Lecture note 2

LR, Wald and LM tests principles in the regression model.

The likelihood-ratio (LR) test was motivated in Lecture 1, with reference to the first chapters in the book by Hendry and Nielsen.

The following there gives a little bit of details W and LM tests in the regression model, with reference to the Davidson and MacKinnon book. The notation deviates somewhat from Davidson and MacKinnon p 422-426 (in part because of higher generality there).

Throughout this note we assume independent and identical normal distribution for the disturbances

The Wald approach: Testing H_0 based on unrestricted estimation

A general representation of linear hypotheses:

$$H_0\text{:}\quad \mathbf{R}\boldsymbol{\beta}= \mathbf{q}.$$

The number of restrictions is r, so $\mathbf{R}$ is $r \times k$ and $\mathbf{q}$ is $r \times 1$. Clearly $r < k$.

To develop a so called *Wald* test statistics of H_0 we need (only) to estimate the model *without imposing* H_0. (aka unrestricted estimation). We start from the "discrepancy vector" $\mathbf{m}$

$$\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{q} = \mathbf{m}$$

If $\boldsymbol{\beta}$ is close to hypothesized value of β in H_0, then $\mathbf{m}$ is small. To derive a formal test we need to know the distribution of $\mathbf{m}$ under the null hypothesis. Since $\mathbf{m}$ is a linear combination of the normally distributed $\hat{\boldsymbol{\beta}}$, $\mathbf{m}$ is also normal with;

$$E\left[\mathbf{m} \mid \mathbf{X}\right] = \mathbf{0}$$

and

$$\begin{aligned}
Var\left[\mathbf{m} \mid \mathbf{X}\right] &= Var\left[\mathbf{R}\hat{\boldsymbol{\beta}} \mid \mathbf{X}\right] \\
&= E\left[(\mathbf{R}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}))(\mathbf{R}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}))' \mid \mathbf{X}\right] \\
&= \mathbf{R}E\left[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})' \mid \mathbf{X}\right]\mathbf{R}' \\
&= \mathbf{R}Var\left[\hat{\boldsymbol{\beta}} \mid \mathbf{X}\right]\mathbf{R}^{'} \\
&= \sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'
\end{aligned}$$

The so called Wald criterion is a *quadratic form* in the r deviations located in the $\mathbf{m}$ vector. It is defined by

$$W = \mathbf{m}'\left\{Var\left[\mathbf{R}\hat{\boldsymbol{\beta}} \mid \mathbf{X}\right]\right\}^{-1}\mathbf{m}$$

We know that a sum of r squared independent standard normal variables (i.e., each $N(0,1)$) is $\chi^2(r)$ distributed. This generalizes to W which therefore becomes

$$W \sim \chi^2(r \mid \mathbf{X}{,}H_0)$$

Since σ^2 is unknown it must be estimated, we use

$$\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$$

Since $\mathbf{MX} = \mathbf{0}$, we have

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{M}\boldsymbol{\varepsilon},$$

as also shown in Lecture note 1, and

$$\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 = \frac{1}{n-k}\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$$

The statistic

$$(n-k)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}}{\sigma^2}$$

is a quadratic form in the standard normal vector ε/σ^2. However this Chi-square does not have n degrees of freedom but $n-k$. This is due to $\mathbf{M}$ being symmetric idempotent which means that the characteristic roots of $\mathbf{M}$ are either 0 or 1. There are $n-k$ roots that are equal to 1, and k roots that are zero. As a result $(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon})/\sigma^2$ is a sum of $n-k$ independent standard normal variables:

$$(n-k)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}}{\sigma^2} \sim \chi^2(n-k \mid \mathbf{X})$$

$$F = \frac{W}{r}\frac{(n-k)}{(n-k)\frac{\hat{\sigma}^2}{\sigma^2}} = \frac{W}{r}\frac{\sigma^2}{\hat{\sigma}^2} \sim F(r, n-K \mid \mathbf{X}, H_0).$$

Using the definition of W we see that the unknown σ^2 disappears from the expression for F.

$$F = \frac{\mathbf{m}'\left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}\mathbf{m}}{\sigma^2 r}\frac{(n-k)}{(n-k)\frac{\hat{\sigma}^2}{\sigma^2}} = \frac{\mathbf{m}'\left[\hat{\sigma}^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}\mathbf{m}}{r} \tag{1}$$

compare (10.60) in DM, despite the slight difference in notation.

In the simplest case there is only one restriction on a single parameter. Let the hypothesis be H_0: $\beta_2 = 0$. This implies that $\mathbf{R}$ is $(1 \times k)$ and $\mathbf{q}$ =0 so that

$$\mathbf{m} = \hat{\beta}_2 - 0.$$

and

$$Var[\mathbf{m} \mid \mathbf{X}] = Var[\hat{\beta}_2 \mid \mathbf{X}] = \sigma^2\left\{(\mathbf{X}'\mathbf{X})^{-1}\right\}_{diag\ 2x2}$$

where $\left\{(\mathbf{X}'\mathbf{X})^{-1}\right\}_{diag\ 2x2}$ denotes the second element along the diagonal of $(\mathbf{X}'\mathbf{X})^{-1}$. The Wald criterion for this case becomes

$$W = (\hat{\beta}_2 - 0)\frac{1}{Var[b_1 \mid \mathbf{X}]}(\hat{\beta}_2 - 0) = \frac{(\hat{\beta}_2 - 0)^2}{\sigma^2\left\{(\mathbf{X}'\mathbf{X})^{-1}\right\}_{diag\ 2x2}}$$

and the F statistic which results from substituting σ^2 by $\hat{\sigma}^2$ is:

$$F = \frac{(\hat{\beta}_2 - 0)^2}{\hat{\sigma}^2\left\{(\mathbf{X}'\mathbf{X})^{-1}\right\}_{diag\ 2x2}} = \frac{\frac{(\hat{\beta}_2 - 0)^2}{\left\{(\mathbf{X}'\mathbf{X})^{-1}\right\}_{diag\ 2x2}}}{\hat{\sigma}^2} = \frac{\frac{(\hat{\beta}_2 - 0)^2}{\frac{1}{\sigma^2}\left\{(\mathbf{X}'\mathbf{X})^{-1}\right\}_{diag\ 2x2}}}{\frac{\hat{\sigma}^2}{\sigma^2}}$$

$$= \frac{\frac{(\hat{\beta}_2 - 0)^2}{\frac{1}{\sigma^2}\left\{(\mathbf{X}'\mathbf{X})^{-1}\right\}_{diag\ 2x2}}(n-k)}{\frac{\hat{\sigma}^2}{\sigma^2}(n-k)} \sim F(1, n-k)$$

As you know, in this special case with a single parameter restriction:

$$F = t^2$$

where

$$t = \frac{(\hat{\beta}_2 - 0)}{\sqrt{\hat{\sigma}^2\left\{(\mathbf{X}'\mathbf{X})^{-1}\right\}_{diag\ 2x2}}} = \frac{(\hat{\beta}_2 - 0)}{\sqrt{\widehat{Var}(\hat{\beta}_2)}}$$

i.e., the usual t-statistic which is Student$-t$ distributed under the null hypothesis: $t(n-k) \mid H_0$.

The Lagrange multiplier approach: Test based on restricted estimation

If we impose the r restrictions on the model and then do OLS on the *restricted model* we obtain restricted estimates on the remaining $k-r$ parameters. This is of course the restricted least squares estimator. It can be thought of as a constrained minimization problem, with a Lagrange multiplier that measures how much the restrictions "bite".

If the restrictions holds exactly so that $\mathbf{m} = \mathbf{0}$, then the restricted estimates are the same as the unrestricted and the Lagrange multiplier associated with the constrained minimization problem is zero.

If the restrictions do not hold exactly, then the Lagrange multiplier is different from zero and the inference problem is whether this is due to difference in population parameters or sampling variability.

In practice we typically compute the Lagrange multiplier (LM) test by estimation of both the restricted and the unrestricted model. The sum of squared residuals from the regressions with the k restrictions imposed is denoted $SSR(\tilde{\boldsymbol{\beta}})$ while the unrestrcited estimation gives $SSR(\hat{\boldsymbol{\beta}})$. If these two sums are equal, then $\mathbf{m} = \mathbf{0}$ and the associated Lagrange multiplier is zero.

Moreover, we can construct χ^2 statistics from $SSR(\tilde{\boldsymbol{\beta}})$ $SSR(\hat{\boldsymbol{\beta}})$ with degrees of freedom $n-(k-r)$ and $n-k$ respectively.

$$\frac{SSR(\tilde{\boldsymbol{\beta}})}{\sigma^2} - \frac{SSR(\hat{\boldsymbol{\beta}})}{\sigma^2} = \frac{\varepsilon_*'\mathbf{M}_*\varepsilon_*}{\sigma^2} - \frac{\varepsilon'\mathbf{M}\varepsilon}{\sigma^2} \sim \chi^2(r)$$

since the two statistics are $\chi^2(n-(k-r))$ and $\chi^2(n-k))$ respectively. The subscript $*$ refer to the restricted model/estimation. Therefore:

$$F = \frac{\frac{SSR(\tilde{\boldsymbol{\beta}})-SSR(\hat{\boldsymbol{\beta}})}{k}}{\frac{SSR(\hat{\boldsymbol{\beta}})}{n-k}} = \frac{\frac{\varepsilon_*'\mathbf{M}_*\varepsilon_*}{\sigma^2} - \frac{\varepsilon'\mathbf{M}\varepsilon}{\sigma^2}}{\frac{\varepsilon'\mathbf{M}\varepsilon}{\sigma^2}} \frac{n-k}{k} \sim F(r, n-k \mid H_0). \tag{2}$$

Hence the Wald and LM approaches lead to test statistics that have the same F distribution under the null hypothesis.

The two test statistics are also numerically identical. Intuitively, this is because also in the Wald case we do in fact impose exactly the same restrictions as in the LM case—but not on the model: they are imposed instead *after* estimation of the unrestricted model, in the Wald test statistic.