Exact Tests in the Classical Normal Linear Model

Consider the classical normal linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2\mathbf{I}), \tag{1}$$

where $\boldsymbol{X}$ is $N \times k$, and $\boldsymbol{u}$ is statistically independent of $\boldsymbol{X}$.

In other words, all of the regressors in $\boldsymbol{X}$ are exogenous.

To test a single restriction, partition $\boldsymbol{\beta}$ as $[\boldsymbol{\beta}_1 \vdots \beta_2]$, where $\boldsymbol{\beta}_1$ is a $(k-1)$-vector and β_2 is a scalar.

When $\boldsymbol{X}$ is partitioned conformably with $\boldsymbol{\beta}$, (1) can be rewritten as

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \beta_2\boldsymbol{x}_2 + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2\mathbf{I}), \tag{2}$$

where $\boldsymbol{X}_1$ is $N \times (k-1)$ and $\boldsymbol{x}_2$ is an N-vector, with $\boldsymbol{X} = [\boldsymbol{X}_1 \;\; \boldsymbol{x}_2]$.

By the FWL Theorem, the OLS estimate of β_2 from (2) is the same as the OLS estimate from the FWL regression

$$\boldsymbol{M}_1\boldsymbol{y} = \beta_2\boldsymbol{M}_1\boldsymbol{x}_2 + \text{residuals}, \tag{3}$$

where $\boldsymbol{M}_1 \equiv \mathbf{I} - \boldsymbol{X}_1(\boldsymbol{X}_1^\top\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top$ is the matrix that projects on to $\mathcal{S}^\perp(\boldsymbol{X}_1)$. Here $\boldsymbol{M}_1$ is short for $\boldsymbol{M}_{X_1}$.

To test the hypothesis that $\beta_2 = \beta_2^0$, we have to subtract β_2^0 from $\hat\beta_2$ and divide by the square root of the variance, where

$$\hat\beta_2 = \frac{\boldsymbol{x}_2^\top\boldsymbol{M}_1\boldsymbol{y}}{\boldsymbol{x}_2^\top\boldsymbol{M}_1\boldsymbol{x}_2} \quad \text{and} \quad \operatorname{Var}(\hat\beta_2) = \sigma^2(\boldsymbol{x}_2^\top\boldsymbol{M}_1\boldsymbol{x}_2)^{-1}. \tag{4}$$

Of course, the variance here depends on σ^2, which is unknown. In practice, we will need to replace it by s^2.

But let us assume, for a little while, that we know σ^2. This will yield a test statistic that is usually infeasible.

For a test of $\beta_2 = 0$, the infeasible test statistic is

$$z_{\beta_2} \equiv \frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{\sigma(\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2)^{1/2}}. \tag{5}$$

If the data are actually generated by the model (2) with $\beta_2 = 0$, then

$$\boldsymbol{M}_1 \boldsymbol{y} = \boldsymbol{M}_1(\boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{u}) = \boldsymbol{M}_1 \boldsymbol{u}. \tag{6}$$

Therefore, the right-hand side of equation (5) becomes

$$\frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{u}}{\sigma(\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2)^{1/2}}. \tag{7}$$

We want to show that $z_{\beta_2} \sim \mathrm{N}(0, 1)$. This requires that the numerator of (7) be normally distributed with variance equal to the square of the denominator.

The numerator is just a linear combination of the components of $\boldsymbol{u}$, which is multivariate normal, so z_{β_2} must be normally distributed.

The variance of the numerator of (7) is

$$\mathrm{E}(\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{M}_1 \boldsymbol{x}_2) = \boldsymbol{x}_2^\top \boldsymbol{M}_1 \,\mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top)\boldsymbol{M}_1 \boldsymbol{x}_2 \tag{8}$$

$$= \boldsymbol{x}_2^\top \boldsymbol{M}_1 \sigma^2 \mathbf{I} \boldsymbol{M}_1 \boldsymbol{x}_2 = \sigma^2 \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2. \tag{9}$$

Since the denominator of (7) is just the square root of the variance of the numerator, $z_{\beta_2} \sim \mathrm{N}(0, 1)$ under the null hypothesis.

- In practice, of course, we very rarely know σ^2.
- We need to replace σ in (5) by s, the standard error of (2).
- Recall that $s^2 = \boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}/(N-k) = \mathrm{SSR}/(N-k)$.
- Because s^2 is random and not equal to σ^2, the ***t* statistic** does not follow the $\mathrm{N}(0, 1)$ distribution in finite samples.
- Instead, it follows the $t(N-k)$ distribution.

$$t_{\beta_2} \equiv \frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{s(\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2)^{1/2}} = \left(\frac{\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}}{N-k}\right)^{-1/2} \frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{(\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2)^{1/2}}. \tag{10}$$

If a test statistic has the $t(N-k)$ distribution, we can write it as the ratio of a standard normal variable z to the square root of $\zeta/(N-k)$, where ζ is independent of z and distributed as $\chi^2(N-k)$.

The t statistic (10) can be rewritten as

$$t_{\beta_2} = (\sigma/s) z_{\beta_2} = \frac{z_{\beta_2}}{\left(\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}/((N-k)\sigma^2)\right)^{1/2}}. \tag{11}$$

We have already shown that $z_{\beta_2} \sim \mathrm{N}(0,1)$. It remains to show that $\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}/\sigma^2 \sim \chi^2(N-k)$ and that the numerator and denominator of (11) are independent.

Under any DGP that belongs to (2),

$$\frac{\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}}{\sigma^2} = \frac{\boldsymbol{u}^\top \boldsymbol{M}_X \boldsymbol{u}}{\sigma^2} = \boldsymbol{\epsilon}^\top \boldsymbol{M}_X \boldsymbol{\epsilon}, \quad \text{where} \quad \boldsymbol{\epsilon} \equiv \boldsymbol{u}/\sigma \sim \mathrm{N}(\mathbf{0}, \mathbf{I}). \tag{12}$$

Since $\boldsymbol{M_X}$ is a projection matrix with rank $N-k$, $\boldsymbol{\epsilon}^\top \boldsymbol{M_X} \boldsymbol{\epsilon}$ in (12) is distributed as $\chi^2(N-k)$ by part 2 of Theorem 4.1.

Note that $\boldsymbol{\epsilon}^\top \boldsymbol{M_X} \boldsymbol{\epsilon}$ depends on $\boldsymbol{y}$ only through $\boldsymbol{M_X y}$.

z_{β_2} depends on $\boldsymbol{y}$ only through $\boldsymbol{P_X y}$, since

$$\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y} = \boldsymbol{x}_2^\top \boldsymbol{P_X M}_1 \boldsymbol{y} = \boldsymbol{x}_2^\top (\boldsymbol{P_X} - \boldsymbol{P_X P}_1) \boldsymbol{y} = \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{P_X y}. \tag{13}$$

The first equality uses the fact that $\boldsymbol{x}_2 \in \mathcal{S}(\boldsymbol{X})$. The third equality uses the fact that $\boldsymbol{P_X P}_1 = \boldsymbol{P}_1 \boldsymbol{P_X}$.

We know that $\boldsymbol{M_X y} = \boldsymbol{M_X u}$ and $\boldsymbol{P_X y} = \boldsymbol{X\beta} + \boldsymbol{P_X u}$.

The $N \times N$ matrix of covariances of the components of $\boldsymbol{P_X u}$ and $\boldsymbol{M_X u}$ is

$$\mathrm{E}(\boldsymbol{P_X u u}^\top \boldsymbol{M_X}) = \sigma^2 \boldsymbol{P_X M_X} = \mathbf{O}, \tag{14}$$

because $\boldsymbol{P_X}$ and $\boldsymbol{M_X}$ are complementary projections.

- The vectors $\boldsymbol{P_X u}$ and $\boldsymbol{M_X u}$ have zero covariance because they lie in orthogonal subspaces, namely, the images of $\boldsymbol{P_X}$ and $\boldsymbol{M_X}$.
- Zero covariance implies that $\boldsymbol{P_X u}$ and $\boldsymbol{M_X u}$ are independent, since they are multivariate normal.
- Even though the numerator and denominator of (11) both depend on $\boldsymbol{y}$, they are independent.

Conclusion: The t statistic for $\beta_2 = 0$ in (2) follows the $t(N-k)$ distribution under the null hypothesis.

- One-tailed and two-tailed tests based on t_{β_2} are almost the same as ones based on z_{β_2}.
- We use the $t(N-k)$ distribution instead of the standard normal distribution to compute P values or critical values.
- Both critical values and P values based on $t(N-k)$ will be larger than ones based on $\mathrm{N}(0,1)$, because the randomness in s causes t_{β_2} to be more spread out than z_{β_2}.

Tests of Several Restrictions

Suppose there are r restrictions, with $r \leq k$, of the form $\boldsymbol{\beta}_2 = \mathbf{0}$. The alternative hypothesis is the model

$$\mathrm{H}_1: \quad \boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}). \tag{15}$$

Here $\boldsymbol{X}_1$ is $N \times k_1$, $\boldsymbol{X}_2$ is $N \times k_2$, $\boldsymbol{\beta}_1$ is a k_1-vector, $\boldsymbol{\beta}_2$ is a k_2-vector, $k = k_1 + k_2$, and the number of restrictions $r = k_2$.

The null hypothesis is the model

$$\mathrm{H}_0: \quad \boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}). \tag{16}$$

If $\mathrm{USSR} = \boldsymbol{y}^\top\boldsymbol{M}_X\boldsymbol{y}$, from (15), and $\mathrm{RSSR} = \boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{y}$, from (16), then the ***F* statistic**, which is distributed as $F(r, N-k)$, is

$$F_{\boldsymbol{\beta}_2} \equiv \frac{(\mathrm{RSSR} - \mathrm{USSR})/r}{\mathrm{USSR}/(N-k)}. \tag{17}$$

The USSR can be computed from the FWL regression

$$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \text{residuals}. \tag{18}$$

The TSS from this regression is $\boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{y}$, the ESS is $\boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}\boldsymbol{M}_1\boldsymbol{y}$, and so the SSR is

$$\text{USSR} = \boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{y} - \boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y}. \tag{19}$$

Therefore,

$$\text{RSSR} - \text{USSR} = \boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y}, \tag{20}$$

and the F statistic (17) can be written as

$$F_{\beta_2} = \frac{\boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y}/r}{\boldsymbol{y}^\top\boldsymbol{M}_X\boldsymbol{y}/(N-k)}. \tag{21}$$

In general, $\boldsymbol{M_X y} = \boldsymbol{M_X u}$. Under the null, $\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{u}$, and so

$$F_{\boldsymbol{\beta}_2} = \frac{\boldsymbol{\epsilon}^\top \boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top \boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top \boldsymbol{M}_1\boldsymbol{\epsilon}/r}{\boldsymbol{\epsilon}^\top \boldsymbol{M_X}\boldsymbol{\epsilon}/(N-k)}, \tag{22}$$

where, as before, $\boldsymbol{\epsilon} \equiv \boldsymbol{u}/\sigma$.

The denominator of (22) is $1/(N-k)$ times something that is distributed as $\chi^2(N-k)$.

The quadratic form in the numerator is $\boldsymbol{\epsilon}^\top \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}\boldsymbol{\epsilon}$. It must be distributed as $\chi^2(r)$ because $\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$ is a projection matrix with rank r.

The two χ^2 random variables are independent, because $\boldsymbol{M_X}$ and $\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$ project on to mutually orthogonal subspaces:

$$\boldsymbol{M_X}\boldsymbol{M}_1\boldsymbol{X}_2 = \boldsymbol{M_X}(\boldsymbol{X}_2 - \boldsymbol{P}_1\boldsymbol{X}_2) = \mathbf{O}. \tag{23}$$

Thus (22) is distributed as $F(r, N-k)$ under H_0.

F Tests and *t* Tests

When there is just one restriction, the F statistic (21) is equal to the square of the t statistic (10).

The numerator of (21) simplifies to

$$\boldsymbol{y}^\top \boldsymbol{M}_1 \boldsymbol{x}_2 (\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2)^{-1} \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y} = \frac{(\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y})^2}{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2}, \tag{24}$$

which is the square of the second factor in (10). The square root of the denominator of (21) is

$$\left(\frac{\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}}{N-k}\right)^{1/2}. \tag{25}$$

Combining the signed square root of (24) with (25), we get (10):

$$\sqrt{F_{\beta_2}} = \left(\frac{\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}}{N-k}\right)^{-1/2} \frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{(\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2)^{1/2}}. \tag{26}$$

Examples of the F Test

1. Testing Slope Coefficients in a Classical Normal Linear Model

The null hypothesis H_0 is that $\boldsymbol{\beta}_2 = \mathbf{0}$ in the model

$$\boldsymbol{y} = \beta_1 \boldsymbol{\iota} + \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{27}$$

where $\boldsymbol{\iota}$ is an N-vector of 1s and $\boldsymbol{X}_2$ is $N \times (k-1)$.

The test statistic (21) becomes

$$F_{\boldsymbol{\beta}_2} = \frac{\boldsymbol{y}^\top \boldsymbol{M}_\iota \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_\iota \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_\iota \boldsymbol{y} / (k-1)}{\big(\boldsymbol{y}^\top \boldsymbol{M}_\iota \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{M}_\iota \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_\iota \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_\iota \boldsymbol{y}\big) / (N-k)}. \tag{28}$$

The matrix expression in the numerator here is just the ESS from the FWL regression

$$\boldsymbol{M}_\iota \boldsymbol{y} = \boldsymbol{M}_\iota \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \text{residuals}. \tag{29}$$

The matrix expression in the denominator of (28) is the TSS from this regression, minus the ESS.

Since the centered R^2 from (27) is just the ratio of ESS to TSS,

$$F_{\boldsymbol{\beta}_2} = \frac{N-k}{k-1} \times \frac{R_c^2}{1-R_c^2}. \tag{30}$$

But you should never compute $F_{\boldsymbol{\beta}_2}$ in this way!

2. Testing the Equality of Two Parameter Vectors

We can often divide a sample into two, or possibly more than two, subsamples.

We can ask whether a linear regression model has the same coefficients for both subsamples. The test is often called a **Chow test**.

Suppose there are two subsamples, of lengths N_1 and N_2, with $N = N_1 + N_2$, and both N_1 and N_2 are greater than k. Examples.

We can write

$$\boldsymbol{y} \equiv \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{X} \equiv \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}, \tag{31}$$

where $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are an N_1-vector and an N_2-vector, while $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are $N_1 \times k$ and $N_2 \times k$ matrices.

We can put the subsamples together in the regression model

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \boldsymbol{\beta}_1 + \begin{bmatrix} \mathbf{O} \\ \boldsymbol{X}_2 \end{bmatrix} \gamma + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}). \tag{32}$$

In the first subsample, the regression functions are the components of $\boldsymbol{X}_1 \boldsymbol{\beta}_1$. In the second, they are the components of $\boldsymbol{X}_2(\boldsymbol{\beta}_1 + \gamma)$.

Thus γ is defined as $\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1$.

Define $\boldsymbol{Z}$ as an $N \times k$ matrix with $\mathbf{O}$ in its first N_1 rows and $\boldsymbol{X}_2$ in the remaining N_2 rows.

Then (32) can be rewritten as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_1 + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}). \tag{33}$$

This model has N observations and $2k$ regressors. The null hypothesis is now a set of k zero restrictions, that $\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = \boldsymbol{\gamma} = \mathbf{0}$.

We could run (33) to get the USSR, and then run the restricted model, which is just the regression of $\boldsymbol{y}$ on $\boldsymbol{X}$, to get the RSSR.

But USSR is just the sum of the two SSRs from the two subsample regressions, say SSR_1 and SSR_2.

If RSSR denotes the SSR from regressing $\boldsymbol{y}$ on $\boldsymbol{X}$, then

$$F_\gamma = \frac{(\mathrm{RSSR} - \mathrm{SSR}_1 - \mathrm{SSR}_2)/k}{(\mathrm{SSR}_1 + \mathrm{SSR}_2)/(N-2k)}. \tag{34}$$

This **Chow statistic** is distributed as $F(k, N-2k)$ under the null hypothesis that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.