Econometrics, Lecture 7.
Problems, consequences and cures

Ron Smith
EMS, Birkbeck, University of London

Autumn 2020

Last time

- Asymptotic test procedures
- Wald
- Likelihood Ratio
- Lagrange Multiplier
- Bootstrap
- Now look at what happens when the assumptions fail.
    - Diseases models get, their effects and how to cure them.
    - Next time, how to detect them, symptoms.
    - Do distinguish "true" errors and estimated residuals.
    - Estimated residuals have problems because we have the wrong model.

Gaussian LRM

If y is a $T \times 1$ vector and X is a $T \times k$ full rank matrix of exogenous variables, then conditional on X,

$$y \sim N(X\beta, \sigma^2 I)$$

and the ML estimator is

$$\widehat{\beta} = (X'X)^{-1}X'y$$

a linear function of y, so $\widehat{\beta}$ is also normally distributed:

$$\begin{aligned}\widehat{\beta} &\sim N\{(X'X)^{-1}X'X\beta,\ (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1}\} \\ &\sim N\{\beta, \sigma^2(X'X)^{-1}\}\end{aligned}$$

Estimates

- $\widehat{\beta}$ is unbiased, $E(\widehat{\beta}) = \beta$, only requires exogeneity, with predetermined (e.g. lagged dependent variable) it is consistent
- With normality it is fully efficient, its variance covariance matrix attains the lower bound $\sigma^2(X'X)^{-1}$, with Gauss-Markov assumptions BLUE.
- We generally estimate the variance covariance matrix by $s^2(X'X)^{-1}$, where $s^2 = \widehat{u}'\widehat{u}/(T-k)$, the unbiased estimator.
- The square roots of the diagonal elements of this matrix give the standard errors of the individual regression coefficients, e.g. β_i and the off diagonal elements give the covariances between regression coefficients, e.g. $Cov(\beta_i, \beta_j)$.

X not full rank

- Rank(X)$< k$. If X is not of full rank k, because there is an exact linear dependency between some of the variables, the OLS/ML estimates of β are not defined and there is said to be exact multicollinearity. The model should be respecified to remove the exact dependency, e.g. from the dummy variable trap: including an intercept and four seasonal dummies.
- When there is high, though not perfect, correlation between some of the variables there is said to be multicollinearity. This does not involve a failure of any assumption and you cannot test for it. Ignore VIFs etc. High correlations cause standard errors to be larger, you cannot estimate the effect so precisely. So if coefficients are significant there is no problem.
- Worrying about multicollinearity causes more problems than multicollinearity.

X not exogenous

- If the X are not strictly exogenous, independent of u, the estimates of β are biased.
- If the X are predetermined, uncorrelated with u, (e.g. lagged dependent variables where the disturbance term is not serially correlated), they will remain consistent.
- Otherwise, the estimator will be biased and inconsistent.
- Even if the exogeneity assumption fails least squares gives the best (minimum variance) linear predictor of y.
- Exogeneity fails because of
  - simultaneity: two way causation
  - measurement errors in the independent variable
  - omitted variables.

Omitted variables

- If the DGP is

$$y_t = \alpha + \beta x_t + \gamma z_t + u_t \tag{1}$$

  and you omit z_t, and estimate

$$y_t = a + b x_t + v_t. \tag{2}$$

- The relation between omitted and included variables is:

$$z_t = c + d x_t + w_t \tag{3}$$

- If they are unrelated $d = 0$. Replace z_t in (1) by the RHS of (3) you get:

$$y_t = \alpha + \beta x_t + \gamma(c + d x_t + w_t) + u_t \tag{4}$$

Omitted variables

- Can rewrite (4) as

$$y_t = (\alpha + \gamma c) + (\beta + \gamma' d)x_t + (\gamma w_t + u_t).$$

- Thus in $y_t = a + bx_t + v_t$; $b = (\beta + \gamma d)$ and $v_t = (\gamma w_t + u_t)$.
- The coefficient of x_t in (2) will only be an unbiased estimator of β, the coefficient of x_t in (1) if either $\gamma = 0$ (z_t really has no effect on y_t) or $d = 0$, (there is no correlation between the included and omitted variables).
- Since v_t contains the part of z_t that is not correlated with x_t, w_t, there is no reason to expect w_t to be serially uncorrelated or homoskedastic. Thus misspecification, omission of z_t, may cause the estimated residuals to show these problems. This generalises easily to x_t and z_t being vector

Non-scalar covariance matrix

- If $Var(y \mid X) = E(uu') = \sigma^2\Omega \neq \sigma^2 I$:
  - the variances (diagonal terms of the matrix) are not constant equal to σ^2 (heteroskedasticity) and/or
  - the off diagonal terms, the covariances, are not equal to zero (failure of independence, serial correlation, autocorrelation).
- If so $\widehat{\beta}$ remains unbiased but not minimum variance (efficient).
- $V(\widehat{\beta}) \neq \sigma^2(X'X)^{-1}$, but $V(\widehat{\beta}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$.
- Robust variance-covariance matrices, use estimates of $X'\Omega X$ (White Heteroskedasticity consistent covariance matrices or Newey-West heteroskedasticity and autocorrelation consistent, HAC, ones).
- Residual serial correlation or heteroskedasticity may not indicate a non-scalar covaraiance matrix but that the model is wrongly specified, So often residual serial correlation or heteroskedasticity should lead you to respecify the model rather than to use Generalised Least Squares.

Generalised Least Squares, GLS

- If $y \sim N(X\beta, \sigma^2\Omega)$ its distribution is given by:

$$2\pi^{-T/2} \mid \sigma^2\Omega \mid^{-1/2} \exp\left\{-\frac{1}{2}(y - X\beta)'(\sigma^2\Omega)^{-1}(y - X\beta)\right\}.$$

  If $\Omega = I$, the term in the determinant,
  $\mid \sigma^2\Omega \mid^{-1/2} = (\sigma^2)^{-T/2}$.
- For Ω known the ML Estimator is GLSquares

$$\begin{aligned} \beta^{GLS} &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \\ V(\beta^{GLS}) &= \sigma^2(X'\Omega^{-1}X)^{-1} \end{aligned}$$

- OLS chooses β to make $(\sigma^2)^{-1}X'\widehat{u} = 0$, $\widehat{u} = y - X\widehat{\beta}$
- GLS chooses β to make $(\sigma^2)^{-1}X'\Omega^{-1}\widetilde{u} = 0$, $\widetilde{u} = y - X\beta^{GLS}$.
- GLS is implemented by finding a 'transformation' P such that $P'P = \Omega^{-1}$ and $P\Omega P' = I$. You can always do this as Ω must be a positive-definite symmetric matrix.

Transformed GLS

- Transform the data by premultiplying by P

$$Py = PX\beta + Pu$$

$$y^* = X^*\beta + u^*$$

- Apply OLS to the transformed data, $y^* = Py$, etc. Efficient since

$$\begin{aligned} E(u^* u^{*\prime}) &= E(Puu'P') = PE(uu')P' \\ &= P(\sigma^2\Omega)P' = \sigma^2 P\Omega P' = \sigma^2 I. \end{aligned}$$

Example

- Suppose

$$\begin{aligned} y_i &= \alpha + \beta x_i + u_i, \\ E(u_i^2) &= \sigma^2 x_i^2. \end{aligned}$$

  transform by dividing by x_i

$$\begin{aligned} \frac{y_i}{x_i} &= \frac{\alpha}{x_i} + \beta + \frac{u_i}{x_i} \\ E\left(\frac{u_i}{x_i}\right)^2 &= \frac{\sigma^2 x_i^2}{x_i^2} = \sigma^2 \end{aligned}$$

- Common when x_i is a size measure and y_i / x_i a share.
- Weight for heteroskedasticity, not to match population.

Feasible GLS

- In practice, Ω is rarely known completely, but may be known up to a few unknown parameters which can be estimated and used to form an estimate of Ω, and P.
- This is the Feasible or Estimated GLS estimator. It generally differs from the exact ML estimator.
- We had example with serially correlated errors $u_t = \rho u_{t-1} + \varepsilon_t$.
- The text books give large number of examples of FGLS estimators, differing in the assumed structure of Ω.
- FGLS useful sometimes, but often to fix problems with residuals, better to respecify the model or correct the standard errors than to apply FGLS.

Serial Correlation

- Suppose as before

$$\begin{aligned} y_t &= \beta x_t + u_t;\ \ u_t = \rho u_{t-1} + \varepsilon_t && (5) \\ y_t &= \beta x_t + \rho u_{t-1} + \varepsilon_t \\ y_t &= \beta x_t + \rho\left(y_{t-1} - \beta x_{t-1}\right) + \varepsilon_t \\ y_t - \rho y_{t-1} &= \beta(x_t - \rho x_{t-1}) + \varepsilon_t && (6) \end{aligned}$$

- This is the transformed equation for GLS, ρ known. for FGLS you start with an initial estimate of ρ then iterate.
- But this is a restricted version of

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \alpha_1 y_{t-1} + v_t \qquad (7)$$

- The restriction is $\beta_1 = -\beta_0 \alpha_1$
- Better to start with (7) and test down: general to specific.

u not Gaussian

- If normality does not hold the Least Squares estimator, $\widehat{\beta} = (X'X)^{-1}X'y$, is no longer the Maximum Likelihood estimator and is not fully efficient, but it is the minimum variance estimator in the class of linear unbiased estimators (biased or non-linear estimators may have smaller variances).
- In small samples tests will not have the stated distributions, though asymptotically they will be normal, because of the central limit theorem.
- If the form of the distribution is known (e.g. a t distribution) maximum likelihood estimators can be derived for that particular distribution and they will be different from the OLS estimators.

Next time

- How do we find out whether the assumptions hold?
- Always look at plots of the residuals. Look for outliers, serial correlation, heteroskedasticity.
- There are diagnostic tests which are mainly Lagrange multiplier tests using the estimated residuals.
- The model you estimate is treated as the restricted model and the unrestricted model has the problems.