5 Finite sample tests of linear hypotheses.

5.1 Linear hypotheses

The t-test is appropriate when the null hypothesis is a real valued restriction. However, more generally there may be multiple restrictions on the coefficient vector $\boldsymbol{\beta}$. Suppose we have $p > 1$ restrictions, we can express a linear hypothesis about $\boldsymbol{\beta}$ in the form $\boldsymbol{R}_{p\times k}\boldsymbol{\beta}_{k\times 1} = \boldsymbol{q}_{p\times 1}$.

Example (Nerlove's returns to scale). Nerlove studied the regression of the total cost of electricity production on demand (Q_i) and factor prices (capital, labour and fuel):

$$\log TC_i = \beta_1 + \beta_2 \log Q_i + \beta_3 \log p_{C_i} + \beta_4 \log p_{L_i} + \beta_5 \log p_{F_i} + \varepsilon_i$$

Economic theory suggests that $\beta_2 = \frac{1}{r}$ where r is the degree of returns to scale. To test constant returns we can use $H_0 : \beta_2 = 1$, which is trivially linear in components of $\boldsymbol{\beta}$. Alternatively we can write

$$R\beta = q$$

with $R = (0, 1, 0, 0, 0)$ and $q = 1$.
Further the total cost must be homogenous of degree 1 with respect to factor prices (doubling cost of all inputs doubles total cost). To test this we can consider $H_0 : \beta_3 + \beta_4 + \beta_5 = 1$. If we were to reject this it would suggest model misspecification.
To test these hypotheses simultaneously consider:

$$R\beta = q \quad \text{with} \quad R = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad q = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

To test H_0: $\boldsymbol{R\beta} = \boldsymbol{q}$ vs. H_1: $\boldsymbol{R\beta} \neq \boldsymbol{q}$ we compute the vector $\boldsymbol{R\hat{\beta}} = \boldsymbol{q}$ and reject the null if this vector is "too large" depending on the distribution of $\hat{\boldsymbol{\beta}}$ under H_0.

Definition 5.1.1: Wald statistic

When restrictions are a linear function of coefficients $\boldsymbol{\beta}$, we can write the Wald statistic as

$$W = (R\hat{\beta} - q)'(R\hat{V}_{\hat{\beta}}R')^{-1}(R\hat{\beta} - q)$$

i.e. a weighted Euclidean measure of the length of the vector $R\hat{\beta} - q$.

Note:-

As the Wald statistic is symmetric in the argument $R\hat{\beta} - q$ it treats positive and negative alternatives symmetrically. Thus the inherent alternative is always two-sided.
The Wald statistic is not-invariant to a non-linear transformation/reparametrisation of the hypothesis. For example, asking whether $\beta_1 = 1$ is the same as asking whether $\log \beta_1 = 0$; but the Wald statistic for $\beta_1 = 1$ is not the same as the Wald statistic for $\log \beta_1 = 0$. This is because there is in general no neat relationship between the standard errors of β_1 and $\log \beta_1$, so it needs to be approximated.

Assuming normal regression:

$$\begin{aligned}
\hat{\beta}|X &\sim N(\beta, \sigma^2 (X'X)^{-1}) \\
R\hat{\beta}|X &\sim N(R\beta, \sigma^2 R(X'X)^{-1}R') \\
R\hat{\beta} - q|X &\sim N(R\beta - q, \sigma^2 R(X'X)^{-1}R') \\
&\overset{H_0}{\sim} N(0, \sigma^2 R(X'X)^{-1}R')
\end{aligned}$$

We can thus standardise:

$$(\sigma^2 R(X'X)^{-1}R')^{-\frac{1}{2}}(R\hat{\beta} - q)|X \overset{H_0}{\sim} N(0, I_P)$$

$$(R\hat{\beta} - q)'(\sigma^2 R(X'X)^{-1}R')^{-1}(R\hat{\beta} - q)|X \overset{H_0}{\sim} \chi^2(p) \tag{5.1}$$

However, the true variance σ^2 is unknown, we thus replace it with the estimated $\hat{\sigma}^2$ to obtain the Wald statistic:

$$\begin{aligned}
W &= (R\hat{\beta} - q)'(\hat{\sigma}^2 R(X'X)^{-1}R')^{-1}(R\hat{\beta} - q) \\
&= \frac{(R\hat{\beta} - q)'(\sigma^2 R(X'X)^{-1}R')^{-1}(R\hat{\beta} - q)}{\hat{\sigma}^2/\sigma^2}
\end{aligned}$$

Note that this distribution is not $\chi^2(p)$ since $\hat{\sigma}^2$ is itself a random variable. We must consider the joint distribution of $\boldsymbol{\hat{\sigma}^2}$ and $\boldsymbol{\hat{\beta}}$ to make progress.

5.2 The joint distribution of $\hat{\sigma}^2$ and $\hat{\beta}$

Recall the definition of the variance estimator:

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}$$

To express this in terms of the population $\boldsymbol{\varepsilon}$'s examine the following, where we denote the residual maker matrix by $\mathbf{M_X} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$:

$$\begin{aligned}
(n-k)\hat{\sigma}^2 &= \hat{\varepsilon}'\hat{\varepsilon} \\
&= (\mathbf{M_X y})'\mathbf{M_X y} \\
&= (\mathbf{M_X}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}))'\mathbf{M_X}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{\varepsilon}'\mathbf{M_X'}\mathbf{M_X}\boldsymbol{\varepsilon} \qquad (\text{since } \mathbf{M_X X} = \mathbf{0}) \\
&= \boldsymbol{\varepsilon}'\mathbf{M_X}\boldsymbol{\varepsilon} \qquad (\text{since } \mathbf{M_X'}\mathbf{M_X} = \mathbf{M_X}\mathbf{M_X} = \mathbf{M_X})
\end{aligned}$$

Since $\mathbf{M_X}$ is symmetric, it is positive definite when all eigenvalues are positive. Since it is also idempotent, $\mathbf{M_X^2} = \mathbf{M_X}$, all eigenvalues are either zero or one, meaning $\mathbf{M_X}$ is positive semi-definite.[1]

> **Lemma 5.2.1 (Spectral decomposition).** For every $n \times n$ real symmetric matrix, the eigenvalues are real and the eigenvectors can be chosen real and orthonormal. Thus a real symmetric matrix $\mathbf{A}$ can be decomposed as
>
> $$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}'$$
>
> where $\mathbf{Q}$ is an orthogonal matrix whose columns are the real, orthonormal eigenvectors of $\mathbf{A}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix whose entries are the eigenvalues of $\mathbf{A}$.

[1] Alternatively since $\mathbf{M_X^2} = \mathbf{M_X}$ and $\mathbf{M_X}' = \mathbf{M_X}$, note that $\mathbf{v}'\mathbf{M_X v} = \mathbf{v}'\mathbf{M_X^2 v} = \mathbf{v}'\mathbf{M_X}'\mathbf{M_X v} = (\mathbf{v}'\mathbf{M_X})'(\mathbf{M_X v}) = \|\mathbf{M_X v}\|^2$ for all $\mathbf{v} \in \mathbb{R}^n$.

The spectral decomposition of $\mathbf{M_X}$ is $\mathbf{M_X} = \mathbf{H\Lambda H'}$ where $\mathbf{HH'} = \mathbf{I_n}$ and $\mathbf{\Lambda}$ is diagonal with the eigenvalues of $\mathbf{M_X}$ along the diagonal. Since $\mathbf{M_X}$ is idempotent with rank $n-k$, it has $n-k$ eigenvalues equalling 1 and k eigenvalues equalling 0, so:

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{I}_{n-k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_k \end{bmatrix}$$

In the normal regression $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I_n}\sigma^2)$, we want to find the distribution of $\mathbf{H}'\boldsymbol{\varepsilon}$. A linear combination of normals is also normal, meaning $\mathbf{H}'\boldsymbol{\varepsilon}$ is normal with mean $\mathbb{E}[\mathbf{H}'\boldsymbol{\varepsilon}] = \mathbf{H}'\mathbb{E}[\boldsymbol{\varepsilon}] = 0$ and variance $\mathrm{Var}(\mathbf{H}'\boldsymbol{e}) = \mathbf{H}'\mathbf{I_n}\sigma^2\mathbf{H} = \sigma^2\mathbf{H}'\mathbf{H} = \mathbf{I_n}\sigma^2$. Thus $\mathbf{H}'\boldsymbol{\varepsilon} \sim N(0, \mathbf{I_n}\sigma^2)$.

Let $\mathbf{u} = \mathbf{H}'\boldsymbol{\varepsilon}$, and partition $\underset{n\times 1}{\mathbf{u}} = \begin{bmatrix} \underset{(n-k)\times 1}{\mathbf{u_1}} \\ \underset{k\times 1}{\mathbf{u_2}} \end{bmatrix}$ where $\mathbf{u_1} \sim N(0, \mathbf{I_n}\sigma^2)$, then we have

$$\begin{aligned}
(n-k)\hat{\sigma}^2 &= \boldsymbol{\varepsilon}'\mathbf{M_X}\boldsymbol{\varepsilon} \\
&= \boldsymbol{\varepsilon}'\mathbf{H\Lambda H}'\boldsymbol{\varepsilon} \\
&= \mathbf{u}' \begin{bmatrix} \mathbf{I}_{n-k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_k \end{bmatrix} \mathbf{u} \\
&= [\mathbf{u}_1' \;\; \mathbf{u}_2'] \begin{bmatrix} \mathbf{I}_{n-k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_k \end{bmatrix} \begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \end{bmatrix} \\
&= \mathbf{u}_1'\mathbf{u_1}
\end{aligned}$$

where $\mathbf{u}_1'\mathbf{u_1}$ is the sum of $n-k$ squared normals with mean 0 and variance σ^2. We can transform each normal into a standard normal with division by σ; since each normal is squared we divide by σ^2. A sum of j squared standard normals is distributed χ^2_j, thus $\frac{(n-k)\hat{\sigma}^2}{\sigma^2}$ is distributed χ^2_{n-k}. Since $\boldsymbol{\varepsilon}$ is independent of $\hat{\boldsymbol{\beta}}$ it follows that $\hat{\sigma}^2$ is independent of $\hat{\boldsymbol{\beta}}$ as well.

Theorem 5.2.1. In normal regression,

$$\frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}$$

and is independent of $\hat{\boldsymbol{\beta}}$.

Corollary 5.2.1. In normal regression satisfying GM1-3, the normalised Wald statistic $\frac{W}{p}$, is distributed as $F(p, n-k)$ under the null.

Proof.

$$\frac{W}{p} = \frac{(R\hat{\beta} - q)'(\sigma^2 R(X'X)^{-1}R')^{-1}(R\hat{\beta} - q)/p}{\hat{\sigma}^2/\sigma^2} \sim \frac{\chi^2(p)/p}{\chi^2(n-k)/(n-k)} \sim F(p, n-k).$$

Where we have used 5.1 in the numerator, and Theorem 5.2.1 in the denominator. □

Consider a special case of testing a single restriction, that the j-th coefficient is zero. Then $R\hat{\beta}_j - q = \beta_j$:

$$\hat{\beta}_j | X \overset{H_0}{\sim} N(0, \sigma^2(X'X)^{-1}_{ij})$$

$$\frac{\hat{\beta}_j}{\sqrt{\sigma^2(X'X)^{-1}_{jj}}} | X \overset{H_0}{\sim} N(0, 1)$$

As before σ^2 is unknown, we can substitute in $\hat{\sigma}^2$, but the distribution will change:

$$\begin{aligned} t &= \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 (X'X)^{-1}_{jj}}} \\ &= \frac{\hat{\beta}_j / \sqrt{\sigma^2 (X'X)^{-1}_{jj}}}{\sqrt{\frac{(n-k)\hat{\sigma}^2}{\sigma^2} / (n-k)}} \\ t|X &\overset{H_0}{\sim} \frac{N(0,1)}{\sqrt{\chi^2(n-k)/(n-k)}} \\ &\overset{H_0}{\sim} t(n-k) \end{aligned}$$

Where we are using the fact that the numerator and denominator are independent conditional on X. Note that the square of the t-statistic equals the F-statistic for testing the single restriction.

$$\begin{aligned} t^2(n-k) &= \left(\frac{N(0,1)}{\sqrt{\chi^2(n-k)/(n-k)}} \right)^2 \\ &= \frac{\chi^2(1)/1}{\chi^2(n-k)/(n-k)} \\ &= F(1, n-k) \end{aligned}$$

It is preferable to use the t-statistic since we can test one-sided alternatives, by squaring it we kill the sign of $\hat{\beta}_j$, making it impossible to differentiate between left and right sided alternatives.

5.3 The familiar form of the F-statistic

Consider the following test:

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

Proposition 5.3.1. The normalised Wald statistic is equivalent to the following formula for the F-statistic when testing linear restrictions:

$$F = \frac{W}{p} = \frac{(RSS_r - RSS_u)/p}{RSS_u/(n-k)}$$

Proof. Let us impose the null hypothesis $R\beta = q$ when minimising the sum of squared residuals, denote the solution as the restricted least squares estimator $\tilde{\beta}$:

$$\min_{\beta} (Y - X\beta)'(Y - X\beta) \quad \text{s.t.} \quad R\beta = q$$

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

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \beta} &= -2X'(Y - X\tilde{\beta}) + R'\lambda = 0 \\ &\Rightarrow X'Y - X'X\tilde{\beta} = R' \left(\frac{\lambda}{2} \right) \\ &\Rightarrow (X'X)^{-1}X'Y - (X'X)^{-1}X'X\tilde{\beta} = (X'X)^{-1}R' \left(\frac{\lambda}{2} \right) \end{aligned}$$

Define the usual (unrestricted) OLS estimate as $\hat{\beta} = \hat{\beta}_{OLS} = (X'X)X'Y$

$$
\begin{aligned}
\Rightarrow \hat{\beta} - \tilde{\beta} &= (X'X)^{-1}R'\left(\frac{\lambda}{2}\right) \\
\Rightarrow \tilde{\beta} &= \hat{\beta} - (X'X)^{-1}R'\left(\frac{\lambda}{2}\right) \\
\Rightarrow R\tilde{\beta} &= R\hat{\beta} - R(X'X)^{-1}R'\left(\frac{\lambda}{2}\right)
\end{aligned}
$$

Since $R\tilde{\beta} = q$:

$$
\begin{aligned}
q &= R\hat{\beta} - R(X'X)^{-1}R'\left(\frac{\lambda}{2}\right) \\
R\hat{\beta} - q &= R(X'X)^{-1}R'\left(\frac{\lambda}{2}\right) \\
\Rightarrow (R(X'X)^{-1}R')^{-1}(R\hat{\beta} - q) &= \frac{\lambda}{2}
\end{aligned}
$$

Thus,

$$\tilde{\beta} = \hat{\beta} - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - q)$$

Now from the corresponding restricted and unrestricted residuals,

$$\hat{\varepsilon} = Y - X\hat{\beta}$$

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

Since $\hat{\varepsilon}'X = 0$ [a]

$$
\begin{aligned}
\tilde{\varepsilon}'\tilde{\varepsilon} &= (\hat{\varepsilon} + X(\hat{\beta} - \tilde{\beta}))'(\hat{\varepsilon} + X(\hat{\beta} - \tilde{\beta})) \\
&= \hat{\varepsilon}'\hat{\varepsilon} + \hat{\varepsilon}'X(\hat{\beta} - \tilde{\beta}) + (\hat{\beta} - \tilde{\beta})'X'\hat{\varepsilon} + (\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta}) \\
&= \hat{\varepsilon}'\hat{\varepsilon} + (\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta})
\end{aligned}
$$

and substituting $\hat{\beta} - \tilde{\beta} = (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - q)$,

$$
\begin{aligned}
\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon} &= ((X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - q))'\cancel{X'X(X'X)^{-1}}R'\left(R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - q) \\
&= (R\hat{\beta} - q)'\left(R(X'X)^{-1}R'\right)^{-1}\cancel{R(X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}}(R\hat{\beta} - q) \\
&= (R\hat{\beta} - q)'\left(R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - q)
\end{aligned}
$$

Finally,

$$\frac{W}{p} = \frac{(R\hat{\beta} - q)'\left(R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - q)/p}{\hat{\sigma}^2} = \frac{(\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon})/p}{\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}} = \frac{(RSS_r - RSS_u)/p}{RSS_u/(n-k)}$$

□

[a] I.e.: Unrestricted OLS residuals uncorrelated with regressors, see lecture 2 for an explanation