right)$ are both normally distributed and their covariance TM is $\mathbf{0}$, the vectors of the quadratic forms are independent. The numerator and denominator of F are functions of independent random vectors and are therefore independent. This completes the proof of the F distribution.

Cancelling terms leaves the F statistic for testing a linear hypothesis:

$$F[J, n-K|X] = \frac{(Rb-q)'\left\{R\left[s^2(X'X)^{-1}\right]R'\right\}^{-1}(Rb-q)}{J}$$

Notice that degrees of freedom in the denominator is set for a certain regression, since $(n-K)$ is "used" to make up for substitute s^2 for σ^2. The only change with regard to different hypothesis is the degrees of freedom in the numerator, which depends on the number of restrictions.

For testing one linear restriction of the form

$$H_0 : r_1\beta_1 + r_2\beta_2 + \cdots + r_K\beta_K = r'\beta$$

The F statistic is them

$$F\left[1, n-K\right] = \frac{\left(\sum_j r_j b_j - q\right)^2}{\sum_j \sum_k r_j r_k \text{Est.Cov}\left[b_j, b_k\right]}$$

If the hypothesis is that the j-th coefficient is equal to a particular value, then R has a single row with a 1 in the j-th position and 0s elsewhere, so $R(X'X)^{-1}R'$ is the j-th diagonal element of the inverse matrix $(X'X)^{-1}$, and $Rb-q$ is (b_j-q). The F statistic is then

$$F\left[1, n-K\right] = \frac{\left(b_j - q\right)^2}{\text{Est.Var}\left[b_j\right]}$$

Consider an alternative approach to test a single restriction of testing $r_1\beta_1 + r_2\beta_2 + \cdots + r_K\beta_K = r'\beta$. The sample estimate of $r'\beta$ is

$$r_1 b_1 + r_2 b_2 + \cdots + r_K b_K = \hat{q} = r'b$$

If $\hat{q}$ differs significantly from q, then we conclude that the sample data are not consistent with the hypothesis. It is natural to base the test on

$$t = \frac{\hat{q} - q}{\text{se}\left(\hat{q}\right)}$$

In words, t statistic is the distance in standard error units between the hypothesized function of the true coefficients and the same function of our estimates of them. If the hypothesis is true, then our estimates should reflect that, at least within the range of sampling variability. Thus, if the absolute value of the preceding t ratio is larger than the appropriate critical value, then doubt is cast on the hypothesis.

We require an estimate of the standard error of $\hat{q}$. Since $\hat{q}$ is a linear function of b and we have an estimate of the covariance matrix of b, which is $s^2(X'X)^{-1}$, we can estimate the variance of $\hat{q}$ with

$$\text{Est.Var}\left[\hat{q}|X\right] = r'\left[s^2(X'X)^{-1}\right]r$$

Therefore, we find

$$
\begin{aligned}
t &= \frac{\hat{q} - q}{\mathrm{se}\left(\hat{q}\right)} = \frac{r'b - q}{\sqrt{r'\left[s^2(X'X)^{-1}\right]r}} \\
t^2 &= \frac{\left(\hat{q} - q\right)^2}{\left[\mathrm{se}\left(\hat{q}\right)\right]^2} = \frac{\left(r'b - q\right)\left(r'b - q\right)}{r'\left[s^2(X'X)^{-1}\right]r} \\
&= \frac{\left(r'b - q\right)\left\{r'\left[s^2(X'X)^{-1}\right]r\right\}^{-1}\left(r'b - q\right)}{1} \\
&= F
\end{aligned}
$$

It follows, therefore, that for testing a single restriction, the t statistic is the square root of the F statistic that would be used to test that hypothesis.

3.2 Fit-Based Test

A different approach to hypothesis testing focuses on the fit of the regression. Recall that the least squares vector b was chosen to minimize the sum of squared deviations, $e'e$. Since R^2 equals $1 - \dfrac{e'e}{y'M_0y}$ and $y'M_0y$ is a constant that does not involve b, it follows that b is chosen to maximize R^2. One might ask whether choosing some other value for the slopes of the regression leads to a significant loss of fit. To develop the test statistic, we first examine the computation of the least squares estimator subject to a set of restrictions. We will then construct a test statistic that is based on comparing the R^2 s from the two regressions.

3.2.1 Constrained Estimator

Suppose that we explicitly impose the restriction of the general linear hypothesis in the regression. The restricted least squares estimator is obtained as the solution to

$$
\begin{aligned}
&\min_{b_0} S(b_0) = (y - Xb_0)'(y - Xb_0) \\
&\text{s.t. } Rb_0 = q
\end{aligned}
$$

A Lagrangian function for this problem can be written

$$
\mathcal{L}(b_0, \lambda) = \left(y - Xb_0\right)'\left(y - Xb_0\right) + 2\lambda'\left(Rb_0 - q\right)^2
$$

(Notice that since λ is not restricted, we formulate the constraints in terms of 2λ for convenience of scaling later.)

The solutions b_0 and λ will satisfy the necessary conditions

$$\frac{\partial \mathcal{L}}{\partial b_0} = -2X'\left(y - Xb_0\right) + 2R'\lambda = \mathbf{0}$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = 2\left(Rb_0 - q\right) = \mathbf{0}$$

which can be expressed in the following partitioned matrix equation:

$$\begin{bmatrix} X'X & R' \\ R & \mathbf{0} \end{bmatrix} \begin{bmatrix} b_0 \\ \lambda \end{bmatrix} = \begin{bmatrix} X'y \\ q \end{bmatrix}$$

If, in addition, $X'X$ is not singular, then explicit solutions for b_0 and λ can be obtained by using the partitioned inverse formula:

$$b_0 = b - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}(Rb - q)$$
$$\lambda = \left[R(X'X)^{-1}R'\right]^{-1}(Rb - q)$$

However, the partitioned inverse formula is unfriendly to remember, so another useful way is to smartly solve the necessary condition equations:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial b_0} = \mathbf{0} \\ \frac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{0} \end{cases} \quad \Longrightarrow \begin{cases} -X'y + X'Xb_0 + R'\lambda = \mathbf{0} \\ Rb_0 = q \end{cases}$$

To make use of the restriction that $Rb_0 = q$, if $X'X$ is invertible, we premultiply $R(X'X)^{-1}$ in the first equation:

$$\begin{aligned}
& -X'y + X'Xb_0 + R'\lambda = \mathbf{0} \\
\Longrightarrow & -R(X'X)^{-1}X'y + R(X'X)^{-1}X'Xb_0 + R(X'X)^{-1}R'\lambda = \mathbf{0} \\
\Longleftrightarrow & -Rb + Rb_0 + R(X'X)^{-1}R'\lambda = \mathbf{0} \\
\Longleftrightarrow & -Rb + q + R(X'X)^{-1}R'\lambda = \mathbf{0} \\
\Longrightarrow & \lambda = \left\{R(X'X)^{-1}R'\right\}^{-1}(Rb - q)
\end{aligned}$$

After solving $\lambda = \left\{R(X'X)^{-1}R'\right\}^{-1}(Rb - q)$, we insert this into the first equation to obtain b_0:

$$\begin{aligned}
& -X'y + X'Xb_0 + R'\lambda = \mathbf{0} \\
\Longleftrightarrow & X'Xb_0 = X'y - R'\lambda \\
\Longleftrightarrow & X'Xb_0 = X'y - R'\left\{R(X'X)^{-1}R'\right\}^{-1}(Rb - q) \\
\Longrightarrow & b_0 = (X'X)^{-1}X'y - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}(Rb - q) \\
& \quad = b - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}(Rb - q)
\end{aligned}$$

Intuitively, the constrained solution b_0 is equal to the unconstrained solution b *minus* a term that accounts for the failure of the unrestricted solution to satisfy the constraints.

To get the variance of b_0, as we have done before, we need to work out the random part of b_0. Under the hypothesis of restriction, $R\beta = q$, so

$$\begin{aligned}
b_0 &= b - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}(Rb - q) \\
&= b - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}(Rb - R\beta) \\
&= b - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R\,(b - \beta) \\
&= b - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}X'\varepsilon \\
&= \beta + (X'X)^{-1}X'\varepsilon - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}X'\varepsilon \\
&= \beta + \Big\{(X'X)^{-1}X' - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}X'\Big\}\varepsilon
\end{aligned}$$

Notice that we should not forget b also contains the random part, so we need to segment it into $b = \beta + (X'X)^{-1}X'\varepsilon$. Then we forward to get the variance of b_0:

$$\begin{aligned}
\mathrm{Var}\,[b_0|X] &= \Big\{(X'X)^{-1}X' - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}X'\Big\}\{\mathrm{Var}\,[\varepsilon|X]\}\Big\{(X'X)^{-1}X' - \\
&= \sigma^2\Big\{(X'X)^{-1}X' - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}X'\Big\}\Big\{(X'X)^{-1}X' - (X'X)^{-1} \\
&= \sigma^2\Big\{(X'X)^{-1}X' - (X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}X'\Big\}\Big\{X(X'X)^{-1} - X(X'X)^{-1} \\
&= \sigma^2\left\{(X'X)^{-1}X'X(X'X)^{-1}\right\} \\
&\quad - \sigma^2\Big\{(X'X)^{-1}X'X(X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}\Big\} \\
&\quad - \sigma^2\Big\{(X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}X'X(X'X)^{-1}\Big\} \\
&\quad + \sigma^2\Big\{(X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}X'X(X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X' \\
&= \sigma^2(X'X)^{-1} - \sigma^2\Big\{(X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}\Big\} - \sigma^2\Big\{(X'X)^{-1}R'\{R(X'X \\
&\quad + \sigma^2\Big\{(X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}\Big\} \\
&= \sigma^2(X'X)^{-1} - \sigma^2\Big\{(X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}\Big\} \\
&= \mathrm{Var}\,[b|X] - \sigma^2\Big\{(X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}\Big\}
\end{aligned}$$

We hope to prove that the second term, $\sigma^2\Big\{(X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}\Big\}$, is a non-negative definite matrix. From the core out, since $X'X$ is a positive definite matrix, so $(X'X)^{-1}$ is a positive matrix. Because R has full row rank, $R(X'X)^{-1}R'$ is then a positive definite matrix, so is $\left\{R(X'X)^{-1}R'\right\}^{-1}$. In the same logic, $R'\left\{R(X'X)^{-1}R'\right\}^{-1}R$ is non-negative definite, and finally $(X'X)^{-1}R'\left\{R(X'X)^{-1}R'\right\}^{-1}R(X'X)^{-1}$ is non-negative definite.

3.2.2 Loss-of-Fit Statistics

To develop a test based on the restricted least squares estimator, we consider a single coefficient first and then turn to the general case of J linear restrictions. Consider the change in the fit of a multiple regression when a variable z is added to a model that already contains $K-1$ variables, $\mathbf{x}$. We showed in before that the effect on the fit would be given by

$$R_{Xz}^2 = R_X^2 + (1-R_X^2)r_{yz}^{*2}$$

where R_{Xz}^2 is the new R^2 after z is added, R_X^2 is the original R^2 and r_{yz}^* is the partial correlation between y and z, controlling for $\mathbf{x}$. So as we knew, the fit improves. In deriving the partial correlation coefficient between y and z in preceding sections, we obtained the convenient result

$$r_{yz}^{*2} = \frac{t_z^2}{t_z^2 + (n-K)}$$

where t_z is the square of the t ratio for testing the hypothesis that the coefficient on z is zero in the *multiple* regression of y on X and z. Based on the two equations, we have

$$t_z^2 = \frac{\left(R_{Xz}^2 - R_x^2\right)/1}{\left(1-R_{Xz}^2\right)/(n-K)}$$

We have developed in Wald test that in testing for a single restriction,

$$F\left[1, n-K\right] = t^2\left[n-K\right]$$

We see that the squared t statistic (i.e., the F statistic) is computed using the change in the R^2. By interpreting the preceding as the result of removing z from the regression, we see that we have proved a result for the case of testing whether a single slope is zero. By this construction, we see that for a single restriction, F is a measure of the loss of fit that results from imposing that restriction. Next, we will proceed to the general case of J linear restrictions, which will include one restriction as a special case.

The fit of the restricted least squares coefficients cannot be better than that of the unrestricted solution.

Change in Sum of Squares with Different Coefficient Vectors Suppose that b is the least squares coefficient vector in the regression of y on X, and that c is any other $K \times 1$ vector. The difference in the two sums of squared residual is

$$(y - Xc)' (y - Xc) - (y - Xb)' (y - Xb) = (c - b)' X'X (c - b)$$

As we can see directly, this difference is positive.

Let e_0 equal $y - Xb_0$. Then, using a familiar device,

$$\begin{aligned} e_0 &= y - Xb_0 = y - Xb - X(b_0 - b) \\ &= e - X(b_0 - b) \end{aligned}$$

The new sum of squared deviation is (recall the result that $X'e = 0$)

$$e_0'e_0 = e'e + (b_0 - b)'X'X(b_0 - b) \geq e'e$$

The loss of fit is

$$\begin{aligned} e_0'e_0 - e'e &= (b_0 - b)'X'X(b_0 - b) \\ &= \left((X'X)^{-1}R' \left\{R(X'X)^{-1}R'\right\}^{-1} (Rb - q)\right)' X'X \left((X'X)^{-1}R' \left\{R(X'X)^{-1}R'\right\}^{-1} (Rb - q)\right) \\ &= (Rb - q)' \left\{R(X'X)^{-1}R'\right\}^{-1} R(X'X)^{-1}X'X(X'X)^{-1}R' \left\{R(X'X)^{-1}R'\right\}^{-1} (Rb - q) \\ &= (Rb - q)' \left\{R(X'X)^{-1}R'\right\}^{-1} (Rb - q) \end{aligned}$$

Recall that the expression, $(Rb - q)' \left\{R(X'X)^{-1}R'\right\}^{-1} (Rb - q)$, appears in the numerator of F statistic. Inserting the remaining parts, we obtain

$$F[J, n - K] = \frac{(e_0'e_0 - e'e)/J}{e'e/(n - K)}$$

Finally, by dividing both numerator and denominator of F by $y'M_0y$, we obtain the general result involving R^2:

$$F[J, n - K] = \frac{(R^2 - R_0^2)/J}{(1 - R^2)/(n - K)}$$

This form has some intuitive appeal in that the difference in the fits of the two models is directly incorporated in the test statistic. As an example of this approach, consider the joint test that all the slopes in the model are zero. In this case, $R_0^2 = 0$, the F statistic is then $F[K, n - K] = \dfrac{R^2/K}{(1 - R^2)/(n - K)}$. For imposing a set of exclusion restrictions such as $\beta_k = 0$ for one or more coefficients, the obvious approach is simply to omit the variables from the regression and base the test on the sums of squared residuals for the restricted and unrestricted regressions.

Remarks: The R^2-form of F statistic is based on regressions for the same dependent variable. If the set of restriction requires an equivalent restricted regression with a different dependent variable, then R^2-form is invalid.

Analytically, if we are to test the hypothesis that a subset of coefficients, say β_2, are all zero, which is constructed using $R = [\mathbf{0}; \mathbf{I}]\,, q = \mathbf{0}$, and number of restrictions J is equal to K_2, the number of elements in β_2, then the matrix $R(X'X)^{-1}R'$ is the $K_2 \times K_2$ lower right block of the full inverse matrix. From partitioned inverse formula,

$$R(X'X)^{-1}R' = (X_2'M_1X_2)^{-1}$$

And clearly, $Rb - q = [\mathbf{0}; \mathbf{I}] \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} - \mathbf{0} = b_2$. Jointly,

$$\begin{aligned} e_0'e_0 - e'e &= (Rb - q)' \left\{ R(X'X)^{-1}R' \right\}^{-1} (Rb - q) \\ &= b_2'X_2'M_1X_2b_2 \end{aligned}$$

(Notice that the result has in fact been proved from another perspective in section of goodness of fit.)

The procedure for computing the appropriate F statistic amounts simply to comparing the sums of squared deviations from the “short” and “long” regressions, which we saw earlier.