

Hypothesis Testing

Table of contents

1	Overview	1
2	Linear Hypotheses	2
3	Distribution under H_0	2
4	Single linear hypothesis	3
4.1	Asymptotic distribution	6
5	Multiple Linear Hypotheses	6
5.1	Asymptotic distribution	8
6	Restricted OLS	8
6.1	Testing full model	10
7	Heteroskedasticity	11
	References	11

1 Overview

In this handout we will see how to test linear hypotheses. These could be,

- single hypothesis concerning single parameter;
- single hypothesis concerning multiple parameters;
- multiple hypothesis concerning more than one parameter.

Further reading can be found in:

- Section 7.1-7.4 of Cameron and Trivedi (2005)
- Section 6.2 & 6.3 of Verbeek (2017)

2 Linear Hypotheses

Assume that the underlying model is linear in parameters,

$$Y = X\beta + u$$

Consider the linear hypothesis,

$$R\beta = r$$

where R is an $q \times k$, non-random matrix and r a $q \times 1$ non-random vector of constants.

For example, consider the linear model from Problem Set 2,

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$$

we can express the single ($q = 1$) linear hypothesis $H_0 : \beta_2 = \beta_3 \Rightarrow \beta_2 - \beta_3 = 0$ as

$$H_0 : R\beta = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = 0$$

In the case of a single linear hypothesis, we can also use the notation $c'\beta = r$ where c is a $k \times 1$ non-random vector.

We can express the multiple ($q = 2$) linear hypotheses from this exercise

$$H_0 : \beta_2 = \beta_3 \quad \text{and} \quad \beta_4 + \beta_5 = 1$$

as,

$$R\beta = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

3 Distribution under H_0

Let us assume the CLRM assumptions, including normality and homoskedasticity

- **CLRM 5:** $u|X \sim N(0, \sigma^2 I_n)$

We know that under these assumptions the OLS estimator is normally distributed,

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

This implies that,¹

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

Under the null hypothesis, $H_0 : R\beta = r$,

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

From Handout 3, we know the asymptotic distribution of $\hat{\beta}$ is normal, even if the (conditional) finite sample distribution of u is not. Applying the Delta Method (see Handout 2) to this result,

$$\sqrt{n}(R\hat{\beta}_n - R\beta) \rightarrow_d N(0, \sigma^2 R(E[X_1 X_1'])^{-1}R') \quad \text{as } n \rightarrow \infty$$

Which means that, **under** H_0 ,

$$\sqrt{n}(R\hat{\beta}_n - r) \rightarrow_d N(0, \sigma^2 R(E[X_1 X_1'])^{-1}R') \quad \text{as } n \rightarrow \infty$$

4 Single linear hypothesis

Assuming CLRM 5, we know that for any $k \times 1$ non-random vector c ,

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

Which then implies,

$$\frac{c'\hat{\beta} - c'\beta}{\sqrt{\sigma^2 c'(X'X)^{-1}c}}|X \sim N(0, 1)$$

Under $H_0 : c'\beta = r$,

$$\frac{c'\hat{\beta} - r}{\sqrt{\sigma^2 c'(X'X)^{-1}c}}|X \stackrel{H_0}{\sim} N(0, 1)$$

Note, this statement is only true under H_0 , while the former is true always (under CLRM assumptions).

¹If $X \sim N(\mu, \Omega)$ then the linear transformation of X , $Y = AX + b$, is distributed $Y \sim N(A\mu + b, A\Omega A')$.

When evaluating against the alternative hypothesis, $H_1 : c'\beta \neq r$, the valid two-sided rejection rule is given by,

- Reject H_0 if $|Z\text{-stat}| > z_{1-\alpha/2}$

where α is the chosen significance level (e.g., 1%, 5%, or 10%) and $z_{1-\alpha/2}$ the $1-\alpha/2$ percentile of the standard normal distribution. We can write the two sided rejection in this way given the symmetry of the normal distribution. In general, the rejection rule is,

- Reject H_0 if $Z\text{-stat} < z_{\alpha/2}$ or $Z\text{-stat} > z_{1-\alpha/2}$

However, given the symmetric of $N(0,1)$, $z_{\alpha/2} = -z_{1-\alpha/2}$:

$$\begin{aligned}\alpha &= Pr(Z < z_{\alpha/2}) + Pr(Z > z_{1-\alpha/2}) \\ &= Pr(Z < -z_{1-\alpha/2}) + Pr(Z > z_{1-\alpha/2}) \\ &= Pr(|Z| > z_{1-\alpha/2})\end{aligned}$$

For asymmetric distributions, this is not the case, and you must write the two rejection cases separately.

This test is valid for the true σ^2 , the homoskedastic variance of the error term. We typically do not know this variance and need to replace it with the estimator,

$$s^2 = \frac{1}{n-k} \sum_{i=1}^n \hat{u}_i^2 = \frac{\hat{u}'\hat{u}}{n-k}$$

where \hat{u} is the residual from the linear model: $\hat{u} = Y - X\hat{\beta}$. This begs the question: what is the conditional distribution of,

$$T\text{-stat} = \frac{c'\hat{\beta} - r}{\sqrt{s^2 c'(X'X)^{-1}c}}$$

under H_0 . You will know from undergraduate texts that it is t -distributed. We can demonstrate this result using the idempotent properties of projection matrices.

First-off, substitute back in σ^2

$$T\text{-stat} = \frac{c'\hat{\beta} - r}{\sqrt{\sigma^2 c'(X'X)^{-1}c}} \times \sqrt{\frac{s^2}{\sigma^2}}$$

The T-statistic can be rewritten as the Z-statistic (which has a standard normal distribution under H_0) multiplied by the square-root of the ratio of s^2 and σ^2 . Consider the distribution of this ratio:

$$\frac{s^2}{\sigma^2} = \frac{1}{n-k} \frac{\hat{u}'\hat{u}}{\sigma^2}$$

We know that $\hat{u} = M_X Y = M_X u$, where M_X is idempotent and symmetric. Therefore,

$$\frac{s^2}{\sigma^2} = \frac{1}{n-k} \frac{u' M_X u}{\sigma^2} = \frac{1}{n-k} \left(\frac{u}{\sigma} \right)' M_X \left(\frac{u}{\sigma} \right)$$

By the properties of symmetric matrices (see Linear Algebra notes), M_X has a eigenvector decomposition

$$M_X = C \Lambda C' \quad \text{where} \quad C C' = C' C = I_n$$

Moreover, since M_X is idempotent, all eigenvalues are $\lambda_j \in \{0, 1\}$ for $j = 1, \dots, n$. This implies that,

$$\text{rank}(M_X) = \text{tr}(M_X) = \sum_{j=1}^n \lambda_j$$

We know that the $\text{rank}(M_X) = \text{rank}(I_n - P_X) = \text{rank}(I_n) - \text{rank}(P_X) = n - \text{rank}(X) = n - k$. Therefore, M_X has $n - k$ non-zero eigenvalues. Applying this eigenvector decomposition, we get that,

$$\frac{s^2}{\sigma^2} = \frac{1}{n-k} \left(\frac{u}{\sigma} \right)' C \Lambda C' \left(\frac{u}{\sigma} \right) = \frac{1}{n-k} \left(\frac{C' u}{\sigma} \right)' \Lambda \left(\frac{C' u}{\sigma} \right)$$

Since $u|X \sim N(0, \sigma^2 I_n)$,

$$\frac{C' u}{\sigma} \sim N(0, \underbrace{C' I_n C}_{=C' C = I_n})$$

Thus,

$$\frac{s^2}{\sigma^2} = \frac{1}{n-k} Z' \Lambda Z = \frac{1}{n-k} \sum_{j=1}^n \lambda_j Z_j^2$$

Under the H_0 , the T-statistic's distribution can be described by the ratio of two random variables: $Z \sim N(0, 1)$ and $W \sim \chi_{n-k}^2$.

$$\text{T-stat} = \frac{Z}{\sqrt{W/(n-k)}} \sim T_{n-k}$$

This distribution is known as the t -distribution, with $n - k$ degrees of freedom (dof). Note, the degrees of freedom correspond to the rank of the M_X matrix, since this determines the number of squared standard normal distributions summed in the denominator.

For a two-sided test, we reject H_0 when,

$$|\text{T-stat}| > t_{n-k, 1-\alpha/2}$$

where $t_{n-k, 1-\alpha}$ is the $1 - \alpha/2$ percentile of the t -distribution (with $n - k$ dof). As with the standard normal distribution, the t -distribution is symmetric.

4.1 Asymptotic distribution

We will see that for n large, it does not make a difference that we do not know σ^2 . The fact that $s^2 \rightarrow_p \sigma^2$ (as $n \rightarrow \infty$) means that the T-statistic will be normally distributed in the limit.

To demonstrate this result we just need to concern ourself with,

$$\frac{s^2}{\sigma^2} = \frac{1}{n-k} \sum_{j=1}^n \lambda_j Z_j^2 \sim \chi_{n-k}^2 / (n-k)$$

The $E[\sum_{j=1}^n \lambda_j Z_j^2] = n - k$, since $E[Z_1^2] = 1$ for any standard normal random variable. Moreover, $E[(\sum_{j=1}^n \lambda_j Z_j^2)^2] = 2(n - k)$. Therefore, by the WLLN,

$$\frac{1}{n-k} \sum_{j=1}^n \lambda_j Z_j^2 \rightarrow_p 1$$

As a result, **under** H_0 ,

$$\text{T-stat} \rightarrow_d N(0, 1)$$

5 Multiple Linear Hypotheses

We will follow similar steps to above. For any non-random $q \times k$ matrix R ,

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

Under the CLRM assumptions, this statement is true always, while **under** $H_0 : R\beta = r$,

$$R\hat{\beta}|X \stackrel{H_0}{\sim} N(r, \sigma^2 R(X'X)^{-1}R')$$

To construct a scalar test static for these q hypotheses we will first need to consider the distribution of,

$$(R\hat{\beta} - r)'(\sigma^2 R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) \sim \chi_q^2$$

We can prove this result as follows.

Proof. Suppose $U \sim N(0, \Omega)$. Then, Ω is a positive-definite, symmetric variance-covariance matrix. By symmetry, this matrix has an eigenvector decomposition,

$$\Omega = C\Lambda C' \quad \text{where} \quad CC' = C'C = I_n$$

Given it's positive definiteness, all eigenvalues in λ are strictly positive. We can therefore define,

$$\Lambda^{1/2} = \begin{bmatrix} \lambda_1^{1/2} & 0 & \cdots & 0 \\ 0 & \lambda_2^{1/2} & & \\ \vdots & & \ddots & \\ 0 & & & \lambda_q^{1/2} \end{bmatrix}$$

Likewise,

$$\Lambda^{-1/2} = \begin{bmatrix} \lambda_1^{-1/2} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-1/2} & & \\ \vdots & & \ddots & \\ 0 & & & \lambda_q^{-1/2} \end{bmatrix}$$

This allows us to define,

$$\Omega^{1/2} = C\Lambda^{1/2}C' \quad \text{since} \quad C\Lambda^{1/2}C'C\Lambda^{1/2}C' = C\Lambda C' = \Omega$$

and

$$\Omega^{-1/2} = C\Lambda^{-1/2}C' \quad \text{since} \quad C\Lambda^{-1/2}C'C\Lambda^{-1/2}C' = C\Lambda C' = \Omega^{-1}$$

This allows us to standardize the distribution of U .

$$\Omega^{-1/2}U \sim N(0, \Omega^{-1/2}\Omega\Omega^{-1/2}) = N(0, I_q)$$

and

$$(\Omega^{-1/2}U)' \Omega^{-1/2}U = U' \Omega^{-1}U = \sum_{j=1}^q Z_j^2 \sim \chi_q^2$$

□

Using this result, we can then define the F-statistic as,

$$\text{F-stat} = (R\hat{\beta} - r)'(s^2 R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/q$$

where we replace the unknown σ^2 by s^2 and scale by q , the $\text{rank}(R)$ (or number of hypotheses).

As above, we can rewrite this test-statistic as,

$$\begin{aligned}\text{F-stat} &= \frac{(R\hat{\beta} - r)'(\sigma^2 R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/q}{s^2/\sigma^2} \\ &\sim \frac{\chi_q^2/q}{\chi_{n-k}^2/(n-k)} \\ &= F_{q,n-k}\end{aligned}$$

We reject H_0 when,

$$\text{F-stat} > f_{q,n-k,1-\alpha}$$

where $f_{q,n-k,1-\alpha}$ is the $1 - \alpha$ percentile of the F -distribution with dof q and $n - k$. This is a one-sided test since the F-statistic is strictly positive.

Note, for $q = 1$, the F-statistic is given by the square of the T-statistic.

5.1 Asymptotic distribution

As before, the replacement of σ^2 with s^2 has no bearing on the asymptotic distribution of the test. Since, $p \lim(s_n^2/\sigma^2) = 1$,

$$\text{F-stat} \rightarrow_d \chi_q^2/q \quad \text{as } n \rightarrow \infty$$

6 Restricted OLS

You will likely be familiar with a second description of the F-statistic:

$$\text{F-stat} = \frac{(RSS_R - RSS_U)/q}{RSS_U/(n-k)}$$

This expression is equivalent to above and can be derived by solving the restricted (ordinary) least squares problem.

$$\min_b (Y - Xb)'(Y - Xb) \quad \text{s.t.} \quad Rb = r$$

This is a constrained optimization problem that can be solved using the Lagrangian function,

$$L(b, \lambda) = (Y - Xb)'(Y - Xb) + 2\lambda'(Rb - r)$$

The F.O.C.'s (w.r.t. b and λ), state at the optimum it must be that,

$$\begin{aligned} 0 &= 2X'X\tilde{\beta} - 2X'Y - 2R'\tilde{\lambda} \\ 0 &= R\tilde{\beta} - r \end{aligned}$$

From (1), we have,

$$\begin{aligned} \tilde{\beta} &= (X'X)^{-1}(X'Y - R'\tilde{\lambda}) \\ &= \underbrace{(X'X)^{-1}X'Y}_{\text{OLS}} - (X'X)^{-1}R'\tilde{\lambda} \\ &= \hat{\beta} - (X'X)^{-1}R'\tilde{\lambda} \end{aligned}$$

Pluggin this solution into (2), you get,

$$\begin{aligned} r &= R\hat{\beta} - R(X'X)^{-1}R'\tilde{\lambda} \\ \Rightarrow \tilde{\lambda} &= (R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) \\ \Rightarrow \tilde{\beta} &= \hat{\beta} - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) \end{aligned}$$

The residuals from the restricted model are given by,

$$\begin{aligned} \tilde{U} &= Y - X\tilde{\beta} \\ &= \underbrace{Y - X\hat{\beta}}_{M_X Y} + X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) \\ &= \hat{U} + X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) \end{aligned}$$

The Residual Sum of Squares (RSS) from the restricted model is then given by,

$$\begin{aligned} \tilde{U}'\tilde{U} &= (\hat{U} + X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r))'(\hat{U} + X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)) \\ &= \hat{U}'\hat{U} + (R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) \end{aligned}$$

The cross-product terms cancel, because $\hat{U}'X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) = Y'M_X X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) = 0$. Thus,

$$(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) = \underbrace{\tilde{U}'\tilde{U}}_{RSS_R} - \underbrace{\hat{U}'\hat{U}}_{RSS_U}$$

Since, $s^2 = RSS_U/(n - k)$, a scalar,

$$\text{F-stat} = \frac{(RSS_R - RSS_U)/q}{RSS_U/(n - k)}$$

Where $q = dof_r - dof_u$, the difference in residual degrees of freedom for the two models.

You can estimate the restricted model by imposing the restrictions of H_0 on the model. For example, in Problem Set 2, we tested $H_0 : \beta_2 = \beta_3$ **and** $\beta_4 + \beta_5 = 1$, by estimating the restricted model,

$$(Y - X_4) = \gamma_1 + \gamma_2(X_2 + X_3) + \gamma_5(X_5 - X_4) + \varepsilon$$

6.1 Testing full model

In the unique case where the hypothesis is given by,

$$H_0 : \beta_2 = \dots = \beta_k = 0$$

And the model includes a constant (i.e. $X_1 = \mathbf{1}$). Then, the restricted model amounts to a regression of Y on a constant: $Y = \beta_1 + u$.

$$RSS_R = (Y - \bar{Y})'(Y - \bar{Y}) = TSS$$

In this *unique* case, the F-statistic can be written in terms of R^2 (of the unrestricted model),

$$\text{F-stat} = \frac{R^2/(k - 1)}{(1 - R^2)/(n - k)}$$

This is the F-statistic reported by Stata when you estimate a linear regression model.

7 Heteroskedasticity

Both the t- and F-tests described above have their corresponding finite sample distributions under the assumptions that the error term in the CLRM is normally distributed (conditional on X) and that its variance is homoskedastic.

These distributions are not valid, for the finite sample, if the error terms are heteroskedastic. However, the tests do have the same asymptotic distributions with heteroskedasticity. For this reason, you should not use the t-distribution or F-distribution if your SEs are computed assuming heteroskedasticity. Instead, use the normal and chi-squared distributions.

References

- Cameron, A Colin, and Pravin K Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge university press.
- Verbeek, Marno. 2017. *A Guide to Modern Econometrics*. John Wiley & Sons.