1 Introduction

This handout reviews some of the key points regarding regression algebra and the multivariate normal distribution. It follows closely Goldberger Ch.’s 17 and Ch. 18.

2 Short and Long Regressions

The basic set-up is

\[ y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon \tag{1} \]

where we have partitioned the \( n \times k \) matrix \( X \) into two submatrices \( X_1 \in \mathbb{R}^{n \times k_1} \) and \( X_2 \in \mathbb{R}^{n \times k_2} \).

We can think of two regressions:

1. a short one

\[ y = X_1\beta_1 + \varepsilon \tag{2} \]

and,

2. a long one

\[ y = X_1\beta_1 + X_2\beta_2 + \varepsilon \tag{3} \]

I’ll use the same notation as Goldberger so let \( b_i \) be a vector of OLS parameter estimates for the subvector \( \beta_i \) in a long regression and \( b_i^* \) be the OLS estimates of \( \beta_i \) in a short regression. And, let \( e \) be the residuals from the long regression and \( e^* \) be the residuals from the short regression.
Exercise 1:

Let \( b_1^* \) be the OLS estimates of \( \beta_1 \) from regression 2 and \( b_1 \) be the OLS estimates of \( \beta_1 \) from regression 3. Show

\[
b_1^* = b_1 + (X'_1X_1)^{-1}X'_1X_2b_2
\]

(4)

Exercise 2:

Letting \( e^* \) be the residuals from 2 show that

\[
e^* = M_1X_2b_2 + e
\]

In words, what is \( M_1X_2 \)?

Exercise 3:

Show that

\[
e'^*e^* = b'_2X'_2M_1X_2b_2 + e'e
\]

and interpret this result. What implication does this have for the fit of the long regression relative to the short regression?

Result 1 Some exceptions

1. If \( b_2 = 0 \) then \( b_1^* = b_1 \) and \( e^* = e \).
2. If \( X'_1X_2 = 0 \) then \( b_1^* = b_1 \) but \( e^* \neq e \).

3 Frisch-Waugh-Lovell

Problem 2 proves the Frisch-Waugh-Lovell theorem which can be thought of as an alternative way of getting at the OLS estimator of \( \beta_2 \).

1. Regress each column of \( X_2 \) on \( X_1 \) and save the corresponding set of residuals in a matrix, \( X_2^* \).
2. Regress \( y \) on \( X_1 \) and save its residual as \( y^* \). (In fact, this step is unnecessary and Goldberger refers to this as a double residual regression. Exercise: 4 Prove that this step is in fact unnecessary)
3. Regress \( y^* \) on \( X_2^* \) and the resulting coefficient vector is the same as the OLS coefficients from the original regression in 1.
To see this consider regressing $y$ on $M_1X_2$ (= $X_2^*$ in Goldberger). The coefficient vector is

$$c_2^* = (X_2'M_1X_2)^{-1}X_2'M_1y$$
$$= (X_2'M_1X_2)^{-1}X_2'M_1(X_1b_1 + X_2b_2 + e)$$
$$= (X_2'M_1X_2)^{-1}X_2'M_1(X_2b_2 + e) \quad (M_1X_1 = 0)$$
$$= b_2 \quad \text{(cancelling and noting } M_1e = e \text{ and } X_2'e = 0)$$

For some applications see section 17.4.

4 The CR Model

Recall the set-up

$$E(y) = X\beta = X_1\beta_1 + X_2\beta_2$$
$$V(y) = \sigma^2 I$$

$X$: full rank and nonstochastic

4.1 The Parameters

Exercise 5: (Omitted Variables Bias):

Show that the estimated coefficients from the short regression (2) $b_1^*$ are biased.

Exercise 6:

What is the variance of the short regression coefficients $b_1^*$ and what is its relation relative to the variance of the long regression coefficients $b_1$?

4.2 The Residuals

Exercise 7:

Find the expectation and variance of the short regression residual vector $e^*$.

Exercise 8:

Find the expectation of the sum of squared residuals, $e^te^*$. 
5 The Normal Distribution

You better become REAL familiar with this. There are just a zillion different properties that the normal (univariate and multivariate) distribution has. Here’s a short list of some things that might be worth knowing:

5.1 Univariate Normal Distribution

1. \( X \sim N (\mu, \sigma^2) \) means \( X \) has a univariate normal distribution with mean parameter \( \mu \) and \( \sigma^2 \). The density is of course

\[
f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\}
\]

which is often denoted by \( \phi(x) \) and there is no closed form for the corresponding distribution, \( \Phi(x) \).

2. The distribution is symmetric implying

\[
\Phi(-x) = 1 - \Phi(x)
\]

This is easily seen by thinking of the area under the normal density.

3. Closed under affine transformations. If \( x \sim N (\mu, \sigma^2) \) then \( y = \alpha + \beta x \) is distributed \( N (\alpha + \beta \mu, \beta^2 \sigma^2) \).

4. Is uniquely determined by it’s first two moments.

5. \( \phi'(x) = x \phi(x) \)

6. If \( Z \) is standard normal than all odd moments are equal to 0 and

\[
E (Z^{2k}) = \frac{(2k)!}{2^k \cdot k!}, \quad k = 1, 2, 3, ...
\]

(This can be shown using integration by parts and induction)

5.2 Multivariate Normal Distribution

1. The vector \( \mathbf{x} \in \mathbb{R}^n \) is distributed multivariate normal with mean vector \( \mu \) and variance-covariance matrix \( \Sigma \) and has the corresponding density

\[
f_X(x) = (2\pi)^{-n/2} \times |\Sigma|^{-1/2} \times \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \right\}
\]

where \( |\cdot| \) means determinant.
2. Two normal random variables are independent if and only if they are uncorrelated.

3. Affine transformations of a vector of normal random variables are again normal. So, if \( x \sim \text{MVN} (\mu, \Sigma) \) then \( y = Hx + b \) is distributed multivariate normal with mean \( H\mu + b \) and variance \( H\Sigma H' \)

4. Important!!! Consider a pair of random vectors \( x \) and \( y \) each multivariate normal such that \((x', y')\) has mean and covariance matrix given by

\[
\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}
\]

respectively. Then the distribution of \( x \) conditional on \( y \) is also multivariate normal with mean

\[
\mu_{x|y} = \mu_x + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \mu_y)
\]

and covariance matrix

\[
\Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx}
\]

Note that the conditional covariance matrix does not depend on \( y \) and that while \( \Sigma \) and \( \Sigma_{yy} \) are assumed to be nonsingular, \( \Sigma_{yy}^{-1} \) can be replaced by a pseudo inverse.

5.3 Functions of Normal Random Variables

1. Let \( x \) be a \( k \)-dimensional vector of standard normal random variables. Then \( x'x \) is distributed \( \chi^2 \) with \( k \) degrees of freedom.

2. Extending the above result, if \( x \in \mathbb{R}^n \) is distributed \( \text{MVN} (\mu, \Sigma) \) then

\[
(x - \mu)' \Sigma (x - \mu)
\]

is distributed \( \chi^2_n \)

3. If \( x \in \mathbb{R}^n \) is distributed \( \text{MVN} (0, I) \) and \( M \) is any nonrandom idempotent matrix with rank \( r \leq n \) then \( u'Mu \) is distributed \( \chi^2_r \).

4. Let \( x \in \mathbb{R}^n \) be distributed \( \text{MVN} (0, I) \). Let \( M \) be any nonrandom idempotent matrix with rank \( r \leq n \) and let \( L \) be a nonrandom matrix such that \( LM = 0 \). Then \( a = Mu \) and \( b = Lu \) are independent random vectors.

5. Let \( v^2 \chi^2_n \) and \( w^2 \chi^2_d \) be two independent chi-square random variables. Then

\[
z = \frac{v/n}{w/d}
\]

is distributed Snedecor-\( F : F(n, m) \)
6. Let \( z \sim N(0, 1) \) and \( w \sim \chi^2_n \) independent of \( z \). Then

\[
t = \frac{z}{w/n}
\]

has a Student’s t-distribution with \( n \) degrees of freedom \((t_n)\)

7. If \( u \sim t_n \) then \( u^2 \sim F(1, n) \).