Econometrics, Lecture 5.
Principles of testing and exact procedures

Ron Smith
EMS, Birkbeck, University of London

Autumn 2020

Testing

- Up to now we have focused on estimation of parameters: how to derive estimators and their variances (standard errors) and their properties.
- Now we are going to look at testing (inference).
- For this we need distributions, because we are going to ask whether something could happen by chance.
- We will look at the principles of testing that, according to the American Statistical Association, few understand.
- We will look at "exact" small sample tests t and F where we can derive the distribution and
- Asymptotic approximations, N and chi-squared where we cannot derive the small sample distributions exactly.

Distributions; Normal

- By the central limit theorem, under quite general assumptions, many estimators are normally distributed in large samples, whatever the distribution of the original variables.
- Linear functions of normally distributed variables are normally distributed. If $y \sim IN(\mu, \sigma^2)$, where $\sim$ means is distributed as then $a + by \sim N(a + b\mu, b^2\sigma^2)$. From this $z = (y - \mu)/\sigma$ is standard normal $z \sim IN(0, 1)$.
- The multivariate version is that if the $T \times 1$ vector $Y \sim N(M, \Sigma)$, where M is $T \times 1$, Σ is a $T \times T$ variance covariance matrix, specifying the dependence between observations. Then for given A and B of order $K \times 1$ and $K \times T$:
$$A + BY \sim N(A + BM, B\Sigma B'). \tag{1}$$

Distributions, Chi squared

- Sums of squares of T independent standard normal variables are distributed Chi-squared with T degrees of freedom:

$$\sum_{t=1}^{T}\left(\frac{y_t-\mu}{\sigma}\right)^2 \sim \chi^2(T).$$

- The multivariate version for the $T \times 1$ vector $Y \sim N(M, \Sigma)$ is a quadratic form:

$$(Y-M)'\Sigma^{-1}(Y-M) \sim \chi^2(T). \tag{2}$$

- If the errors $u \sim N(0, \sigma^2 I_T)$ then $u'u/\sigma^2$ is distributed as $\chi^2(T)$,
- But the residuals $\widehat{u}'\widehat{u}/\sigma^2 = u'Mu/\sigma^2$ are $\chi^2(rankM) = \chi^2(T-k)$.
- Alternatively

$$(T-k)(\frac{s^2}{\sigma^2}) \sim \chi^2(T-k).$$

t and F

A (Student's) t distribution with n degrees of freedom is given by

$$t(n) = z / \sqrt{\frac{\chi^2(n)}{n}}.$$

Fisher's F distribution is the ratio of two independent Chi-squared divided by their degrees of freedom.

$$F(n_1, n_2) = \frac{\chi^2(n_1)/n_1}{\chi^2(n_2)/n_2}.$$

and

$$t(n)^2 = z^2 / \frac{\chi^2(n)}{n} = F(1, n)$$

Principles of testing

- Suppose that we have prior information on θ, which suggests that elements of the parameter vector take specified values, such as zero or one or are linked by other restrictions.
- We wish to decide whether to accept or reject this hypothesis, called H_0, not knowing whether it is true or false.
- The possible outcomes are:

	H_o True	H_o False
Accept H_0	$\surd$	Type II error
Reject H_0	Type I error	$\surd$

- Legal analogy: H_0 : defendent innocent

The Neyman-Pearson approach to testing involves:

(a) a null hypothesis, H_0; e.g. for a scalar parameter: $H_0 : \beta = 1$;
(b) an alternative hypothesis, e.g. $H_1 : \beta \neq 1$, this is a two sided alternative, a one sided alternative would be $\beta < 1$;
(c) a test statistic, which does not depend on the true value of the parameters (is pivotal), (e.g. $(\widehat{\beta} - 1)/SE(\widehat{\beta})$, where $SE(\widehat{\beta})$ is the estimated standard error of $\widehat{\beta})$ with a known distribution when the null hypothesis is true (e.g. a central t distribution);
(d) a specified size α, the chosen probability of Type I error (rejecting H_0 when it is true) such as $\alpha = 0.05$, 5%.
(e) critical values, e.g. ± 1.96, so that if the null hypothesis is true the probability of lying outside the set covered by the critical values is α;
(f) a power function giving the probability of rejecting the null as a function of the true (unknown) value of β. The power of a test is the probability of rejecting H_0 when it is false $(1 - P($type two error$))$.

Decisions, decisions!

- The procedure is:
  - to accept (not reject) H_0 if the test statistic lies within the critical values and to reject H_0 if the test statistic lies outside the critical values.
  - to accept (not reject) H_0 if the p-value is greater than α, the chosen size, and reject H_0 if the p-value is less than α.
- If the p value is small, less than the chosen size (probability of rejecting null when true), e.g. 0.05, then the null hypothesis is rejected: a true hypothesis is unlikely to have generated a test statistic of that value.
- The American Statistical Association has a statement on p values and significance testing, since they are very often misunderstood. They say "Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data (for example, the sample mean difference between two compared groups) would be equal to or more extreme than its observed value."

Joint and individual tests can conflict

- Given a model:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t.$$

- We can do 2 individual t tests: $H_0^2 : \beta_2 = 0$ and $H_0^3 : \beta_3 = 0,.$
- We can do a joint F test: $H_0^J : \beta_2 = \beta_3 = 0$. This F test for all the slopes being zero, is done by most programs.
- The t tests could reject $\beta_2 = 0$ and $\beta_3 = 0$, both are significant, but the F test not reject them both zero: $\beta_2 = \beta_3 = 0$. They could be individually significant but jointly insignificant.The variables cancel out. If we have one we need the other, but we can drop both.
- The t tests could not reject $\beta_2 = 0$ and $\beta_3 = 0$, both are insignificant, but the F test reject both being zero. They could be individually insignificant but jointly significant. This is more common and happens when they are highly correlated, we can drop one of them, but not both.

Distinguish significance and importance

- The test asks whether the difference of the estimate from the null hypothesis could have arisen by chance, it does not tell you whether the difference is important.
- A coefficient may be statistically significant because it is very precisely estimated but so small as to be of no economic importance. Conversely the coefficient may be large in economic terms but have large standard errors so not be statistically significant.
- You need to understand the units and the context to judge whether an effect is important in economic terms.
- The test statistic and p value are conditional on the model and data used. Think of a test as informing a decision and consider the costs of the two sorts of mistakes. The costs can be embodied in some form of loss function or utility function.

Exact Tests

- In the LRM with linear restrictions and normal errors we can derive tests, where we know the exact distribution in small samples, rather than having to use asymptotic approximations.
- Suppose, our null hypothesis is a set of m linear restrictions of the form $R\beta = q$ or $R\beta - q = 0$, where R and q are known and of order $m \times k$ and $m \times 1$ respectively.
- The unrestricted model has k parameters, the restricted model $k - m$, each restriction reduces the number of parameters we estimate. In the case where m=k, all the parameters are specified, R is an identity matrix and the restrictions are $\beta = q$.

Example

Suppose that we have a model:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t,$$

with restrictions $\beta_2 = 1$, $\beta_3 = -\beta_4$. So $m = 2$, $k_u = 4$, $k_R = k_u - m = 2$. We write the restrictions in the form $R\beta = q$ as

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

the restricted model is

$$y_t - x_{2t} = \beta_1 + \beta_3 (x_{3t} - x_{4t}) + u_t^*,$$

Revise distributions

- We need distributions related to the normal and their multivariate equivalents: linear functions of multivariate normal are also normal, quadratic forms of standardised normal are Chi-squared.
- $Y \sim N(M, \Sigma)$,

$$A + BY \sim N(A + BM, B\Sigma B').$$

$$(Y - M)'\Sigma^{-1}(Y - M) \sim \chi^2(T).$$

- Residuals $\widehat{u}'\widehat{u}/\sigma^2 = u'Mu/\sigma^2$ are $\chi^2(rankM) = \chi^2(T - k)$.
- Central limit theorem means that in large samples estimators are normally distributed

Linear restrictions

If $y \sim N(X\beta, \sigma^2 I)$ and $\widehat{\beta} = (X'X)^{-1}X'y$, then

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

the restrictions are linear, so

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

Under $H_0 : R\beta - q = 0$

$$(R\widehat{\beta} - q) \sim N(0, \sigma^2 R(X'X)^{-1}R')$$

and

$$(R\widehat{\beta} - q)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - q) \sim \chi^2(m). \qquad (3)$$

Notice that this is a special case of the asymptotic Wald test statistic below and is of the same form.

Test statistic

(3) is not yet a test statistic because it depends on the unknown σ^2, but we know $(T-k)s^2/\sigma^2 \sim \chi^2(T-k)$ and that for independent Chi-squares:

$$\frac{\chi^2(m)/m}{\chi^2(T-k)/(T-k)} \sim F(m, T-k)$$

so

$$\frac{(R\widehat{\beta}-q)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\widehat{\beta}-q)/m}{[(T-k)s^2/\sigma^2]/(T-k)} \sim F(m, T-k)$$

and since the two unknown σ^2 cancel:

$$\frac{(R\widehat{\beta}-q)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta}-q)/m}{s^2} \sim F(m, T-k). \quad (4)$$

Alternative expression

It is easier to calculate (4) by writing it another way. Define the unrestricted and restricted estimated equations as

$$y = X\widehat{\beta} + \widehat{u}; \quad \text{and} \quad y = X\beta^* + u^*$$

then (4) can be written

$$\begin{aligned} \frac{(u^{*\prime}u^* - \widehat{u}'\widehat{u})/m}{\widehat{u}'\widehat{u}/(T-k)} &\sim F(m, T-k), \\ \frac{(RSSR - USSR)/m}{USSR/(T-k)} &\sim F(m, T-k) \end{aligned}$$

the ratio of (a) the difference between the restricted and unrestricted sum of squared residuals $(RSSR - USSR)$ divided by the number of restrictions m to (b) the unbiased estimate of the unrestricted variance: $USSR/(T-k)$.

F test for no regression

- Computer programs automatically print out a test for the hypothesis that all the slope coefficients in a linear regression are zero.
- The unrestricted estimate is

$$y_t = \widehat{\beta}_1 + \sum_{i=2}^{k} \widehat{\beta}_i x_{it} + \widehat{u}_t,$$

- The restricted is

$$y_t = \beta_1^* + u_t^*,$$

  where β_1^* is just the estimated mean.
- The hypothesis is $H_0 : \beta_2 = \beta_3 = ... = \beta_k = 0$.
- The test statistic is $F(k-1, T-k)$.

Non nested tests

Hypothesis tests require the two models being compared to be 'nested': one model (the restricted model) must be a special case of the other (the unrestricted or maintained model). In many cases we want to compare 'non-nested' models, e.g.

$$
\begin{aligned}
M_1 &: \quad y_t = a_1 + b_1 x_t + u_{1t} \\
M_2 &: \quad y_t = a_2 + c_2 z_t + u_{2t}
\end{aligned}
$$

where x_t and z_t are different scalar variables. There are no restrictions on M_1 that will give M_2 and vice-versa. We could nest them both in a general model:

$$M_3 : y_t = a_3 + b_3 x_t + c_3 z_t + u_{3t}.$$

The restriction $c_3 = 0$ gives M_1; so rejecting the restriction $c_3 = 0$ rejects M_1
The restriction $b_3 = 0$ gives M_2; so rejecting the restriction $b_3 = 0$ rejects M_2.

Four possible outcomes

:

1. Reject M_1, do not reject $M_2 : c_3 \neq 0; b_3 = 0;$
2. Reject M_2, do not reject $M_1 : b_3 \neq 0; c_3 = 0;$
3. Reject both; $b_3 \neq 0; c_3 \neq 0;$
4. Do not reject either: $b_3 = 0; c_3 = 0.$

There are a range of other non-nested tests available (Microfit has a large selection) but they all give rise to the same four possibilities. If x_t and z_t are highly correlated case 4 is quite likely as noted above.

Model selection criteria

- Model selection criteria like the AIC and BIC can be used both for nested or non-nested models.
- When comparing nested models the BIC can be interpreted as adjusting the size of the test (probability of type I error) with the number of observations. Suppose we have two models, $M1$ and $M2$, such that $M1$ has k parameters and is nested in $M2$ which has an extra variable and $k+1$ parameters. An LR test at the 5% level chooses $M2$ if $2(MLL_2 - MLL_1) > 3.84$. The BIC chooses $M2$ if $2(MLL_2 - MLL_1) > \ln T$.
- This is sensible. As the sample size grows the standard error of the parameter falls and with a large enough sample any hypothesis will be rejected at a constant size even if the deviation from the hypothesis is tiny.

Next time

- Often we cannot work out the exact small sample distributions of the test statistics and we have to use asymptotic approximations, the distribution of the test statistic if the sample is large.
- The normal distribution is the asymptotic approximation to the t distribution and the sample is large enough for the approximation to be good if $T > 30$. Much larger samples are need to give good approximations in other cases.
- There are three asymptotic tests: Wald, Likelihood Ratio and Lagrange Multiplier. They make different approximations