1 Lecture 2: General Linear Model

Last semester did econometirc only using scalars. Now we have gone through some matrix algebra we can rewrite in matrices. This makes many of the results easy to derive, is more general, is the exposition used in article and books, it is how the estimators and tests etc are computed.

Consider the simple model

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

i=1,...,n.

We can rewrite this as:

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} x_{11} \\ x_{12} \\ . \\ . \\ x_{1n} \end{bmatrix} + \beta_2 \begin{bmatrix} x_{21} \\ x_{22} \\ . \\ . \\ x_{2n} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_n \end{bmatrix}$$

or

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & . & . \\ . & . & . \\ . & . & . \\ 1 & x_{1n} & x_{2n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ . \\ . \\ . \\ u_n \end{bmatrix}$$

which we write as:

$$y = X\beta + u$$

y is nx1 X is nxk β is kx1u is nx1

We want estimates of β and u

$$y = X\widehat{\beta} + \widehat{u}$$

Consider the Maximum likelihood estimator: We need a probability model to give: $f(y)$; $E(y) = X\beta$ and $var(y) = E(\mathrm{u}\, u')$

Note that the variance of y is an nxn variance covariance matrix with variances on the diagonal.

$$E(\mathrm{u}\, u') = E \begin{bmatrix} u_1^2 & . & . & u_1 u_n \\ . & . & . & . \\ . & . & . & . \\ u_n u_1 & . & . & u_n^2 \end{bmatrix}$$

simplest case is the one used before with constant variane and zero covariance

$$E(\mathrm{u}\, u') = E \begin{bmatrix} \sigma^2 & 0 & . & 0 \\ . & . & . & . \\ . & . & \sigma^2 & . \\ 0 & .0 & . & \sigma^2 \end{bmatrix} = \sigma^2 I$$

Assuming y and so u are normally distributed:

$$y \sim N(X\beta, \sigma^2 I) \Longrightarrow E(y) = X\beta$$
$$u \sim N(0, \sigma^2 I) \Longrightarrow E(u) = E(y - E(y)) = 0$$
$$L = \Pi_i^n f(u_i)$$

since independent errors -because of assumption of zero covaraice and normality

$$L = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{-u'u}{2\sigma^2}\right)$$
$$L = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{-(y - X\beta)'(y - X\beta)}{2\sigma^2}\right)$$

Maximise L with respect to β involves maximising $-(y - X\beta)'(y - X\beta)$ which is the same as minimising $(y - X\beta)'(y - X\beta)$ which is minimising the sum of squared residuals, the same as OLS.

$$\begin{aligned} RSS &= \mathrm{u}\,u' = (y - X\beta)'(y - X\beta) \\ &= y'y - \beta'X'y - y'X\beta + \beta'X'X\beta \\ &= y'y - 2\beta'X'y - +\beta'X'X\beta \end{aligned}$$

Now transpose of a scalar is a scalar so

$$y'X\beta = (y'X\beta)' = \beta'X'y$$

so

$$\frac{\delta RSS}{\delta\beta} = -2X'y + 2X'X\beta = 0$$

and the normal equations are then

$$(X'X)\beta = X'y$$

with

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

In the two variable case:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix}$$

$$X'X = \begin{bmatrix} 1 & 1 & . & . & 1 \\ x_1 & x_2 & . & . & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$$

and

$$X'y = \begin{bmatrix} 1 & 1 & . & . & 1 \\ x_1 & x_2 & . & . & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

giving

$$\begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

which gives the more familiar normal equations

$$\begin{aligned} n\beta_1 + \beta_2 \sum x &= \sum y \\ \beta_1 \sum x + \beta_2 \sum x^2 &= \sum xy \end{aligned}$$

So $y \sim N(X\beta, \sigma^2 I)$, X non stochastic and $(X'X)^{-1}$ non singular means $\widehat{\beta}_{ML} = (X'X)^{-1}X'y = \widehat{\beta}_{OLS}$, but only if the assumptions made hold.

$\widehat{\beta}_{ML}$ as we know has properties of consistency and asymptotic efficiency.

Can show as in scalar case that $\widehat{\beta}_{OLS}$ is in fact BLUE -it is the minimum variance estimator in the class of linear unbiased estimators. NB dont need normality for this.

$$\widehat{y} = X\widehat{\beta} + u$$

$$\begin{aligned} E(u) &= 0 \\ E(\mathrm{u}\,u') &= \sigma^2 I \\ & \quad X \textit{ is non stochastic} \\ E(X'u) &= X'E(u) = 0 \\ Rank(X) &= k \Longrightarrow (X'X)^{-1} exists \end{aligned}$$

1. Clearly $\widehat{\beta}$ is linear

2. Simple to show unbiased

$$\begin{aligned} \widehat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + u) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u \end{aligned}$$

$$\begin{aligned} \Longrightarrow \quad & E(\widehat{\beta}) = \beta + E\left[(X'X)^{-1}X'u\right] \\ = \quad & \beta + (X'X)^{-1}X'E(u) \\ = \quad & \beta \end{aligned}$$

3.Variance:

$$Var(\widehat{\beta}) = E(\widehat{\beta} - E(\widehat{\beta}))(\widehat{\beta} - E(\widehat{\beta}))'$$

a kxk variance covariance matrix.
as $E(\widehat{\beta}) = \beta$

$$Var(\widehat{\beta}) = E(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)'$$

From before

$$\begin{aligned} \widehat{\beta} &= \beta + (X'X)^{-1}X'u \\ so\ \widehat{\beta} - \beta &= (X'X)^{-1}X'u \end{aligned}$$

so

$$\begin{aligned} Var(\widehat{\beta}) &= E\left[(X'X)^{-1}X'\,\mathrm{u}\,u'X(X'X)^{-1}\right] \\ and\ as\ E(X'u) &= X'E(u) \\ Var(\widehat{\beta}) &= (X'X)^{-1}X'E(\mathrm{u}\,u')X(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

4. Consider $\widetilde{\beta} = Wy$ where W is a kxn matrix

$$\begin{aligned} \widetilde{\beta} &= \left[(X'X)^{-1}X' + c\right] y \\ where\ c &= W - (X'X)^{-1}X' \end{aligned}$$

Now

$$\begin{aligned} \widetilde{\beta} &= (X'X)^{-1}X'y + cy \\ &= \widehat{\beta} + cy \\ &= \widehat{\beta} + c(X\beta + u) \\ &= \widehat{\beta} + cX\beta + cu \\ &= \beta + (X'X)^{-1}X'u + cX\beta + cu \\ &= \beta + cX\beta + [(X'X)^{-1}X' + c]u \end{aligned}$$

which means

$$\begin{aligned} E(\widetilde{\beta}) &= \beta + cX\beta + [(X'X)^{-1}X' + c]E(u) \\ &= \beta + E(\widetilde{\beta}) = \beta \end{aligned}$$

so $E(\widetilde{\beta}) = \beta$ only if $cX\beta = 0$

Now

$$\begin{aligned}
var(\widetilde{\beta}) &= E(\widetilde{\beta} - \beta)(\widetilde{\beta} - \beta)' \\
&= E[((X'X)^{-1}X' + c)u][((X'X)^{-1}X' + c)u]' \\
&= E[(X'X)^{-1}X' + c]\,\mathrm{u}\,u'[(X'X)^{-1}X' + c]' \\
&= \sigma^2[(X'X)^{-1}X' + c][(X'X)^{-1}X' + c]' \\
&= \sigma^2[(X'X)^{-1}X'X(X'X)^{-1} + cc' + (X'X)^{-1}X'c' + cX(X'X)^{-1}]
\end{aligned}$$

now as $cX = 0 = X'c'$

$$var(\widetilde{\beta}) = \sigma^2[(X'X)^{-1} + cc']$$

so if we look at the differences between the variances

$$\begin{aligned}
var(\widetilde{\beta}) - var(\widehat{\beta}) &= \sigma^2[(X'X)^{-1} + cc'] - \sigma^2[(X'X)^{-1} \\
&= \sigma^2 cc'
\end{aligned}$$

Now $cc'$is positive semi definite -in particular its diagonal is non negative as it is squared terms. So the variance of each $\widetilde{\beta}$ must be greater than or equal to the variance of $\widehat{\beta}$. So $\widehat{\beta}$ is the minimum variance estimator.

If also assume normality then $\widehat{\beta}$ can be shown to be minimum variance estimator among all unbised estimators, not just linear ones.

Relation between $\widehat{u}$ and v

$$\begin{aligned}
\widehat{u} &= y - X\widehat{\beta} \\
&= X\beta + u - X\widehat{\beta} \\
&= X\left(\beta - \widehat{\beta}\right) + u \\
&= u - X(X'X)^{-1}X'u \\
&= \left[I_n - X(X'X)^{-1}X'\right]u
\end{aligned}$$

which we write as Nu

Now $N = \left[I_n - X(X'X)^{-1}X'\right]$ maps the disturbance into the estimated residuals and is an nxn symmetric idempotent matrix so $N^2 = N$ and

$$\begin{aligned}
Rank(N) &= Trace(I_n) - Trace(X(X'X)^{-1}X') \\
&= Trace(I_n) - Trace((X'X)^{-1}X'X) \\
&= n - k
\end{aligned}$$

Thus N is singular as $n - k < n$

$M = X(X'X)^{-1}X'$ is also idempotent $N = [I_n - M]$and $MN = 0$ and

$$\begin{aligned}
\widehat{u}'\widehat{u} &= u'N'Nu = u'Nu \textit{ a scalar} \\
E(\widehat{u}'\widehat{u}) &= E(u'Nu\ ) \\
&= E(Trace(N\,\mathrm{u}\,u') \\
&= Trace(N)E(\mathrm{u}\,u') \\
&= TraceNI\sigma^2 \\
&= (n-k)\sigma^2
\end{aligned}$$

Thus

$$E\left(\frac{\widehat{u}'\widehat{u}}{n-k}\right) = E(\widehat{\sigma}^2) = \sigma^2$$

Note that the estimated residuals are orthogonal to X

$$\begin{aligned}
X'\widehat{u} &= X'(y - X\widehat{\beta}) \\
&= X'y - X'X\widehat{\beta} \\
&= X'y - X'X(X'X)^{-1}X'y \\
&= X'y - X'y \\
&= 0
\end{aligned}$$

So or least squares the $\widehat{u}$ are completely uncorrelated with the explanatory variables

$$\begin{aligned}
\sum \widehat{u}_i &= 0 \\
\sum X_i\widehat{u}_i &= 0
\end{aligned}$$

For two variables, the constant and $x_1$this means we have 2 restrictions on the estimated residuals. So if we know n-2 residuals we can work out the last 2. Generally if we know n-k we can derive the other k. $\widehat{u}'\widehat{u}$ is a singular matrix.

To summarise: We have a variety of assumptions

1. $E(u) = 0$
2. a.) $E(\mathrm{u}\,u') = \sigma^2 I$ which means $E(u_i^2) = \sigma^2$ and $E(u_iu_j) = 0$ for $i \neq j$ or the weaker alternative
   b.) $E(\mathrm{u}\,u') = \sigma^2\Omega$ where Ω is positive definite
3. a.) X non stochastic, independent of u $\Longrightarrow E(X'u) = 0$
   b.) X stochastic but not correlated with $u \Longrightarrow$

$$\underset{n\rightarrow\infty}{p\lim}\ \frac{X'u}{n} = 0$$

4. $Rank(X) = k \Longrightarrow (X'X)^{-1} exists$
5. u is normally distributed

Gauss Markov requires	1, 2a, 3a, 4
Maximum Likelihood/OLS requires	1, 2a, 4, 5
Hypothesis testing also need 5	

1.1 What happens when assumptions fail.

1. If X is not of full rank k, because there is an exact linear dependency between some of the variables, then the ML estimates of β are not defined and there is said to be exact multicollinearity. $X'X$ is singular and not invertable. The model should be respecified to remove the exact dependency.

   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.

2. If the X are not strictly exogenous the estimates of β are biased, though if the X are predetermined (e.g. lagged dependent variables) and the disturbance term is not serially correlated, they will remain consistent. Otherwise, they will be inconsistent.

   In certain circumstances failure of the exogeneity assumptions can be dealt with by the method of Instrumental Variables.

3. If normality does not hold and the form of the distribution is not known 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, the tests below will not have the stated distributions, though asymptotically they will be normal. If the form of the distribution is known (e.g. a t distribution) maximum likelihood estimators can be derived for the particular distribution.

4. If $y \sim N(X\beta, \sigma^2\Omega)$, that is its variance covariance matrix is not $\sigma^2 I$, there are two possible problems:

   - – the variances (diagonal terms of the matrix) are not constant and equal to σ^2 (heteroskedasticity)
     – and/or the off diagonal terms, the covariances, are not equal to zero (failure of independence, serial correlation, autocorrelation).
     – Under these circumstances, $\widehat{\beta}$ remains unbiased but is not minimum variance (efficient). Its variance-covariance matrix is not $\sigma^2(X'X)^{-1}$, but $\sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$. Corrected variance-covariance matrices are available in most packages (White Heteroskedasticity consistent covariance matrices or Newey-West autocorrelation consistent ones).
     – Notice that residual serial correlation or heteroskedasticity may indicate not that there is some covariances between the true disturbances but that the model is wrongly specified, e.g. variables are omitted, see below.

– When it is appropriate to model the disturbance structure in terms of Ω, Generalised Least Squares can be used. In most cases, residual serial correlation or heteroskedasticity should lead you to respecify the model not use Generalised Least Squares.

5. Omitted variables. Suppose the data are generated by

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

and you omit z_t, an hx1 vector and estimate

$$y_t = \alpha' x_t + v_t. \tag{2}$$

What is the relationship between the estimates? Suppose we describe the relation between the omitted and included right hand side variables by the multivariate regression model:

$$z_t = Bx_t + w_t \tag{3}$$

where B is an hxk matrix. This is just a set of h regressions in which each z_t is regressed on all k x_t. If you replace z_t above you get:

$$y_t = \beta' x_t + \gamma'(Bx_t + w_t) + u_t$$
$$y_t = (\beta' + \gamma' B)x_t + (\gamma' w_t + u_t).$$

Thus $\alpha = (\beta' + \gamma' B)$ and $v_t = (\gamma' w_t + u_t)$. The coefficient of x_t in will only be an unbiased estimator of β, the coefficient of x_t in if either $\gamma = 0$ (z_t really has no effect on y_t) or $B = 0$, (there is no correlation between the included and omitted variables). Notice that v_t also contains the part of z_t that is not correlated with x_t, w_t, and 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.

1.2 Generalised Least Squares

$$\begin{aligned} y &= X\beta + u \\ E(\mathrm{u}\,u') &= \Omega \end{aligned}$$

Ω is an arbitrary positive definite matrix
Premultiply by $\Omega^{-1/2}$

$$\begin{aligned} y^* &= X^*\beta + u^* \\ y^* &= \Omega^{-1/2} y \\ x^* &= \Omega^{-1/2} X \\ u^* &= \Omega^{-1/2} u \\ E(u^* u^{*\prime}) &= \Omega^{-1/2} E(\mathrm{u}\,u')\Omega^{-1/2} = I \end{aligned}$$

So BLUE of β is

$$\begin{aligned}\beta_{GLS} &= (X^{*\prime}X^*)^{-1}(X^{*^{\prime}}y^*) \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y\end{aligned}$$

which is the generalised least squares estimator. A feasible GLS estimator results from replacing Ω with a consistent estimator.

Now

$$var(\widehat{\beta}_{GLS}) = \sigma^2(X^{*\prime}X^*)^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}$$

while variance for OLS is

$$var(\widehat{\beta}_{OLS}) = \sigma^2(X'X)^{-1}$$

If we were to estimate OLS when GLS is appropriate $E(u) = 0$ will still be true so the estimator is unbiased, but its variance is now

$$\begin{aligned}var(\widehat{\beta}_{OLS}) &= (X'X)^{-1}X'E(\mathrm{u}\,u')X'(X'X) - 1 \\ &= (X'X)^{-1}X'\Omega X(X'X)^{-1}\end{aligned}$$

Generally this means it is inefficient, though it is not necessarily the case.