Econometrics, Lecture 6.
Asymptotic test procedures

Ron Smith
EMS, Birkbeck, University of London

Autumn 2020

Last time

- Introduced the Neyman-Pearson testing procedure
- Looked at the exact tests for linear restrictions in the LRM with normal errors.
  - For individual hypotheses the exact test was the t test
  - For joint hypotheses the exact test was the F test
- Looked at 2 different ways of writing the F test, in terms of $R\widehat{\beta} - q$, and in terms of the sum of squared residuals for the restricted and unrestricted models
- The number of parameters estimated in the unrestricted model equals the number of restrictions plus the number of parameters in the restricted model.
- Discussed non-nested testing and model selection criteria: AIC and BIC..

Principles

- If we cannot derive exact results for our tests, we often use asymptotic approximations relying on the central limit theorem.
- For t ratios the asymptotic distribution is normal with 5% critical value of 1.96 whereas for 10 degrees of freedom the exact critical value is 2.23. So we are more likely to reject using the asymptotic approximation.
- That is generally the case.
- The exact distribution is the F the asymptotic is the Chi-squared. Sometimes programs give both. Where available use F which makes small sample corrections.

Unrestricted estimates

- Model with known distribution, with pdf $f(y_t, \theta)$ where θ is a $k \times 1$ vector, sample $t = 1, 2, ..., T$, and log-likelihood function $LL(\theta) = \sum_{t=1}^{T} \ln f(y_t, \theta)$.
- The ML estimates maximise $LL(\theta)$, i.e. the $\widehat{\theta}$, which makes the $k \times 1$ vector

$$\frac{\partial LL(\theta)}{\partial \theta} = S(\widehat{\theta}) = 0$$

  $S(\widehat{\theta})$ is the score vector, the derivatives of the LL wrt each of the k elements of the vector θ evaluated at the values, $\widehat{\theta}$, which make $S(\widehat{\theta}) = 0$.
- These are the unrestricted estimates. The value of the Log-likelihood at $\widehat{\theta}$ is $LL(\widehat{\theta})$.

Restricted Estimates

- Given $m \leq k$ restrictions (maybe non-linear) $H_0 : R(\theta) = 0$, where $R(\theta)$ is an $m \times 1$ vector. If $m = k$, no parameters to estimate.
- The restricted estimator maximises

$$\$ = LL(\theta) - \lambda' R(\theta)$$

where λ is a $m \times 1$ vector of Lagrange Multipliers, so $\underset{(1 \times m)(m \times 1}{\lambda' R(\theta)}$.
- The first order condition, FOC, is a $k \times 1$ vector (remember matrix derivatives)

$$\frac{\partial \$}{\partial \theta} = \frac{\partial LL(\theta)}{\partial \theta} - \frac{\partial R(\theta)}{\partial \theta}\lambda = 0$$

$$\underset{(k \times m)(m \times 1)}{\frac{\partial R(\theta)}{\partial \theta}\lambda}$$

Restricted and unrestricted estimators

- The restricted estimate is θ^*. The $kx1$ vector $S(\theta^*)$ is the restricted (or efficient) score. $F(\theta^*)$ is $\partial R(\theta)/\partial\theta$ at θ^*.

$$\frac{\partial \$}{\partial \theta} = S(\theta^*) - F(\theta^*)\lambda^* = 0,$$

- Notice that at θ^*

$$\frac{\partial LL(\theta)}{\partial \theta} = S(\theta^*) = F(\theta^*)\lambda^* \neq 0$$

$\widehat{\theta}$	θ^*
$LL(\widehat{\theta}) \geq$	$LL(\theta^*)$
$S(\widehat{\theta}) = 0$	$S(\theta^*) \neq 0$
$R(\widehat{\theta}) \neq 0$	$R(\theta^*) = 0$

Implications

- If the hypotheses (restrictions) are true:
- (a) the two log-likelihoods should be similar, i.e. $LL(\widehat{\theta}) - LL(\theta^*)$ should be close to zero;
- (b) the unrestricted estimates should satisfy the restrictions $R(\widehat{\theta})$ should be close to zero (note $R(\theta^*)$ is exactly zero by construction);
- (c) the restricted score, $S(\theta^*)$, should be close to zero (note $S(\widehat{\theta})$ is exactly zero by construction) or equivalently the Lagrange Multipliers λ^* should be close to zero, the restrictions should not be binding.

Test procedures

- To judge 'close to zero' we use the asymptotic equivalents of the distributions used for the LRM.
- Asymptotically, by the CLT, the ML estimator is multivariate normal
$$\widehat{\theta} \;\tilde{a}\; N(\theta, I(\theta)^{-1})$$
- asymptotically the scalar quadratic form is chi-squared
$$(\widehat{\theta} - \theta)' I(\theta)(\widehat{\theta} - \theta) \;\tilde{a}\; \chi^2(k).$$
- asymptotically $R(\widehat{\theta})$ is also normal
$$R(\widehat{\theta}) \;\tilde{a}\; N(R(\theta), F(\theta)' I(\theta)^{-1} F(\theta))$$
where $F(\theta) = \partial R(\theta)/\partial\theta$.
- Similar to
$$(R\widehat{\beta} - q) \;\sim\; N(R\beta - q, \sigma^2 R(X'X)^{-1} R')$$

Test statistics

We have three procedures to get asymptotic test statistics for the m restictions $H_0 : R(\theta) = 0$; each asymptotically distributed (central) $\chi^2(m)$, if $H_0 : R(\theta) = 0$ is true:

- Likelihood Ratio Tests

$$LR = 2(LL(\widehat{\theta}) - LL(\theta^*)) \sim \chi^2(m)$$

- Wald Tests

$$W = R(\widehat{\theta})'[F(\widehat{\theta})'I(\widehat{\theta})^{-1}F(\widehat{\theta})]^{-1}R(\widehat{\theta}) \sim \chi^2(m)$$

  where the term in $[...]$ is an estimate of the variance of $R(\widehat{\theta})$ and $F(\theta) = \partial R(\theta)/\partial\theta$.

- Lagrange Multiplier (or Efficient Score) Tests where $\partial LL(\theta)/\partial\theta = S(\theta)$

$$LM = S(\theta^*)'I(\theta^*)^{-1}S(\theta^*) \sim \chi^2(m).$$

Remarks

- LR test is easy to calculate when both the restricted and unrestricted models have been estimated.
- The Wald test only requires the unrestricted estimates, but is not invariant to how you write non-linear restrictions.
- The Lagrange Multiplier test only requires the restricted estimates. In the LRM, the LM test is usually calculated using regression residuals. This is discussed below under diagnostic tests.
- For the LRM, the inequality W$>$LR$>$LM holds, so you are more likely to reject using W.

Non-linear restrictions

- Suppose $m = 1$, a single restriction that the product of two coefficients equals a third.
- We can write the same hypothesis in two different ways:
  - $H_0 : R_m(\theta) = \theta_1\theta_2 - \theta_3 = 0$.
  - $H_0 : R_d(\theta) = \theta_1 - \theta_3/\theta_2 = 0$
- These will give different values of the W test statistic. Using multiplication, $R_m(\theta)$, rather than division, $R_d(\theta)$, is usually better.

Source of this Non-linear restrictions

- Suppose we have AR1 (autoregressive of order 1) errors

$$
\begin{aligned}
y_t &= \beta x_t + u_t;\ \ u_t = \rho u_{t-1} + \varepsilon_t && (1)\\
y_t &= \beta x_t + \rho u_{t-1} + \varepsilon_t \\
y_t &= \beta x_t + \rho\left(y_{t-1} - \beta x_{t-1}\right) + \varepsilon_t \\
y_t &= \beta x_t - \rho\beta x_{t-1} + \rho y_{t-1} + \varepsilon_t && (2)\\
y_t &= \beta_0 x_t + \beta_1 x_{t-1} + \alpha_1 y_{t-1} + v_t && (3)
\end{aligned}
$$

- The restriction: $\beta_1 = -\beta_0\alpha_1$ on (3) gets (2) or (1). (1) has 2 parameters, (3) has 3.
- Number of restrictions is number of parameters in the unrestricted model minus the number of parameters in the restricted model.

Non-linear estimation

- 3 models with ε_{it}, $i = 1, 2, 3$ white noise

$$\begin{aligned} y_t &= \beta x_t + \varepsilon_{1t} &(4)\\ y_t &= \beta x_t + \rho u_{t-1} + \varepsilon_{2t} &(5)\\ y_t &= \beta_0 x_t + \beta_1 x_{t-1} + \alpha_1 y_{t-1} + \varepsilon_{3t} &(6) \end{aligned}$$

- How do you estimate (5) same as (1) above?

$$\begin{aligned} y_t &= \beta x_t + u_t; \;\; u_t = \rho u_{t-1} + \varepsilon_{2t}\\ y_t &= \beta x_t + \rho u_{t-1} + \varepsilon_{2t}\\ y_t &= \beta x_t + \rho\left(y_{t-1} - \beta x_{t-1}\right) + \varepsilon_{2t}\\ y_t - \rho y_{t-1} &= \beta(x_t - \rho x_{t-1}) + \varepsilon_{2t} \end{aligned}$$

Start with an initial estimate of ρ. transform the data and iterate. Feasible Generalised Least Squares.

Non-linear functions of coefficients

- ARDL model

$$y_t = \alpha_0 + \beta_0 x_t + \beta_1 x_{t-1} + \alpha_1 y_{t-1} + u_t$$

- In long-run equilibrium, no shocks $u_t = 0$, and variables constant $y_t = y_{t-1} = y$, $x_t = x_{t-1} = x$, so

$$\begin{aligned} y &= \alpha_0 + \beta_0 x + \beta_1 x + \alpha_1 y \\ y - \alpha_1 y &= \alpha_0 + (\beta_0 + \beta_1) x \\ y &= \frac{\alpha_0}{(1-\alpha_1)} + \frac{(\beta_0 + \beta_1)}{(1-\alpha_1)} x \end{aligned}$$

- Long run coefficient of x is $(\beta_0 + \beta_1)/(1-\alpha_1)$.
- We get standard errors from, delta method

$$R(\widehat{\theta}) \ \widetilde{a} \ \ N(R(\widehat{\theta}), F(\widehat{\theta})' I(\widehat{\theta})^{-1} F(\widehat{\theta}))$$

where $F(\widehat{\theta}) = \partial R(\theta)/\partial \theta$ evaluated at $\widehat{\theta}$.

Monte Carlo methods

- If the asymptotic distribution provides a poor approximation in sample size you are using or one cannot derive the asymptotic distribution, you use numerical procedures.
- Monte Carlo methods simulate the distributions by drawing random numbers.
- The t statistic for testing $H_0 : \rho = 1$, against $H_1 : \rho < 1$ in the model $y_t = \rho y_{t-1} + \varepsilon_t$ has a non standard distribution.
- Draw a sequence of $T + N$ random variables, ε_t^1, from some distribution.
- Then starting from $y_0 = 0$, generate $y_1 = y_0 + \varepsilon_1$, $y_2 = y_1 + \varepsilon_2$, ...,$y_{T+N} = y_{T+N-1} + \varepsilon_t$. Drop first N observations to remove effect of initial condition.
- This gives you a sample of T observations, y_t^1 from which you estimate $\widetilde{\rho}^1$.its se and t statistic $t\left(\widetilde{\rho}^1 = 1\right)$.
- Do this R times, getting $t\left(\widetilde{\rho}^r = 1\right)$. From this empirical distribution get the CV below which α, (e.g. 5%) of the t statistics fall.

Bootstrap

- This uses your estimates, say $Y_t = \alpha + \beta X_t + u_t$, $t = 1, 2, ..., T$.
- From the estimated residuals $\widehat{u}_t$, you randomly choose, **with replacement**, a new sample,$\widetilde{u}_t^1$ of T observations.
- You then construct a new sample of $\widetilde{Y}_t^1 = \widehat{\alpha} + \widehat{\beta} X_t + \widetilde{u}_t^1$. Using this sample, you get a new estimate of β, $\widetilde{\beta}^1$.
- You repeat this R times, getting $\widetilde{\beta}^r$, $r = 1, 2, ..., R$ where R is a large number. You then have a sampling distribution for your estimate from which you can calculate standard errors, confidence intervals, etc.
- This procedure will reflect any non-normality or heteroskedasticity in the residuals on the distribution. But it will not reflect any serial correlation in the residuals, because the order has been lost. There are other forms of bootstrap which can allow for that.

Next time

- We know how to estimate $\widehat{\beta}$ and its standard errors and their properties if certain assumptions hold.
- What happens when the assumptions fail to hold?
- What causes them to fail to hold?
- What can you do to fix it?
- The time after: how do you test for failure of the assumptions: diagnostic tests?