1 General linear Model Continued..

We have

$$\begin{aligned} y &= X\beta + u \\ & \quad X \text{ non random} \\ u &\backsim N(0, \sigma^2 I_n) \\ \widehat{\beta} &= (X'X)^{-1}X'y \end{aligned}$$

We know

$$\begin{aligned} E(\widehat{\beta}) &= \beta \\ Var(\widehat{\beta}) &= \sigma^2 (X'X)^{-1} \end{aligned}$$

We saw that

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

so $\widehat{\beta} - \beta$ is a linear function of a normally distributed random vector u and so is itself normal.

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

From before

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{\widehat{u}'\widehat{u}}{n-k} = \frac{u'Nu}{n-k} \\ where\ N &= \left[I_n - X(X'X)^{-1}X'\right] \end{aligned}$$

Now we can show that for any symmetric idempotent matrix with $\varepsilon \backsim N(0, \sigma^2 I_n)$

$$\frac{\varepsilon Q \varepsilon}{\sigma^2} \backsim \chi^2_{trace(Q)}$$

Since N is symmetric and idempotent

$$\frac{u'Nu}{\sigma^2} = \frac{(n-k)\widehat{\sigma}^2}{\sigma^2} \backsim \chi^2_{n-k}$$

Now

$$\begin{aligned} \widehat{\beta} - \beta &= (X'X)^{-1}X'u \\ &\Longrightarrow X(\widehat{\beta} - \beta) = X(X'X)^{-1}X'u = Mu \\ &\Longrightarrow (\widehat{\beta} - \beta)'X'X(\widehat{\beta} - \beta) = u'M^2u = u'Mu \end{aligned}$$

as M is idempotent. So

$$\frac{(\widehat{\beta} - \beta)'X'X(\widehat{\beta} - \beta)}{\sigma^2} = \frac{u'Mu}{\sigma^2} \backsim \chi^2_k$$

As $MN = 0$ these two results are independent

$$\frac{u'Mu/k}{u'Nu/n-k} \backsim \frac{\chi^2_k}{\chi^2_{n-k}} \backsim F_{(k,n-k)}$$

That is

$$\frac{(\widehat{\beta}-\beta)'X'X(\widehat{\beta}-\beta)/n}{(y-X\widehat{\beta})'(y-X\widehat{\beta})/n-k} \backsim F_{(k,n-k)}$$

which can be used to construct significance tests on β. replace the value of β given by the null and compute the ratio.

More usually we dont test the hypothesis on the whole parameter space. Consider the linear restriction

$$R\beta = d$$

Want to test

$$\begin{aligned} H_0 &: R\beta = d \\ H_1 &: R\beta \neq d \end{aligned}$$

R is an rxk matrix of constants with $Rank(R) = r$ and d is an rx1 vector. Many economic hypotheses can be put on this form.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

To test

$$\begin{aligned} \beta_1 &= 0 : R = (0100) : d = 0 \\ \beta_2 &= -\beta_3 : R = (0011) : d = 0 \\ \beta_1 + \beta_2 + \beta_3 &= 1 : R = (0111) : d = 1 \end{aligned}$$

In each case there is one restriction and $Rank(r) = 1$. For

$$\beta_0 = 0; \beta_1 = 1; \beta_2 = -\beta_3$$

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} : d = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

and the $Rank(R) = 3$. In all of these cases the restricted model can be estimated by imposing the restrictions on the data used to estimate the model

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \\ y_i &= \beta_0 + x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \text{ when } \beta_1 = 1 \\ y_i - x_{i1} &= \beta_0 + \beta_2 (x_{2i} - x_{3i}) + u_i \text{ when } \beta_2 = -\beta_3 \\ y_i - x_{i1} &= \beta_2 (x_{2i} - x_{3i}) + u_i \text{ when } \beta_0 = 0 \end{aligned}$$

Note that in each case after you impose r restrictions only estimate k-r parameters. So each restriction reduces the number of parameters by one. What we are doing is minimising $\widehat{u}'\widehat{u}$ subject to the restrictions $R\beta = d$

As a Lagrangian

$$\begin{aligned} L &= \widehat{u}'\widehat{u} + \lambda(R\beta - d) \\ &= (y - X\beta)'(y - X\beta) - \lambda(R\beta - d) \end{aligned}$$

The restricted estimator can be written

$$\beta^* = \widehat{\beta} - (X'X)^{-1}R\lambda$$

So define unrestricted and restricted residual sum of squares

$$\begin{aligned} URSS &= (y - X\widehat{\beta})'(y - X\widehat{\beta}) \\ RRSS &= (y - X\beta^*)'(y - X\beta^*) \end{aligned}$$

Then if the null is true

$$\frac{(RRSS - URSS)/r}{URSS/n - k} \backsim F_{(r,n-k)}$$

which is the familiar F test. This measures whether an increase in RSS per degree of freedom from imposing the restrictions is large relative to the variance. If it is large reject the restriction, if it is small accept the restriction. What is large depends on the significance level chosen.Note that it is distributed F only if the null is true. If $R\widehat{\beta} \neq d$ then β^* will be a long way from $\widehat{\beta}$.

So an alternative way of looking at this is that if the restrictions are not true $R\widehat{\beta} - d$ will be far from zero. To measure how far we could standardise by the estimated variance of $R\widehat{\beta} - d$ namely $\widehat{V}(\ R\widehat{\beta} - d)$. Then

$$\frac{1}{r}\left(R\widehat{\beta} - d\right)'\widehat{V}(R\widehat{\beta} - d)^{-1}(R\widehat{\beta} - d) \backsim F_{(r,n-k)}$$

with

$$\widehat{V}(R\widehat{\beta} - d)^{-1} = \frac{1}{\sigma^2}\left[R(X'X)^{-1}R'\right]^{-1}$$

An important special case is where the hypothesis to be tested is that a particular β equals a particular value, eg $\beta_1 = \overline{\beta}_1$

Then $R = [0100...] : d = \overline{\beta}_1$

$$R\widehat{\beta} - d = \widehat{\beta}_1 - \overline{\beta}_1$$

with $\widehat{V}(R\widehat{\beta} - d)$ the estimated variance of $\widehat{\beta}_1$

So

$$\frac{\left(\widehat{\beta}_1 - \overline{\beta}_1\right)^2}{Var(\widehat{\beta}_1)} \backsim F_{(1,n-k)}$$

Now if $w^2 \backsim F_{(1,n-k)} \Longrightarrow w \backsim t_{n-k}$ so

$$\frac{\widehat{\beta}_1 - \overline{\beta}_1}{se(\widehat{\beta}_1)} \backsim t_{n-k}$$

which is the commonly used t test

These are small sample tests and require the model be linear and the errors normal. It these dont hold then have to use asymptotic tests.

The 3 widely asymptotic tests are:

- – The likelihood ratio test: uses restricted and unrestricted
  – The Wald test: uses only the unrestricted
  – The Lagrange Multiplier (LM) test: uses only restricted

These are asymptotically equivalent and are each distributed χ^2_r (where r is the number of restrictions) if the null hypothesis is true.

All can be written as measures of distance standardised by a variance covariance matrix, each differing by which distance is measured.

The ML estimates are those which maximise $LL(\theta)$, i.e. the $\widehat{\theta}$, which make

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

where $S(\widehat{\theta})$ is the score vector, the derivatives of the LL with respect to each of the k elements of the vector θ evaluated at the values, $\widehat{\theta}$, which make $S(\theta) = 0$. We will call these the unrestricted estimates and the value of the Log-likelihood at $\widehat{\theta}$, $LL(\widehat{\theta})$.

Suppose theory suggests $m \leq k$ prior restrictions of the form $R(\theta) = 0$. If $m = k$, theory specifies all the parameters and there are none to estimate. The restricted estimates maximise

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

where λ is a $m \times 1$ vector of Lagrange Multipliers. The solution to

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

gives the restricted estimator θ^*, we can write this

$$S(\theta^*) - F(\theta^*)\lambda^* = 0$$

where $S(\theta^*)$ is the $k \times 1$ Score vector evaluated at the restricted estimates and $F(\theta^*)$ is the $k \times m$ matrix of the derivatives of the restrictions with respect to the parameters evaluated at the restricted estimates. Notice that at θ^* the derivative of the Log-likelihood function with respect to the parameters is not equal to zero but to $F(\theta^*)\lambda^*$. The value of the Log-likelihood at θ^* is $LL(\theta^*)$ which is less than or equal to $LL(\widehat{\theta})$.

If the hypotheses (restrictions) are true:

1. the two log-likelihoods should be similar, i.e. $LL(\widehat{\theta}) - LL(\theta^*)$ should be close to zero;

2. the unrestricted estimates that satisfy the restrictions $R(\widehat{\theta})$ should be close to zero (note $R(\theta^*)$ is exactly zero by construction);

3. 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.

   ILLUSTRATE WITH DIAGRAM FROM KENNEDY

These implications are used as the basis for three types of test procedures. The issue is how to judge 'close to zero'? To judge this we use the asymptotic equivalents of the linear distributional results used above in the discussion of the properties of the Linear Regression Model.

Asymptotically the ML estimator is normal

$$\widehat{\theta} \sim N(\theta, I(\theta)^{-1})$$

asymptotically the scalar quadratic form is chi-squared

$$(\widehat{\theta} - \theta)' I(\theta)(\widehat{\theta} - \theta) \sim \chi^2(k).$$

and asymptotically $R(\widehat{\theta})$ is also normal

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

This gives us three procedures for generating asymptotic test statistics for the m restictions $H_0 : R(\theta) = 0$; each of which are distributed $\chi^2(m)$, when the null hypothesis is true:

1. Likelihood Ratio Tests

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

2. Wald Tests

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

3. Lagrange Multiplier (or Efficient Score) Tests

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

- The Likelihood ratio test is straightforward to calculate when both the restricted and unrestricted models have been estimated.

- The Wald test only requires the unrestricted estimates.

- The Lagrange Multiplier test only requires the restricted estimates.
- For the linear regression model, the inequality W>LR>LM holds, so you are more likely to reject using W.
- In the LRM, the LM test is usually calculated using regression residuals.
- The Wald test is not invariant to how you write non-linear restrictions. Suppose $m = 1$, and $R(\theta)$ is $\theta_1\theta_2 - \theta_3 = 0$. This could also be written $\theta_1 - \theta_3/\theta_2 = 0$ and these would give different values of the test statistic. The former form, using multiplication rather than division, is usually better.

These formulas are not used to compute the test. The LR test is usually

$$-2\log\lambda = 2\left[LL(\widehat{\theta}, y) - LL(\theta^*, y)\right] \backsim \chi^2_r$$

while the LM and W are normally calculated from tests on the regression residuals.

1.1 Instrumental Variables

We know that if the regressors are not independent of disturbances then OLS estimates are biased and inconsistent

$$\begin{aligned} y &= X\beta + u \\ &\Longrightarrow \widehat{\beta} = \beta + (X'X)^{-1}X'u \\ &\Longrightarrow p\lim\left(\widehat{\beta}\right) = \beta + p\lim\left(\frac{1}{n}X'X\right)^{-1} p\lim\left(\frac{1}{n}X'u\right) \end{aligned}$$

Assume that

$$p\lim\left(\frac{X'X}{n}\right) = \Sigma_{XX}$$

a positive definite matrix of full rank
and

$$p\lim\left(\frac{X'u}{n}\right) = \Sigma_{Xu} \neq 0$$

then

$$p\lim\left(\widehat{\beta}\right) = \beta + \Sigma_{XX}^{-1}\Sigma_{Xu}$$

So the correlation of the disturbance term with one or more of the regressors will make OLS inconsistent.

Such correlations can be caused by measurement error in one or more regressors, but there are other possibilities:

- lagged dependent variables
- autoregressive disturbance
- simultaneity

Can get consistent estimator by instrumental variables. Consider:

$$y = X\beta + u$$

with

$$var(u) = \sigma^2 I$$

but

$$p\lim\left(\frac{X'u}{n}\right) \neq 0$$

Suppose we can find a data matrix Z of order nxl where $l > k$

1. variables in Z correlated with variables in X

$$p\lim\left(\frac{Z'X}{n}\right) = \Sigma_{ZX}$$

   a finite matrix of full rank

Variables in Z are in the limit uncorrelated with u

$$p\lim\left(\frac{Z'u}{n}\right) = 0$$

then take

$$y = X\beta + u$$

premultiply both sides

$$\begin{aligned} Z'y &= Z'X\beta + Z'u \\ var(Z'u) &= \sigma^2 Z'Z \end{aligned}$$

If we use GLS

$$\begin{aligned} \widehat{\beta}_{GLS} &= \widehat{\beta}_{IV} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y \\ &= (X'PX)^{-1}X'Py \\ \text{with } P &= Z(Z'Z)^{-1}Z' \end{aligned}$$

then

$$Var\left(\widehat{\beta}_{IV}\right) = \sigma^2(X'PX)^{-1}$$

$$\widehat{\sigma} = (y - \widehat{\beta}_{IV})'(y - \widehat{\beta}_{IV})/n$$

using $n-k$ or n does not matter asymptotically).

$$\widehat{\beta}_{IV} = \beta + \left(\frac{1}{n}X'PX\right)^{-1}\left(\frac{1}{n}X'Pu\right)$$

Now

$$\frac{1}{n}X'PX = \left(\frac{1}{n}X'Z\right)\left(\frac{1}{n}Z'Z\right)^{-1}\left(\frac{1}{n}Z'X\right)$$

Assume middle term has plim Σ_{ZZ}^{-1} so

$$p\lim\left(\frac{1}{n}X'PX\right) = \Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}$$

which will be finite non singular matrix.
Similarly

$$p\lim\left(\frac{1}{n}Z'Pu\right) = \Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{Zu} = 0$$

as instruments z are assumed uncorrelated with u in the limit. So the IV estimator is consistent

If $l = k$ so that Z has the same number of columns as X then $X'Z$ is kxk and nonsingular. Then all simplifies to

$$\begin{aligned}\widehat{\beta}_{IV} &= (Z'X)^{-1}Z'y \\ var(\widehat{\beta}_{IV}) &= \sigma^2(Z'X)^{-1}(Z'Z)(X'Z)^{-1}\end{aligned}$$

if $l < k$ then singular

Two Stage Least Squares A form of IV can be seen as result of double application of OLS

1. Regress each of the variables in the X matrix on Z to give $\widehat{X}$

$$\begin{aligned}\widehat{X} &= Z(Z'Z)^{-1}Z'X \\ &= PX\end{aligned}$$

2. Regress y on $\widehat{X}$ to obtain β

$$\begin{aligned}\widehat{\beta}_{TSLS} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'y \\ &= (X'PX)^{-1}(X'Py) \\ &= \widehat{\beta}_{IV}\end{aligned}$$

As before

$$\begin{aligned}var(\widehat{\beta}_{IV}) &= \sigma^2(X'PX)^{-1} \\ \widehat{\sigma}^2 &= (y - X\widehat{\beta}_{IV})'(y - X\widehat{\beta}_{IV})/n\end{aligned}$$

Choice of instruments

- can use variables from X
- any variable thought exogenous and independent of disturbances
- lagged values

When some of the X are used need to partition:

$$\begin{aligned} X &= [X_1\ X_2] \text{ and } Z = [X_1\ Z_1] \\ & X_1 \text{ is } nxr \\ X_2 \text{ is } nx(k-r) \text{ with } r &< k \\ & Z_1 \text{ is } nx(p-r) \end{aligned}$$

can show

$$\widehat{X} = \left[X_1\ \widehat{X}_2\right]$$

where

$$\widehat{X}_2 = Z(Z'Z)^{-1}Z'X_2$$

So variables in X_1 serve as instruments for themselves and the remaining second stage regressors are the fitted values of X_2 obtained form regressing X_2 on the full set of instruments Z

Minimum number of instruments is k including any variables that serve as their own instruments.

Asymptotic efficiency increases with the number of instruments, but so does finite sample bias

if select n instruments then $P = I$ and get the biased and inconsistent OLS estimates

1.2 Multicollinearity

Have assumed in the past that the explanatory variables were linearly independent. This means that $(X'X)^{-1}$ exists.

Perfect multicollineairty implies that $(X'X)$ will be a singular matrix with rank less than k. Which implies do not have unique solutions to the normal equations

As

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

then if $(X'X)$ is singular $\widehat{\beta}$ cannot be estimated. In fact it means that not all regression parameters (the whole vector of parameters) are estimable, only certain linear functions of $\widehat{\beta}_i$.