Econometrics, Lecture 2
The Linear Regression Model, LRM, in matrix notation

Ron Smith
EMS, Birkbeck, University of London

Autumn 2020

Last time

- Considered the bivariate LRM, and its estimates

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_t + u_t \\ Y_t &= \widehat{\beta}_1 + \widehat{\beta}_2 X_t + \widehat{u}_t \end{aligned}$$

- Made assumptions, $E(u_t) = 0$, $E(u_t^2) = \sigma^2$, $E(u_t u_{t-i}) = 0$; X_t varied and independent of u_t.
- Got estimators of the coefficients, β_j, $j = 1, 2$, by method of moments and least squares to get,

$$\begin{aligned} \widehat{\beta}_1 &= \overline{Y} - \widehat{\beta}_2 \overline{X} \\ \widehat{\beta}_2 &= \frac{\sum_t x_t y_t}{\sum_t x_t^2} \end{aligned}$$

  where $y_t = Y_t - \overline{Y}$ and $x_t = X_t - \overline{X}$
- showed $\widehat{\beta}_j$ are unbiased, derived $V(\widehat{\beta}_2) = \sigma^2 / \sum_t x_t^2$ and standard errors.
- Now do the same but in matrix algebra to handle multiple regression.

Multiple regression

with k explanatory variables takes the form

$$Y_t = \beta_1 + \beta_2 X_{2t} + ... + \beta_k X_{kt} + u_t$$

where $X_{1t} = 1$ all t. This can be written in vector form as:

$$\underset{1\times 1}{Y_t} = \underset{1\times k}{\beta'}\underset{k\times 1}{X_t} + \underset{1\times 1}{u_t}$$

where β and X_t are $k \times 1$ vectors. Or in matrix form as

$$\underset{T\times 1}{y} = \underset{T\times k}{X}\underset{k\times 1}{\beta} + \underset{T\times 1}{u}$$

where y and u are $T \times 1$ vectors and X is a $T \times k$ matrix.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ .. \\ Y_T \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & .. & X_{k1} \\ 1 & X_{22} & .. & X_{k2} \\ .. & .. & .. & .. \\ 1 & X_{2T} & .. & X_{kT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ .. \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ .. \\ u_T \end{bmatrix}.$$

Matrix Algebra

- Transpose switches rows to columns. If β is 2×1, vectors are column vectors, β' is a row vector, 1x2.
- Try to write the dimensions underneath and use examples

$$\underset{1\text{x}2}{\beta'}\underset{2\text{x}1}{X_t} = \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix} \begin{bmatrix} 1 \\ X_{2t} \end{bmatrix}$$

- Must be conformable, inner dimensions must match.
- If A is a $n \times m$ matrix, and B is an $m \times k$ matrix the transpose of the product $(AB)'$ is $B'A'$

$$\begin{matrix} (A & B)' \\ (n \times m & m \times k)' \end{matrix} = \begin{matrix} B' & A' \\ (k \times m & m \times n) \end{matrix}$$

  a $k \times n$ matrix the product of a $k \times m$ matrix with a $m \times n$ matrix. $A'B'$ is not conformable.
- I_T is a $T \times T$ identity matrix with $1s$ on the diagonal and zeros on the off diagonals

$$I_2 = \{$$

Sums of squares and var-cov matrices

- Distinguish $\underset{(1\times T)(T\times 1)}{u'u} = \sum u_t^2$ the scalar sum of squared errors and $\underset{(T\times 1)(1\times T)}{uu'}$ a $T \times T$ matrix.
- Suppose u a 2×1 vector.

$$uu' = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} u_1^2 & u_1 u_2 \\ u_2 u_1 & u_2^2 \end{bmatrix}$$

$$u'u = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = u_1^2 + u_2^2$$

Assumptions about the errors

The assumption $E(u_t) = 0$ can be written in matrix notation $E(u) = 0$, while

$$\begin{aligned} E(u_t^2) &= \sigma^2 \\ E(u_t u_{t-i}) &= 0; \; i \neq 0 \end{aligned}$$

can be written

$$\begin{aligned} E(u\ u') &= E \begin{bmatrix} u_1^2 & u_1 u_2 & .. & u_1 u_T \\ u_2 u_1 & u_2^2 & .. & u_2 u_T \\ .. & .. & .. & .. \\ u_T u_1 & u_T u_2 & .. & u_{T^2} \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & .. & 0 \\ 0 & \sigma^2 & .. & 0 \\ .. & .. & .. & .. \\ 0 & 0 & .. & \sigma^2 \end{bmatrix} \\ E(u\ u') &= \sigma^2 I \end{aligned}$$

I is a $T \times T$ identity matrix with 1s on the diagonal and zeros on the off diagonals so $\sigma^2 I_T$ has σ^2 on the diagonal and zeros on the off-diagonals.

Assumptions about the X

- No exact multicollinearity. The matrix of explanatory variables, X is of full rank k. $X'X$ is non singular. These are all equivalent statements. In the bivariate case this implies that the variance of X_t is non-zero).
- The explanatory variables in X are either
  - (a) non stochastic
  - (b) (strictly) exogenous, distributed independently of the errors u_t or
  - (c) pre-determined, uncorrelated with the errors u_t, e.g. lagged dependent variables.
- All three imply that $E(X'u) = 0$; which in the bivariate case is $E(u_t) = 0$ and $E(X_t u_t) = 0$. Exogeneity is discussed in more detail in week 9.
- $\widehat{\beta}$ is unbiased if (a) or (b) hold, biased but consistent if (c) holds.

Estimators

We use 3 procedures.

- Method of moments chooses $\widehat{\beta}$ to make a property assumed to hold in the population hold in the sample.
- Least squares chooses $\widehat{\beta}$ to minimise $\sum \widehat{u}_t^2$.
- Maximum Likelihood, ML, chooses the $\widehat{\beta}$ most likely to have generated the observed sample. ML also requires an additional assumption about the conditional distribution of y (distribution of u).

In the case of the linear regression model with normally distributed errors, the three procedures lead to the same estimator. This is not generally the case.

Method of moments

The MoM estimator finds the estimator $\widehat{\beta}$ that makes the sample equivalent of $E(X'u) = 0$ which is $X'\widehat{u} = 0$, hold.

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

as long as $(X'X)$ is non-singular, which is ensured by the assumption that the rank of $X = k$.
If $(X'X)$ is singular $\widehat{\beta}$ does not exist. You will know when this assumption fails.

Least squares

Finds the $\widehat{\beta}$ that minimizes $\sum \widehat{u}_t^2 = \widehat{u}'\widehat{u}$ in:

$$\underset{T\times 1}{y} = \underset{T\times k}{X}\underset{k\times 1}{\widehat{\beta}} + \underset{T\times 1}{\widehat{u}}$$

$$\begin{aligned}
\widehat{u}'\widehat{u} &= \left(y - X\widehat{\beta}\right)'\left(y - X\widehat{\beta}\right) && (1)\\
&= y'y - \widehat{\beta}'X'y - y'X\widehat{\beta} + \widehat{\beta}'X'X\widehat{\beta} && (2)\\
&= y'y - 2\widehat{\beta}'X'y + \widehat{\beta}'X'X\widehat{\beta}. && (3)
\end{aligned}$$

- If A is a $n \times m$ matrix, and B is an $m \times k$ matrix the transpose of the product $(AB)'$ is $B'A'$ a $k \times n$ matrix the product of a $k \times m$ matrix with a $m \times n$ matrix. $A'B'$ is not conformable.
- $y'X\beta = \beta'X'y$ because both are scalars (1×1 matrices). Scalars are always equal to their transpose.
- The term $\beta'X'X\beta$ is a quadratic form, i.e. of the form $x'Ax$ above.

Least Squares 2

- The $k \times 1$ vector of derivatives is

$$\begin{aligned} \frac{\partial \widehat{u}'\widehat{u}}{\partial \widehat{\beta}} &= -2X'y + 2X'X\widehat{\beta} = 0 \qquad (4) \\ \widehat{\beta} &= (X'X)^{-1}X'y \end{aligned}$$

  as in the method of moments case.
- The second derivative is $2X'X$ which is a positive definite matrix, so this is a minimum.
- Matrix, A, is positive definite if for any a, $a'Aa > 0$. Matrices with the structure $X'X$ are always positive definite, since they can be written as a sum of squares. Define $z = Xa$, then $z'z = a'X'Xa = \sum z_t^2 > 0$.
- Matrices like $X'X$ are symmetric, so $(X'X)' = X'X$

Expected Value

$$
\begin{aligned}
\widehat{\beta} &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) \\
\widehat{\beta} &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\
&= \beta + (X'X)^{-1}X'u
\end{aligned} \tag{5}
$$

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

since β is not a random variable, and if X and u are independent

$$E((X'X)^{-1}X'u) = E((X'X)^{-1}X')E(u) = 0$$

since $E(u) = 0$. Thus $E(\widehat{\beta}) = \beta$ and $\widehat{\beta}$ is an unbiased estimator of β.

Variance Covariance matrix 1

The variance-covariance matrix of $\widehat{\beta}$ is a $k \times k$ matrix

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

since $\widehat{\beta}$ is unbiased. But from (5) we have

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

so

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

since $E(uu') = \sigma^2 I$, σ^2 is a scalar, and $(X'X)^{-1}X'X = I$.

Variance Covariance matrix 2

- Compare this to the scalar case $\sigma^2(\sum x_i^2)^{-1}$.
- We derive the variance covariance matrix conditional on the observed sample, which is why we can take the expected value inside in line 3.
- We estimate $V(\widehat{\beta})$ by

$$\widehat{V(\beta)} = s^2(X'X)^{-1}$$

  where $s^2 = \widehat{u}'\widehat{u}/(T-k)$.
- The square roots of the *ith* diagonal element of $s^2(X'X)^{-1}$ gives the standard errors of $\widehat{\beta}_i$ the *ith* elements of $\widehat{\beta}$, which is reported by computer programs.

Non scalar covariance matrix

- Note that the conditional variance covariance matrix of y is the variance covariance matrix of $u: Var(y \mid X) = E(uu')$ since

$$E(y - E(y \mid X)E(y - E(y \mid X)' = E(uu')$$

- If $Var(y \mid X) = E(uu') = \sigma^2\Omega$, not $\sigma^2 I$, this can arise because 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 $\sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$.
- Use robust standard errors.

Predicted Values and residuals

- Predicted values: $\widehat{y} = X\widehat{\beta}$. Estimated residuals: $\widehat{u} = y - \widehat{y} = y - X\widehat{\beta}$ are uncorrelated with the explanatory variables by construction, we got the $\widehat{\beta}$ that made $X'\widehat{u} = 0$.
- $X'\widehat{u} = 0$ is a set of k equations of the form:

$$\sum_{t=1}^{T} \widehat{u}_t = 0; \sum_{t=1}^{T} x_{2t}\widehat{u}_t = 0; \ldots; \sum_{t=1}^{T} x_{kt}\widehat{u}_t = 0.$$

- The residuals

$$\begin{aligned} \widehat{u} &= y - X\widehat{\beta} = y - X(X'X)^{-1}X'y \\ &= (I - X(X'X)^{-1}X')y = My; \end{aligned} \tag{6}$$

$$\begin{aligned} P_x &= X(X'X)^{-1}X' \\ M &= I - P_x \end{aligned}$$

- P_x is called a projection matrix. Both M and P_x are idempotent, equal to their product $P_x P_x = P_x$ and $MP_x = 0$.

Residuals and errors

- Using (6)

$$\widehat{u} = My = M(X\beta + u) = MX\beta + Mu = Mu,$$

since $MX\beta = (I - P_x)X\beta = X\beta - X(X'X)^{-1}X'X\beta = X\beta - X\beta = 0.$

- The dependent variable is split into two orthogonal components, the projection of y on X and the orthogonal remainder.

$$y = \widehat{y} + \widehat{u} = P_x y + My$$

- Notice that while the estimated residuals are a transformation of the true disturbances, $\widehat{u} = Mu$,we cannot recover the true disturbances from this equation since M is singular, rank T-k.

Estimating the variance

The sum of squared residuals is:

$$\sum_{t=1}^{T} \widehat{u}_t^2 = \widehat{u}'\widehat{u} = u'M'Mu = u'Mu.$$

To calculate the expected value of the sum of squared residuals, note that $\widehat{u}'\widehat{u}$ is a scalar, thus equal to its trace, the sum of its diagonal elements. Thus using the properties of traces we can write

$$\begin{aligned} E(\widehat{u}'\widehat{u}) &= E(u'Mu) = E(tr(u'Mu)) = E(tr(Muu')) \\ &= tr(M\sigma^2 I) = \sigma^2 tr(M) = \sigma^2(T-k). \end{aligned}$$

Thus the unbiased estimate of σ^2 is $s^2 = \widehat{u}'\widehat{u}/(T-k)$. The last step uses the fact that the Trace of M is

$$\begin{aligned} tr\left[I_T - X(X'X)^{-1}X'\right] &= tr(I_T) - tr\left(X(X'X)^{-1}X'\right) \\ &= tr(I_T) - tr\left((X'X)^{-1}X'X\right) \\ &= tr(I_T) - tr(I_K) = T - k \end{aligned}$$

Goodness of fit

The standard error of regression, SER, $s = \sqrt{\widehat{u}'\widehat{u}/(T-k)}$, is a measure of the average size of the errors. Other measures that are used are

$$R^2 = 1 - \frac{\sum \widehat{u}_t^2}{\sum (Y_t - \overline{Y})^2}$$

which increases if you add another variable and the adjusted R^2

$$\overline{R}^2 = 1 - \frac{\sum \widehat{u}_t^2/(T-k)}{\sum (Y_t - \overline{Y})^2/(T-1)}$$

which increases if the SER reduces, which it will if you add a variable with a t ratio greater than one.

Gauss-Markov Theorem

- This shows that among the class of linear, unbiased estimator $\widehat{\beta}$ has the smallest variance, if

$$\begin{aligned} E(u) &= 0, \\ E(u\ u') &= \sigma^2 I_T, \end{aligned}$$

  X is of rank k and exogenous, independent of u.
- Notice we do not assume normality.
- This is sometimes expressed as $\widehat{\beta}$ is the best (minimum variance) linear unbiased estimator, BLUE. There may be non-linear or biased estimators with smaller variance.
- There are a number of different ways to prove this.It is a tutorial class exercise.

Next time

- So far we have not assumed anything about distributions.
- Will next look at a different approach to regression as a conditional expectation got from conditioning a joint distribution
- If the random variables are joint normal the conditional expectation is linear.
- If $z \sim N(0,1)$

$$f(z) = (2\pi)^{-1/2} \exp\left(-\frac{z^2}{2}\right)$$

- We will look at maximum likelihood the 3rd of our procedures for finding estimators and other distributions related to the normal.