AAE 637 Lab 3: Hypothesis testing[*]

2/11/2015

The Wald test

You are familiar with the t-test. Usually the null hypothesis is that $\beta = 0$. If the null is $\beta = \beta_0$:

$$t_n \left(\beta_0 \right) = \frac{\hat{\beta} - \beta_0}{s \left(\hat{\beta} \right)}$$

where $s \left(\hat{\beta} \right)$ the standard error (i.e. the square root of the variance) of the estimate. For a two-sided test we reject the null hypothesis at the 5% confidence level if $|t_n| \geq 1.96$.

The Wald test is a generalization of the t-test so we can conduct tests dealing with multiple parameters and/or multiple restrictions. Let's square the t-stat:

$$\left(t_n \left(\beta_0 \right) \right)^2 = \frac{\left(\hat{\beta} - \beta_0 \right)^2}{s^2 \left(\hat{\beta} \right)}$$

$\hat{\beta} - \beta_0$ is assumed to be normally distributed, so $\left(\hat{\beta} - \beta_0 \right)^2$ is distributed χ^2 ("chi-squared"), with 1 degree of freedom. With $F_q \left(u \right)$ denoting the χ_q^2 density function, the critical value v for rejection of the null hypothesis at significance level α satisfies $\alpha = 1 - F_q \left(v \right)$. For $\alpha = 0.05$ and $q = 1$, this critical value is 3.84, i.e. we reject the null if $\left(t_n \left(\beta_0 \right) \right)^2 \geq 3.84$.
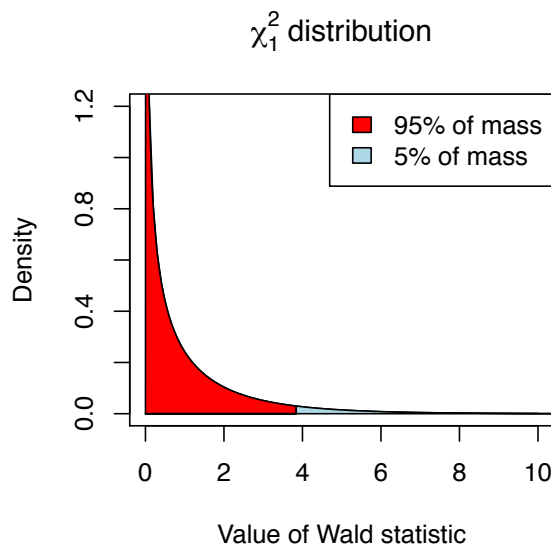
χ_1^2 distribution

For the Wald statistic, we can express any set of q linear restrictions on K parameter values with matrix notation:

$$\underset{q \times K}{\mathbf{R}} \underset{K \times 1}{\boldsymbol{\beta}} = \underset{q \times 1}{\mathbf{c}}$$

Say we have four parameters and we wish to test the joint hypothesis that $\beta_1 + \beta_2 - \beta_3 = 1$ and $\beta_1 = 10$.

Let's make the first row of $\mathbf{R}$:

$\beta_1 + \beta_2 - \beta_3 = \beta_1 \cdot 1 + \beta_2 \cdot 1 + \beta_3 \cdot (-1) + \beta_4 \cdot 0$

and the second row:

$\beta_1 = \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + \beta_4 \cdot 0$

$$\text{So } \mathbf{R} = \left[\begin{array}{cccc} 1 & 1 & -1 & 0 \\ 1 & 0 & 0 & 0 \end{array} \right], \mathbf{c} = \left[\begin{array}{c} 1 \\ 10 \end{array} \right]$$

The Wald statistic is given by:

$$W_n = \left[\underset{q \times K}{\mathbf{R}} \underset{K \times 1}{\hat{\boldsymbol{\beta}}} - \underset{q \times 1}{\mathbf{c}} \right]' \left(\underset{q \times K}{\mathbf{R}} \underset{K \times K}{\hat{\mathbf{V}}} \underset{K \times q}{\mathbf{R}'} \right)^{-1} \left[\underset{q \times K}{\mathbf{R}} \underset{K \times 1}{\hat{\boldsymbol{\beta}}} - \underset{q \times 1}{\mathbf{c}} \right]$$

$W_n \sim \chi_q^2$

$\hat{\mathbf{V}}$ is the estimated asymptotic variance-covariance matrix of the parameters $\hat{\boldsymbol{\beta}}$. You must use the correct number of degrees of freedom on χ^2 to do the Wald test and/or compute the p-value. The degrees of freedom here are the number of restrictions (i.e. rows of $\mathbf{R}$), not the number of parameters being tested.

The key thing here is that the covariance across the parameter estimates must be taken into account when you compare parameter estimates within the same model and/or dataset.

To get the p-value in MATLAB, input something like:

```
1 - chi2cdf(w_test_output, size(R_mat, 1))
```

The F test

The F statistic can be used to test multiple hypotheses as well, but it is extremely similar to the Wald stat. In fact, when the homoskedastic estimate of $\mathbf{V}$ is used in the Wald stat, we have:

$$F_n = \frac{W_n}{q}$$

However, the F statistic is compared against the F distribution, which incorporates a "small sample" correction. $F_n \sim F[q, n - K]$ where K is the number of estimated parameters and n is the number of observations. As $n \to \infty$, the F distribution converges to a χ_q^2 distribution if multiplied by q.

The F statistic can also be expressed in terms of the sum of squared errors of an OLS regression:

$$F_n = \frac{\left(SSE\left(\tilde{\boldsymbol{\beta}}_{cls}\right) - SSE\left(\hat{\boldsymbol{\beta}}\right) \right)/q}{SSE\left(\hat{\boldsymbol{\beta}}\right)/(n - K)}$$

Where $SSE\left(\tilde{\boldsymbol{\beta}}_{cls}\right)$ is the sum of squared errors for the constrained model. The F test is often used to check whether a model has any explanatory power at all, in which case $SSE\left(\tilde{\boldsymbol{\beta}}_{cls}\right)$ would be the total sum of squares. This type of F test was more relevant in the past when only small datasets were available.

The delta method

Say we know that the asymptotic distribution of a $K \times 1$ consistent set of estimators $\hat{\boldsymbol{\beta}}$ is jointly normal, i.e. as $n \to \infty$,

$$\sqrt{n} \left(\underset{K \times 1}{\hat{\boldsymbol{\beta}}} - \underset{K \times 1}{\boldsymbol{\beta}} \right) \xrightarrow{d} N \left(\underset{K \times 1}{\mathbf{0}}, \underset{K \times K}{\mathbf{V}} \right)$$

Further suppose that what we really want to get out of our estimation is some continuously differentiable transformation $\mathbf{g} : \mathbb{R}^K \to \mathbb{R}^L$ of our parameter estimates, not the estimates themselves. In addition to obtaining the transformed estimates, we need to know the uncertainty of the transformed estimates. The delta method gives us the following asymptotic approximation of the variance of the transformed estimates:

$$\sqrt{n} \left(\underset{L \times 1}{\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right)} - \underset{L \times 1}{\mathbf{g}\left(\boldsymbol{\beta}\right)} \right) \xrightarrow{d} N \left(\underset{L \times 1}{\mathbf{0}}, \underset{L \times K}{\left[\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{g}\left(\boldsymbol{\beta}\right)\right]} \underset{K \times K}{\mathbf{V}} \underset{K \times L}{\left[\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{g}\left(\boldsymbol{\beta}\right)\right]'} \right)$$

where $\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{g}\left(\boldsymbol{\beta}\right)$ is the Jacobian matrix:

$$\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{g}\left(\boldsymbol{\beta}\right) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} g_1\left(\boldsymbol{\beta}\right) & \cdots & \frac{\partial}{\partial \beta_K} g_1\left(\boldsymbol{\beta}\right) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \beta_1} g_L\left(\boldsymbol{\beta}\right) & \cdots & \frac{\partial}{\partial \beta_K} g_L\left(\boldsymbol{\beta}\right) \end{bmatrix}$$

The delta method works because it is a first-order Taylor Series approximation of the true variance.

Now say we can consistently estimate $\mathbf{V}$. Then our estimator for the variance of $\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right)$ is simply the "plug-in" estimator:

$$\left[\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right)\right] \hat{\mathbf{V}} \left[\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right)\right]'$$

Example:

Say $\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0.5 \\ 1.5 \end{bmatrix}$, $\hat{\mathbf{V}} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 2 \end{bmatrix}$, $\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right) = \begin{bmatrix} \hat{\beta}_1/\hat{\beta}_2 \\ \left(\hat{\beta}_2\right)^2 \end{bmatrix}$

Then: $\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right) = \begin{bmatrix} 1/\hat{\beta}_2 & -\hat{\beta}_1/\left(\hat{\beta}_2\right)^2 \\ 0 & 2\hat{\beta}_2 \end{bmatrix}$

This implies that the variance-covariance estimate for $\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right)$ is:

$$\hat{\mathbf{V}}_{\mathbf{g}(\boldsymbol{\beta})} = \begin{bmatrix} 1/1.5 & -0.5/(1.5)^2 \\ 0 & 2 \cdot 1.5 \end{bmatrix} \begin{bmatrix} 1 & -0.5 \\ -0.5 & 2 \end{bmatrix} \begin{bmatrix} 1/1.5 & -0.5/(1.5)^2 \\ 0 & 2 \cdot 1.5 \end{bmatrix}' = \begin{bmatrix} 0.691 & -2.\bar{3} \\ -2.\bar{3} & 18 \end{bmatrix}$$

Now say my null hypothesis is $\mathbf{g}\left(\boldsymbol{\beta}\right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

Then the Wald stat is:

$$W_n = \left(\begin{bmatrix} 1/3 \\ 2.25 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right)' \hat{\mathbf{V}}_{\mathbf{g}(\boldsymbol{\beta})}^{-1} \left(\begin{bmatrix} 1/3 \\ 2.25 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right) \approx 1.038$$

Using the χ_2^2 distribution, this yields a p-value of 0.60, so we cannot reject the null hypothesis.

When the parameter restrictions being tested are nonlinear, different algebraic formulations of the restriction can affect the value of the Wald statistic. This is undesirable. Next lab we will discuss likelihood-based tests that do not suffer from this problem.