Econometrics, Lecture 8. Diagnostic tests

Ron Smith
EMS, Birkbeck, University of London

Autumn 2020

Last time

- We considered the diseases that models suffered: failure of the assumptions, model misspecification
- We looked at the effect of these diseases for the properties of the estimators of the coefficients and standard errors (variance covariance matrix) in the LRM
- We looked at ways that we might cure the disease, fix the problems
- Often the best way was to try and respecify the model
- Now we look at the symptoms: how to diagnose the disease.
- Always look at the patient: inspect graphs of actual and predicted values and residuals. Practice looking at the graphs to learn to recognise problems.
- Remember similar symptoms can come from lots of different diseases.

Diagnostic tests

- The estimates are only valid if a number of assumptions hold.
- Tests of those assumptions are called diagnostic or misspecification tests.
- Failure on a particular diagnostic test (rejection of the null that the model is well specified) only indicates that the model is sick, it does not tell you what the illness is.
- For instance, if you have chosen the wrong functional form you may fail tests for serial correlation.
- Most of these tests are LM tests, which use the residuals from the first stage regressions to ask if the residuals have the properties we would expect if the assumptions were true.
- The null hypothesis is always that the assumptions are true, the model is well specified. Thus if the p value for the test is greater than 0.05, you can accept the hypothesis that the model is well specified at the 5% level.

Tests

There are a very large numbers of these tests.

- Tests for serial correlation and non-linearity, use the residuals as the dependent variable in an auxilliary regressions;
- Tests for heteroskedasticity, use the squared residuals as the dependent variable in an auxilliary regressions;
- Tests for normality check that the third and fourth moments of the residuals have the values they should have under normality.
- For each null, e.g. constant variance (homoskedasticity) there are a large number of different alternatives (ways that the variance changes) thus lots of different tests for heteroskedasticity of different forms.
- Although the justification of these tests is asymptotic, versions which use F tests and degrees of freedom adjustment seem to work well in practice.

Structural stability

- The assumption that the parameters are constant over the sample is crucial and there are a variety of tests for constancy. Two are special cases of the F test for linear restrictions.
- It may happen in cross section - rich and poor countries, men and women, have different parameters - or time series - the parameters change att a breakpoint in time. We will assume time series, but the procedure is the same in cross section as long as you can divide the sample into the two groups that may be different.
- We have data for $t = 1, 2, ..., T$; believe the relationship may have shifted at T_1 and both sub-samples have more than k observations.

Models

- Unrestricted assumes parameters have changed so estimates separate regressions for $t = 1, 2, ..., T_1$ and $t = T_1 + 1, T_1 + 2, ..., T$;
- Define $T_2 = T - T_1$: X_1 a $T_1 \times k$ matrix, X_2 a $T_2 \times k$ matrix, etc. The subperiod models are:

$$\begin{aligned} y_1 &= X_1\beta_1 + u_1 \\ y_2 &= X_2\beta_2 + u_2 \end{aligned}$$

The USSR$= \left(\widehat{u}_1'\widehat{u}_1 + \widehat{u}_2'\widehat{u}_2\right)$ with d of f$=T - 2k$.

- $H_0 : \beta_1 = \beta_2$, k restrictions.
- The restricted model uses X a $T \times k$ matrix for the whole period
$$y = X\beta + u$$
- The RSSR$= \widehat{u}'\widehat{u}$ with d of f$=T - k$.

Test statistic

- We assume $u_i \sim IN(0, \sigma^2)$, $i = 1, 2$; the variances are the same in both periods.
- The formula for the F test is the usual one, where $m = k_u - k_r$ the difference between the number of parameters in the unrestricted $(k_u = 2k)$ and restricted $(k_r = k)$

$$\frac{(RSSR - USSR)/m}{USSR/(T - k_u)} \sim F(m, T - k_u)$$

- The null hypothesis is that $\beta_1 = \beta_2$, k restrictions, so $(k_u = 2k)$ $(k_r = k)$
- The test statistic is

$$\frac{[\widehat{u}'\widehat{u} - (\widehat{u}_1'\widehat{u}_1 + \widehat{u}_2'\widehat{u}_2)]/k}{(\widehat{u}_1'\widehat{u}_1 + \widehat{u}_2'\widehat{u}_2)/(T - 2k)} \sim F(k, T - 2k).$$

This is known as Chow's first or breakpoint test.

Predictive Failure test

Chow also suggested a second 'predictive failure' test for when $T_2 < k$ though it can be used when $T_2 > k$. The test statistic is:

$$\frac{[\widehat{u}'\widehat{u} - \widehat{u}_1'\widehat{u}_1)]/T_2}{\widehat{u}_1'\widehat{u}_1/(T_1 - k)} \sim F(T_2, T_1 - k).$$

This tests the hypothesis that in

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ X_2 & I_{T_2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \delta \end{bmatrix} + \begin{bmatrix} u_1 \\ 0 \end{bmatrix}$$

δ the $T_2 \times 1$ vector of forecast errors are not significantly different from zero. This has a dummy variable for each observation in the second period.

Variance equality

- Chow's first test assumes that the variances in the two periods are the same.
- This can be tested with the Variance Ratio or 'Goldfeld-Quandt' test:

$$\frac{s_1^2}{s_2^2} = \frac{\widehat{u}_1'\widehat{u}_1/(T_1-k)}{\widehat{u}_2'\widehat{u}_2/(T_2-k)} \sim F(T_1-k, T_2-k).$$

- Put the larger variance on top so the F statistic is greater than unity.
- Although this is an F test, it is not a test of linear restrictions on the regression parameters like the other F tests.
- It is a test for a specific form of heteroskedasticity, tests for other forms are given below.

Dummy variable form

- If the variances are equal, the two equations can be estimated together using dummy variables

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

- This gives the same coefficient estimates as two separate regressions, but different standard errors, since it imposes equality of variances, the separate regressions do not. In this form you can use robust variance covariance matrices for the tests.
- This can be rewritten

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ X_2 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 - \beta_1 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}. \qquad (1)$$

Then you can test which coefficients differ between periods. gretl uses this form.

Example

- For instance, suppose that $k = 3$, and we define dummy variables, $D1_t = 1$ period 1, $D1_t = 0$ period 2. $D2_t = 1$ period 2, $D2_t = 0$ period 1.
- 2 separate regressions give same coefficients as estimating on the whole period

$$\begin{aligned} y_t &= \beta_{11} D_{1t} + \beta_{12} D_{1t} x_{2t} + \beta_{13} D_{1t} x_{3t} \\ &\quad + \beta_{21} D_{2t} + \beta_{22} D_{2t} x_{2t} + \beta_{23} D_{2t} x_{3t} + u_t, \end{aligned}$$

$$y_t = \beta_{11} + \beta_{12} x_{2t} + \beta_{13} x_{3t} + \gamma_1 D2_t + \gamma_2 D2_t x_{2t} + \gamma_3 D2_t x_{3t} + u_t,$$

where $\gamma_1 = \beta_{21} - \beta_{11}$; $\gamma_2 = \beta_{22} - \beta_{12}$; $\gamma_3 = \beta_{23} - \beta_{13}$.

- Then if $H_0 : \gamma_2 = \gamma_3 = 0$ was not rejected, you would just need to include the dummy variable for $D2_t$ an intercept shift, but not the two $D2$ interaction terms.

Unknown break points

- Use recursive residuals such as the CUSUM and CUSUMSQ diagrams. These are presented as graphs of the statistics within two lines. If the graphs cross the lines it indicates structural instability.
- Recursive estimates, where the parameters are estimated on the first $k+1$ observations, the first $k+2$ and so on up to T.
- Rolling window estimates, which keeps a constant sub sample that moves through the whole sample.
- Quandt-Andrews unknown break point tests identify the most likely place for a break, but will have much less power than tests with a known break-point. You have to allow for the multiple testing problem.
- A sequence of 4 independent tests at the 5% level have an almost 20% chance $\left(1-0.95^4\right)=0.81$ of rejecting a true null.

Serial Correlation: DW

- For $u_t = \rho u_{t-1} + \varepsilon_t$
- The Durbin Watson statistic is

$$DW = \frac{\sum(\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum \widehat{u}_t^2} \approx 2(1-\rho)$$

- It should be close to 2. But
  - Critical values depend on X (it is a bounds test)
  - Not appropriate when there is a lagged dependent variable
  - Only tests for first order serial correlation.
- Lagrange Multiplier test more flexible

Serial Correlation: LM

- Suppose the DGP is with ε_t $IN(0, \sigma^2)$;

$$\begin{aligned} y_t &= \beta' x_t + u_t; \; u_t = \rho u_{t-1} + \varepsilon_t \\ y_t &= \beta' x_t + \rho u_{t-1} + \varepsilon_t \end{aligned}$$

- but we estimate

$$y_t = \widehat{b}' x_t + \widehat{u}_t$$

$$\widehat{u}_t = \beta' x_t + \rho u_{t-1} + \varepsilon_t - \widehat{b}' x_t$$

$$\widehat{u}_t = (\beta - \widehat{b})' x_t + \rho u_{t-1} + \varepsilon_t$$

- Test the hypothesis, $\rho = 0$, no serial correlation by an auxilliary regression of $\widehat{u}_t$ on x_t and $\widehat{u}_{t-1}$:

$$\widehat{u}_t = c' x_t + \rho \widehat{u}_{t-1} + v_t$$

and testing $\rho = 0$ with a t test.

- We replace u_{t-1} by $\widehat{u}_{t-1}$ and set missing residual $\widehat{u}_0$ to expected value zero.

Higher order serial correlation

- We can add more lagged residuals and test the joint hypothesis that all the coefficients of the lagged residuals are zero, with an F test.
- For quarterly data, there may be $4th$ order serial correlation, i.e. test all $\rho_i = 0, i = 1, 2, ..., 4$ in:

$$\widehat{u}_t = c' x_t + \rho_1 \widehat{u}_{t-1} + \rho_2 \widehat{u}_{t-2} + \rho_3 \widehat{u}_{t-3} + \rho_4 \widehat{u}_{t-4} + \varepsilon_t$$

- This is a different alternative hypothesis to that of no first order serial correlation, but the null hypothesis is the same.
- Including the x_t allows for $E(\widehat{b}) \neq \beta$ if lagged dependent variable in x_t.

Heteroskedasticity

- Estimate a 1st stage LRM get $\widehat{u}_t$ then run 2nd stage (auxilliary) regressions using $\widehat{u}_t^2$:

$$\widehat{u}_t^2 = \alpha + b' z_t + v_t$$

- $H_0 : b = 0$ since the expected value of $\widehat{u}_t^2$ should be constant: $\alpha = \sigma^2$. Test with F test.
- Under H_1, the variance, $\widehat{u}_t^2$, change with z_t.
- There are lots of ways that the variance could change, thus lots of possible candidates for z_t.
  - The most common is $z_t = x_t$, the regressors;
  - It could be the squares and cross-products of the regressors, often called the White test, can use a lot of d of f;
  - it could be the squared fitted values, the RESET version;
  - it could be $\widehat{u}_{t-1}^2$ testing for ARCH (autoregressive conditional heteroskedasticity); etc.

Normality

- If the residuals are normal then their
  - coefficient of skewness (third moment) should be zero and
  - coefficient of kurtosis (fourth moment) three.
- These are tested jointly by the Jarque-Bera test

$$T\left\{\frac{\mu_3^2}{6\mu_2^3}+\frac{1}{24}\left(\frac{\mu_4}{\mu_2^2}-3\right)^2\right\}\sim\chi^2(2)$$

where $\mu_j = \sum_{t=1}^{T}\widehat{u}_t^j\,/\,T$:

Quadratics

- Suppose we explain log wages, w_i, for men $i = 1, 2, ..., N$ by age, A, and years of education, E. Wages rise with age, peak then fall. Similarly with education: in the past getting a UK PhD cuts expected lifetime earnings by about 20%.
- In addition, the variables interact, wages peak later for more educated. This suggests a model of the form:

$$w_i = a + bA_i + cA_i^2 + dE_i + eE_i^2 + fE_iA_i + u_i$$

- Note the effect of age on wages varies with age and education

$$\frac{\partial w}{\partial A} = b + 2cA + fE$$

- We expect, $b, d, f > 0$ and $c, e < 0$.
- Regression coefficients measure the effect of changing one variable, holding the other variables constant. Cannot really change A_i holding A_i^2 constant.

Optimal age and education

- The age at which earnings is maximised is given by the solution to:

$$\frac{\partial w}{\partial A} = b + 2cA + fE = 0 \tag{2}$$

$$A^* = -\frac{b + fE}{2c}.$$

which if the estimated coefficients have the expected signs is positive (since $c < 0$) and peak earning age increases with education.

- The education at which earning is maximised is the solution to

$$\frac{\partial w}{\partial E} = d + 2eE + fA = 0$$

$$E^* = -\frac{d + fA}{2e}$$

- Interpretation?

Testing for non-linearity

- Adding squares and cross-products can use up degrees of freedom. If there are k original variables, there are k(k+1)/2 regressors in the augmented one, for k=2:3; k=3:6; k=4:10; k=5:15 regressors.
- Instead, we estimate a first stage linear regression:

$$y_t = \widehat{\beta}' x_t + \widehat{u}_t \tag{3}$$

  with fitted values $\widehat{y}_t = \widehat{\beta}' x_t$; and run a second stage regression:

$$\widehat{u}_t = b' x_t + c\widehat{y}_t^2 + v_t$$

  and test whether $c = 0$.
- Tests using powers of the fitted values are known as Ramsey RESET tests.

RESET tests

- Here $\widehat{y}_t^2$ is being used to proxy squares and cross-products of the x_t :

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

$$\begin{aligned} \widehat{y}_t^2 &= (\widehat{\beta}_1 + \widehat{\beta}_2 x_{2t} + \widehat{\beta}_3 x_{3t})^2 \\ &= \widehat{\beta}_1^2 + \widehat{\beta}_2^2 x_{2t}^2 + \widehat{\beta}_3^2 x_{3t}^2 + 2\widehat{\beta}_1 \widehat{\beta}_2 x_{2t} + ... \end{aligned}$$

- Some programs do the test differently, running

$$y_t = d' x_t + e\widehat{y}_t^2 + v_t$$

and tests $e = 0$. Noting that

$$y_t = \widehat{y}_t + \widehat{u}_t = \widehat{\beta}' x_t + \widehat{u}_t,$$

$d' = (b + \widehat{\beta})'$ so they give identifical test statistics.

- Higher powers of $\widehat{y}_t$ can also be added.

Next time

- We have covered all the main elements of the underlying principles:
  - how to estimate parameters and their standard errors,
  - how to check that the model is well specified,
  - how to choose between models
  - how to conduct tests.
- For the rest of the term we are going to be applying these principles to a range of different models.
- Next up is univariate time series ARIMA models of stochastic processes very widely used in forecasting.