ECON 61001: *Lecture 1*

Alastair R. Hall

The University of Manchester

Outline of this lecture

Linear regression model

- Model
- Assumptions

- Ordinary Least Squares
- Sum of squares decomposition

- Interpretation of OLS coefficients - the Frisch-Waugh-Lovell Theorem

- asset returns: $R - R_f = \beta_0(R_m - R_f) + error$

- returns to education:
  $ln(w) = \beta_{0,1} + \beta_{0,2} * ed + \beta_{0,3} * exp + \beta_{0,4} * exp^2 + error$

- aggregate production function:
  $ln(Q) = \beta_{0,1} + \beta_{0,2} * ln(L) + \beta_{0,3} * ln(K) + error$

- change in inflation: $\Delta inf = \beta_{0,1} + \beta_{0,2} * \Delta inf(-1) + error$

All have common structure: linear in the parameters, and additive error

Data types and notation

Economic data typically comes in four types:

- Cross-section - covered in course
- Time series - covered in course
- Panel data
- Repeated cross-section

Notation for sample:

- Cross-section: $i = 1, 2, \ldots, N$
- Times series: $t = 1, 2, \ldots, T$

For first part of course, results apply equally to both types of data and use (default) of t notation.

When discuss large sample properties arguments are different and will i or t notation as reminder of sample structure.

Linear regression model

Wish to model relationship between y_t (“dependent variable”) and $k \times 1$ vector x_t (“explanatory variables”).

Assume:

$$y_t = x_t'\beta_0 + u_t$$

and

- y_t, x_t are observable but the error term u_t is not.
- β_0 is an unknown $k \times 1$ vector of “regression coefficients” (parameters).

Observe $\{y_t, x_t; t = 1, 2 \ldots T\} \rightarrow$ estimate of β_0.

More convenient to express model in matrix notation:

$$y = X\beta_0 + u$$

where

- y is $T \times 1$ with t^{th} element y_t
- X is $T \times k$ with t^{th} row x_t'
- u is $T \times 1$ with t^{th} element u_t

Classical assumptions

- *CA1*: true model is: $y = X\beta_0 + u$.
- *CA2*: X is *fixed in repeated samples*.
- *CA3*: X is rank k.
- *CA4*: $E[u] = 0$.
- *CA5*: $Var[u] = \sigma_0^2 I_T$.
- *CA6*: $u \sim$ Normal.

Implications for y:

- $y \sim N(X\beta_0, \sigma_0^2 I_T)$.

Estimation problem

Consider here estimation of β_0 based on sample (y, X) using Ordinary Least Squares (OLS).

Define $u(\beta) = y - X\beta$ and note $u(\cdot) : \mathcal{S} \times \mathcal{B} \to \mathbb{R}^T$.

OLS minimand is:

$$Q_T(\beta) \;=\; u(\beta)'u(\beta) \;=\; \sum_{t=1}^{T}(y_t - x_t'\beta)^2$$

(Note: $Q_T : \mathcal{S} \times \mathcal{B} \to [0, \infty)$.)

OLS estimator of β_0 is:

$$\hat{\beta}_T \;=\; argmin_{\beta \in \mathcal{B}} Q_T(\beta)$$

First order conditions (FOC):

$$\frac{\partial Q_T(\beta)}{\partial \beta}\Big|_{\beta=\hat{\beta}_T} = 0$$

Second order conditions (SOC):

$$\frac{\partial^2 Q_T(\beta)}{\partial \beta \partial \beta'}\Big|_{\beta=\hat{\beta}_T} = \text{positive definite (p.d.)}$$

We have

$$Q_T(\beta) \;=\; y'y \;-\; 2y'X\beta \;+\; \beta'X'X\beta$$

and so from Lemma 2.2(i)-(ii)[LN p.12]:

$$\frac{\partial Q_T(\beta)}{\partial \beta} \;=\; -2X'y \;+\; 2X'X\beta$$

and from Lemma 2.2(iii)

$$\frac{\partial^2 Q_T(\beta)}{\partial \beta \partial \beta'} \;=\; 2X'X$$

FOC $\rightarrow X'(y \;-\; X\hat{\beta}_T) \;=\; 0$ and so using *CA3*,

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

and (using *CA3*) SOC satisfied.

Model involves decomposition of y:

$$y = E[y] + u$$

OLS affects a similar decomposition:

$$y = \hat{y} + e$$

where

- $\hat{y} = X\hat{\beta}_T$, vector of predicted values for y.
- $e = y - X\hat{\beta}_T$, vector of OLS residuals.

Note FOC $\Rightarrow X'e = 0$ and so

$$\hat{y}'e = 0$$

OLS affects a similar decomposition of the variation of y in models that include an intercept. So now set: $X = [\iota_T, X_2]$.

The decomposition of the variation of y is as follows:

$$TSS \;=\; ESS \;+\; RSS$$

where:

- TSS = "Total sum of squares" $= \sum_{t=1}^{T}(y_t - \bar{y})^2$
- ESS = "Explained sum of squares" $= \sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2$
- RSS = "Residual sum of squares" $= \sum_{t=1}^{T} e_t^2 = e'e$.

This leads to the multiple correlation coefficient, R^2:

$$R^2 \;=\; \frac{ESS}{TSS}$$

which is proportion of variation in y explained by linear regression on X.

We now develop a useful interpretation of OLS coefficients based on the Frisch-Waugh-Lovell (FWL) Theorem.

To present the FWL Thm we need to partition the regressors and coefficient vector conformably:

$$\underset{(T\times k)}{X} = \left(\underset{(T\times k_1)}{X_1}, \underset{(T\times k_2)}{X_2} \right), \qquad \beta_0 = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \begin{matrix} (k_1 \times 1) \\ (k_2 \times 1) \end{matrix},$$

and write model as

$$y = X\beta_0 + u = X_1\beta_1 + X_2\beta_2 + u$$

Let $\hat{\beta}_{T,2}$ be the OLS estimator of β_2 in this model.

Now consider the alternative strategy for estimation of β_2 (here $x_{2,\ell}$ denotes the ℓ^{th} column of X_2).

Step 1: Regress y on X_1 via OLS and denote the associated vector of OLS residuals by w.

Step 2: For each $\ell = 1, 2, \ldots, k_2$, regress $x_{2,\ell}$ on X_1 via OLS and denote the associated vector of OLS residuals by d_ℓ.

Step 3: Regress w on D, where $D = (d_1, d_2, \ldots, d_{k_2})$, via OLS and denote the resulting vector of coefficient estimators by $\hat{b}$ that is, $\hat{b} = (D'D)^{-1}D'w$.

FWL Theorem: $\hat{\beta}_{T,2} = \hat{b}$.

Implications of FWL:

- Consider case where $X_1 = [x_1, x_2, \ldots, x_{k-1}]$, and $X_2 = x_k$.

- w and $D = d$ represent the parts of y and x_k that cannot be linearly explained by X_1.

- *Step 3* captures the relationship between y and x_k once they have both been purged of any linear dependence they have on X_1.

FWL Thm

- $\hat{b} = 0 \Rightarrow$ any relationship between y and x_k can be accounted for by their joint dependence on X_1.

- $\hat{b} \neq 0 \Rightarrow y$ and x_k are linearly related in a way that cannot be explained purely by their joint dependence on X_1.

- $\hat{\beta}_k$ captures *partial effect* = the unique contribution (relative to the other regressors in the model) of x_k to the (linear) explanation of y.

Terminology:

- *Steps 1* and *2* are often referred to as *"partialling out"* the effect of X_1.

- The regression in *Step 3* is said to capture the relationship between y and x_k *controlling* for X_1.

Example

In mid-1980's there were two changes to federal highway regulations in US.

- Jan 1986: seat belt law passed
- May 1987: states allowed to raise highway speed limit from 55mph to 65mph

McCarthy (1994) investigates whether these changes affected the number of traffic fatalities in California.

Analysis uses: monthly data, Jan 1981 - Dec 1989.

Example

Dependent variable, y_t: % of highway accidents that resulted in one or more fatality.

Explanatory variables:

- $belt_t$: dummy variable equal to 1 for $t \geq 1986.1$
- mph_t: dummy variable equal to 1 for $t \geq 1987.5$

Example

- $\hat{y}_t = 0.914 - 0.064 * belt_t$
- $\hat{y}_t = 0.893 - 0.024 * mph_t$
- $\hat{y}_t = 0.914 - 0.102 * belt_t + 0.057 * mph_t$

Now introduce controls:

- Linear time trend and monthly dummies:

$$\hat{y}_t = \text{controls} - 0.014 * belt_t + 0.078 * mph_t$$

- Plus state unemployment rate and number of weekends in month:

$$\hat{y}_t = \text{controls} - 0.030 * belt_t + 0.0671 * mph_t$$

However..

Sometimes the inclusion of controls can undermine the inference of interest.

Example: impact of climate variable, C_i, on economic activity, y_i, based on cross-sectional country data

$$y_i \;=\; \alpha + \gamma C_i + error,$$

Suppose include controls (institutional measures, population *etc*)

$$y_i = \alpha + \gamma C_i + z_i'\delta + error,$$

If controls depend on climate then their inclusion masks the impact of climate on economic activity $\rightarrow$ problem known as over-controlling, (Dell, Jones & Olken, 2014).

Further reading

See:

- Lecture Notes Sections 1.1-1.3, 2.1-2.3
- Greene

  - Linear regression model - Chapter 2 (Discussion of assumptions more general than Lecture 1 but does match Lecture 3)
  - OLS - Chapter 3 (Material in Section 3.4 not covered in lecture but read for concept of partial regression)