Steven M Bergantino
14.382
Handout #5

OLS: HYPOTHESIS TESTING

Assumptions

(A0) $Y = X\beta + \varepsilon$

(A1) $\text{rank}(X) = k$

(A2') $\varepsilon \sim N(0, \sigma^2 I) \mid X$

(1) TESTING LINEAR RESTRICTIONS

We want to test linear restrictions on β of the form:

$$H_0: R\beta = C \qquad \text{vs.} \qquad H_1: R\beta \neq C$$

where R is a $J \times k$ matrix w/ $J \leq k$ and $\text{rank}(R) = J$. Each row of R represents a linear restriction on the parameter vector β. To see this observe that the first row of the equality in H_0 can be written as:

$$r_1'\beta = r_{11}\beta_1 + r_{12}\beta_2 + \dots + r_{1k}\beta_k = c_1$$

Thus, we require the linear combination $r_{11}\beta_1 + \dots + r_{1k}\beta_k$ of the elements of β to sum to c_1. Some examples are

- $r_1' = (1, 0, \dots, 0)$, $c_1 = \bar{\beta}_1 \Rightarrow \beta_1 = \bar{\beta}_1$
- $r_1' = (1, -1, 0, \dots, 0)$, $c_1 = 0 \Rightarrow \beta_1 - \beta_2 = 0$ or $\beta_1 = \beta_2$
- $r_1' = (1, 1, 0, \dots, 0)$, $c_1 = 1 \Rightarrow \beta_1 + \beta_2 = 1$

We will use $\hat{\beta}$ and its derived distribution to test these restrictions. Recall from (T4) of HO4 that under assumptions (A0) - (A2')

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1}) \Rightarrow R\hat{\beta} \sim N(R\beta, \sigma^2 R(X'X)^{-1}R')$$

$$(R\hat{\beta} - R\beta)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - R\beta) \sim \chi^2_J$$

This last result will generate the various test statistics we will use to test H_0. Before we proceed, however, let us note that the variance matrix of $R\hat{\beta}$ is invertible iff R has full row rank J (i.e. there are no redundant restrictions). To see this, observe that:

$$\text{rank}(R(X'X)^{-1}R') = \text{rank}([R(X'X)^{-1/2}][R(X'X)^{-1/2}]') = \text{rank}(R(X'X)^{-1/2}) = \text{rank}(R)$$

where the first equality follows from the fact that $(X'X)$ symmetric, PD $\Rightarrow (X'X)^{-1/2}$ symmetric, PD exists; the second equality from (T0) HO1 ($\text{rank}(AA') = \text{rank}(A)$); and the third equality also from (T0) HO1 ($\text{rank}(AB) = \text{rank}(A)$ for B square, nonsingular). Since $R(X'X)^{-1}R'$ is $J \times J$, it is invertible iff $\text{rank}(R) = J$, that is, iff there are no redundant restrictions. If there were, we could just remove the redundant rows from R and proceed. We now present three <u>numerically</u> equivalent test-statistics that can be used to do an F-test of H_0 vs. H_a.

(T1) <u>Theorem</u>: (Wald Statistic): The statistic

$$\frac{W}{J} \equiv \frac{(R\hat{\beta}-c)'[s^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta}-c)}{J} \sim F(J, n-k) \quad \text{under } H_0$$

<u>Proof</u>:

$$\frac{W}{J} = \frac{(R\hat{\beta}-c)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta}-c)/J}{(n-k)s^2/(n-k)\sigma^2} \overset{\text{under } H_0}{=} \frac{\chi^2_J/J}{\chi^2_{n-k}/(n-k)} \sim F(J, n-k)$$

Since, by (T4) HO 4, the denominator is a $\chi^2_{n-k}/(n-k)$ which, by the independence of $\hat{\beta}$ and s^2 ((T4) HO4) is independent of the numerator, and this is a χ^2_J/J under H_0 since $c = R\beta$. //

(T2) <u>Theorem</u>: (F-Statistic): Let $\hat{\varepsilon}_u = Y - X\hat{\beta}$ be the <u>unrestricted</u> vector of residuals from the OLS regression of Y on X and let $\hat{\varepsilon}_R = Y - X\hat{\beta}_R$ be the <u>restricted</u> vector of residuals from the restricted regression of Y on X: $\min_{\beta} (Y - X\beta)'(Y - X\beta)$ s.t. $R\beta = c$ ($\hat{\beta}_R$ is the solution to the restricted minimization problem and so is refered to as the restricted OLS estimator. Then the statistic

$$F = \frac{(\hat{\varepsilon}_R'\hat{\varepsilon}_R - \hat{\varepsilon}_u'\hat{\varepsilon}_u)/J}{\hat{\varepsilon}_u'\hat{\varepsilon}_u/(n-k)} = \frac{W}{J}$$

<u>Proof:</u>
Since the denominator is $\hat{\varepsilon}_u'\hat{\varepsilon}_u/(n-k) = S^2$, we need only show that

$$\hat{\varepsilon}_R'\hat{\varepsilon}_R - \hat{\varepsilon}_u'\hat{\varepsilon}_u = (R\hat{\beta}-c)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-c)$$

To do this observe that:

$$\hat{\varepsilon}_R = Y - X\hat{\beta}_R = Y - X\hat{\beta} + X(\hat{\beta}-\hat{\beta}_R) = \hat{\varepsilon}_u + X(\hat{\beta}-\hat{\beta}_R)$$

Since $\hat{\varepsilon}_u'X = 0$ the two parts of the last sum are orthogonal and so we have

$$\hat{\varepsilon}_R'\hat{\varepsilon}_R = \hat{\varepsilon}_u'\hat{\varepsilon}_u + (\hat{\beta}-\hat{\beta}_R)'(X'X)(\hat{\beta}-\hat{\beta}_R)$$

Now suppose that $\hat{\beta}_R = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-c)$. Then we would have

$$\hat{\varepsilon}_R'\hat{\varepsilon}_R - \hat{\varepsilon}_u'\hat{\varepsilon}_u = (\hat{\beta}-\hat{\beta}_R)'(X'X)(\hat{\beta}-\hat{\beta}_R)$$

$$= (R\hat{\beta}-c)'[R(X'X)^{-1}R']^{-1}\underbrace{R(X'X)^{-1}\overbrace{(X'X)(X'X)^{-1}}^{I}R'[R(X'X)^{-1}R']^{-1}}_{I}(R\hat{\beta}-c)$$

$$= (R\hat{\beta}-c)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-c) \;//$$

We will now solve for $\hat{\beta}_R$ to show that it satisfies the above relation. The lagrangian for the restricted estimator is

$$\mathcal{L}(\beta) = (Y-X\beta)'(Y-X\beta) + 2\lambda'(R\beta-c)$$

<u>FOC's:</u>

(β): $-2[X'(Y-X\beta) - R'\lambda] = 0 \Rightarrow \hat{\beta}_R = \underbrace{(X'X)^{-1}X'Y}_{\hat{\beta}} - (X'X)^{-1}R'\lambda$

(λ): $2(R\beta-c) = 0$

Substituting (δ) into (λ) gives

$$c = R\hat{\beta}_R = R(\hat{\beta} - (X'X)^{-1}R'\lambda) \Rightarrow \lambda = [R(X'X)^{-1}R']^{-1}(R\hat{\beta} - c)$$

$$\Rightarrow \hat{\beta}_R = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - c) \,//$$

Finally, we have the following result:

(C2) Corollary: (R^2 version of F-statistic) The statistic F from theorem (T2) can be written as (dividing numerator and denominator by $Y'Q_0Y$):

$$F = \frac{(R_u^2 - R_R^2)/J}{(1 - R_u^2)/(n-k)}$$

where $R_u^2 = 1 - \hat{\varepsilon}_u'\hat{\varepsilon}_u / Y'Q_0Y$ and $R_R^2 = 1 - \hat{\varepsilon}_R'\hat{\varepsilon}_R / Y'Q_0Y$ are the (centered) R^2 from the unrestricted and restricted regressions, respectively.

Note: In practice, running the restricted regression on a computer may involve a transformation of the LHS variable from Y to some $\tilde{Y} \neq Y$. In such cases, most regression software will report the restricted R^2 as $1 - \hat{\varepsilon}_R'\hat{\varepsilon}_R / \tilde{Y}'Q_0\tilde{Y}$ which is not the correct R_R^2 to plug into the above formula. The formulas in (T1) and (T2), however, are unaffected by such transformations.

We now give some examples of tests of linear restrictions.

(E1) Ex: Testing the Significance of the Regression

Suppose X includes a column of ones so that we can write the regression model as:

$$Y = \alpha i + \tilde{X}\delta + \varepsilon$$

where $X = [i \ \tilde{X}]$ and $\beta' = [\alpha \ \delta']$. The standard output of most regression packages includes a test of the hypothesis

$$H_0: \delta = 0 \quad \text{vs.} \quad H_1: \delta \neq 0$$

where $\delta \neq 0$ means that at least one of the slope terms is nonzero

Consider the R^2 form of the associated F-test. We have

Unrestricted Regression $Y = \alpha i + \tilde{X}\delta + \varepsilon \Rightarrow R_u^2$
Restricted Regression: $Y = \alpha i + \varepsilon \Rightarrow R_R^2 = 0$

where $R_R^2 = 0$ since $\hat{\alpha}_R = \bar{Y} \Rightarrow \hat{\varepsilon}_R = Y - \bar{Y}i = Q_0 Y \Rightarrow R_R^2 = 1 - \hat{\varepsilon}_R' \hat{\varepsilon}_R / Y'Q_0Y = 1 - Y'Q_0Y / Y'Q_0Y = 1 - 1 = 0$. Thus, the F-statistic for H_0 is:

$$F = \frac{R_u^2/(k-1)}{(1-R_u^2)/(n-k)} \sim F(k-1, n-k) \text{ under } H_0$$

where $k-1$ is the number of elements in δ. (We have imposed $k-1$ restrictions by requiring each element of δ to be 0 under H_0).

<u>Note:</u> It is possible for F to be quite large and to reject the <u>joint</u> null that all th slope coefficients are zero while at the same time accepting the null $H_0: \delta_k = 0$ vs. $H_1: \delta_k \neq 0$ in <u>separate</u> tests of each element of δ (eg t-tests). This may happen, for example, when there is a severe multicollinearity problem, making it difficult to get precise estimates of the individual δ_k's but not preventing us from seeing that the X_k's as a group explain a significant portion of the variation in y.

(E2) <u>Ex:</u> Testing a Single Linear Restriction

$$H_0: r'\beta = c \qquad \text{vs.} \qquad H_1: r'\beta \neq c$$

Looking at the Wald Statistic version of the test we have

$$W = \frac{(r'\hat{\beta} - c)^2}{S^2 r'(X'X)^{-1} r} \sim F(1, n-k) \text{ under } H_0 \Rightarrow \sqrt{W} = \frac{r'\hat{\beta} - c}{\sqrt{S^2 r'(X'X)^{-1} r}} \sim t_{n-k} \text{ under } H_0$$

Thus, we could also do a t-test of the restriction. Further, the t-test allows us to do one-sided tests. Notice that the F-tests are inherently two-sided tests as are all the joint tests you will encounter in this course.

(E3) <u>Ex:</u> T-test on a Single Parameter

If r has a 1 in the kth spot and 0's everywhere else and if $C = \bar{\beta}_k$ then our test reduces to a t-test of the null $\beta_k = \bar{\beta}_k$, and the test statistic is:

$$T = \frac{\hat{\beta}_k - \bar{\beta}_k}{SE(\hat{\beta}_k)} \sim t_{n-k} \quad \text{under } H_0$$

$\Rightarrow T^2 = W = F \sim F(1, n-k)$ (Note that T^2, like $|T|$, can only be used for a two-sided test).

(E4) <u>Ex:</u> Tests of Structural Change (Chow Tests)

Consider dividing the sample into two (or more) pieces and testing whether or not the coefficients on some subset of the columns in X differ across subsamples. Specifically, let $X = [W \ Z]$ and $\beta' = (\delta' \ \gamma')$ and consider a test of whether or not δ differs over subsamples 1 and 2, say. Then, if we let D_i be a dummy variable for sample i and let WD_i denote the matrix W, w/ all values in sample $j \neq i$ set to zero, then we can write the unrestricted model as:

$$\left.\begin{array}{lr} \text{Unrestricted Model:} & Y = (WD_1)\delta_1 + (WD_2)\delta_2 + Z\gamma + \varepsilon \\ & \text{or} \\ \text{Since } D_1 = \iota - D_2 & Y = W\delta_1 + (WD_2)(\delta_2 - \delta_1) + Z\gamma + \varepsilon \\ & \text{or} \\ \quad '' & Y = (WD_1)(\delta_1 - \delta_2) + W\delta_2 + Z\gamma + \varepsilon \end{array}\right\} \begin{array}{c}\hat{\varepsilon}_u \\ \text{and} \\ R^2_u\end{array}$$

All three regression specifications will yield the same residuals and R^2 for use in constructing the appropriate F-statistic but the first specification yields $\hat{\delta}_1$ and $\hat{\delta}_2$ w/o any further computation. The restricted model is

Restricted Model: $Y = W\delta + Z\gamma + \varepsilon \Rightarrow \hat{\varepsilon}_R$ and R^2_R

(T2) and (C2) then provide for simple tests of:

$$H_0: \delta_1 = \delta_2 \quad \text{vs} \quad H_1: \delta_1 \neq \delta_2$$

(E5) <u>Ex:</u> Some Common Null Hypotheses:

Unrestricted Model: $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t$

(a) H_0: $\beta_1 = \beta_2$ (1 restriction)
Restricted Model: $Y_t = \beta_0 + \beta_1(X_{1t} + X_{2t}) + \beta_3 X_{3t} + \varepsilon_t$

(b) H_0. $\beta_1 = \beta_2 = 0$ (2 restrictions)
Restricted Model: $Y_t = \beta_0 + \beta_3 X_{3t} + \varepsilon_t$

(c) H_0: $\beta_1 + \beta_2 = 1$ (1 restriction)
Restricted Model: $(Y_t - X_{2t}) = \beta_0 + \beta_1(X_{1t} - X_{2t}) + \beta_3 X_{3t} + \varepsilon_t$
or $\rightarrow$ $(Y_t - X_{1t}) = \beta_0 + \beta_2(X_{2t} - X_{1t}) + \beta_3 X_{3t} \quad \varepsilon_t$

Notice that the LHS variable has changed so that the standard regression output for these equations will give the wrong R^2_R for use in the R^2 version of the F-test.

<u>Note:</u> All of the F-tests discussed above are of the form, reject H_0 if $F \geq C_\alpha$ where C_α is chosen so that $P(F \geq C_\alpha) = \alpha$ for a random variable $F \sim F(J, n-k)$. The logic behind this follows directly from the form of the test statistics. For the Wald Statistic the logic is that in large samples $\hat{\beta} \simeq \beta$ so $R\hat{\beta} \simeq R\beta$. Thus, if H_0 is true, $R\hat{\beta} - c \simeq 0$ and hence $W/J \simeq 0$. If H_0 is false, however, then $R\hat{\beta} - c \neq 0$ and so $W/J > 0$ since W/J can be written as the inner product of a $J \times 1$ vector with itself. For the residual and R^2 tests, the logic is even more transparent. The restricted regression will always entail a loss of fit (ie $\hat{\varepsilon}_R'\hat{\varepsilon}_R \geq \hat{\varepsilon}_U'\hat{\varepsilon}_U$ and $R^2_R \leq R^2_U$) so $F \geq 0$. But if H_0 is true, this loss of fit will be neglible and so $F \simeq 0$.

<u>Note:</u> Since nonlinear functions of normal random variables are not normal, we cannot easily extend our results to tests of <u>nonlinear restrictions</u> on β unless the sample size is large. For once we establish the asymptotic normality of $\hat{\beta}$ (ie $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$ for some V), we will be able to use the Delta Method ((T16) H03) to conduct Chi-Squared tests of nonlinear restrictions on β that

will be "approximately" valid in large samples.

(2) TESTING SEQUENCES OF NESTED HYPOTHESES

Sometimes we may want to test a set of restrictions on β and conditional on these restrictions being accepted, test further restrictions on β. For example, given the model

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

we may wish to test whether or not $\beta_4 = 0$ and if this hypothesis is accepted then proceed to test if $\beta_3 = 0$ as well. More generally, we may wish to conduct tests on a sequence of nested hypotheses:

$H_0^1: R_1\beta = c_1$ vs $H_1^1: R_1\beta \neq c_1$

$H_0^2: R_1\beta = c_1$ and $R_2\beta = c_2$ vs. $H_1^2: R_1\beta = c_1$ but $R_2\beta \neq c_2$

where R_1 is $J_1 \times k$ and R_2 is $J_2 \times k$ with $\text{rank}(R_1) = J_1$, $\text{rank}(R_2) = J_2$ and $\text{rank}(R) = J_1 + J_2 \leq k$ where R is formed by stacking R_1 on top of R_2. Thus, (R_1, c_1) contain J_1 linear restrictions on β while (R_2, c_2) contain J_2 additional (ie non-redundant) and mutually compatible (w/ (R_1, c_1)) linear restrictions. We already know how to conduct the individual tests and compute their size. The question is: How do we test the sequence of restrictions and what is the overall size of the test? The answer is, as we would expect, we conduct the least restrictive test first and, conditional on acceptance of the null hypothesis of this test, proceed to the more restrictive test. Specifically, we have:

(T3) Theorem: Let $(F_i, c_{\alpha i}, \alpha_i)$ be the F-statistic, critical value and size of the test H_0^i vs H_1^i $i = 1, 2$. Then F_1 and F_2 are independent and the composite test of

$$H_0^2 \quad \text{vs} \quad H_1^1 \text{ or } H_1^2$$

given by

(i) Accept H_1^1 if $F_1 \geq C_{\alpha_1}$
(ii) Accept H_1^2 if $F_1 < C_{\alpha_1}$ but $F_2 \geq C_{\alpha_2}$
(iii) Accept H_0^2 if $F_1 < C_{\alpha_1}$ and $F_2 < C_{\alpha_2}$

has size $\alpha_1 + \alpha_2 - \alpha_1\alpha_2$

Proof:
We won't prove independence (ask Whitney) but the size of the composite test is:

$$\text{Size} = \mathbb{P}(\text{Reject } H_0^2 \mid H_0^2) = \mathbb{P}(F_1 \geq C_{\alpha_1} \mid H_0^2) + \mathbb{P}(F_1 < C_{\alpha_1}, F_2 \geq C_{\alpha_2} \mid H_0^2)$$

by independence

$$= \mathbb{P}(F_1 \geq C_{\alpha_1} \mid H_0^2) + \mathbb{P}(F_1 < C_{\alpha_1} \mid H_0^2)\mathbb{P}(F_2 \geq C_{\alpha_2} \mid H_0^2)$$

$$= \alpha_1 + (1-\alpha_1)\alpha_2 = \alpha_1 + \alpha_2 - \alpha_1\alpha_2$$

since $F_1 \sim F(J_1, n-k)$ under H_0^1 and H_0^2 //

(3) CHOOSING BETWEEN NONNESTED MODELS

So far we have discussed tests of linear restrictions on the model $Y = X\beta + \varepsilon$. Our tests have been of the form

$$H_0: R\beta = c \qquad \text{vs} \qquad H_A: R\beta \neq c$$

Equivalently, we could think of them as tests of competing models

$$H_0: \tilde{Y} = \tilde{X}\underset{J-K \times 1}{\tilde{\beta}} + \varepsilon \qquad \text{vs.} \qquad H_A: Y = X\underset{K \times 1}{\beta} + \varepsilon$$

where the model $\tilde{Y} = \tilde{X}\tilde{\beta} + \varepsilon$ is nested inside the model $Y = X\beta + \varepsilon$ in that it embodies a set of J linear restrictions on β. To see this more clearly, observe that, by reshuffling the columns of R and corresponding rows of β, we can write $R = [R_1\ R_2]$ where R_1 is a $J \times J$ invertible matrix and R_2 is $J \times (K-J)$, and can write $\beta' = (\beta_1'\ \beta_2')$. Thus, the restrictions $R\beta = c$ can be reformulated as:

$$C = [R_1 \ R_2]\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = R_1\beta_1 + R_2\beta_2 \Rightarrow \underset{J\times 1}{\beta_1} = R_1^{-1}C - R_1^{-1}R_2\underset{(K-J)\times 1}{\beta_2}$$

Imposing these restrictions on the model we have

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon = X_1(R_1^{-1}C - R_1^{-1}R_2\beta_2) + X_2\beta_2 + \varepsilon$$

$$\Rightarrow \quad \underbrace{(Y - X_1R_1^{-1}C)}_{\tilde{Y}} = \underbrace{(X_2 - X_1R_1^{-1}R_2)}_{\tilde{X}}\underbrace{\beta_2}_{\tilde{\beta}} + \varepsilon$$

Note: $\tilde{Y} = Y \Leftrightarrow C = 0$ Thus, the R^2 version of the F-test will be useful only when $C=0$ (eg for H_0: $\beta_1 = \beta_2$ but not for H_0. $\beta_1 + \beta_2 = 1$).

Suppose, however, that we wish to test nonnested models against each other. Specifically, suppose we wish to conduct the test of

$$H_0: Y = X\beta + \varepsilon \quad \text{vs.} \quad H_1: Y = Z\gamma + \nu$$

where X and Z may contain common variables but neither is contained inside the other. Following are some tests for such cases. They are all structured around the ecompassing principle, nesting both models inside an encompassing model.

Davidson-MacKinnon (Estimation and Inference in Econometrics, 1993)

This test is based on the augmented regression:

$$Y = X\beta + Z_1\gamma_1 + \eta$$

where Z_1 contains the columns of Z that are not also in X. The test is carried out by doing the F-test of $\gamma_1 = 0$ vs. $\gamma_1 \neq 0$ from the augmented regression.

J-Test (Davidson-MacKinnon EMA 1981)

This test is based on the encompassing model

$$Y = (1-\alpha) X\beta + \alpha Z\gamma + \varepsilon$$

If H_0 is true then $\alpha = 0$. The test is implemented as follows.

(1) Regress Y on $Z \Rightarrow \hat{\gamma}$
(2) Regress Y on X and $Z\hat{\gamma} \Rightarrow \hat{\alpha}$
(3) Compute the usual t-statistic $T = \hat{\alpha}/SE(\hat{\alpha})$ from the regression in (2). $T \xrightarrow{d} N(0,1)$ under H_0 and can be used to test H_0 vs. H_A.

The problem w/ this test is that it is possible to reject both, neither, or either one of the two hypotheses. Asymptotically, however, this is not a problem since if H_A is true, then $\mathbb{P}\{|T| \geq Z_{.025}\} \longrightarrow 1$ as $n \to \infty$.