Contents

Preface

This manual contains the solutions to all of the problems in the second edition of my MIT Press book, *Econometric Analysis of Cross Section and Panel Data*. In addition to the problems printed in the text, I have included some “bonus problems” along with their solutions. Several of these problems I left out due to space constraints and others occured to me since the book was published. I have a collection of other problems, with solutions, that I have used over the past 10 years for problem sets, takehome exams, and in class exams. I am happy to provide these to instructors who have adopted the book for a course.

I solved the empirical examples using various versions of Stata, ranging from 8.0 through 11.0. I have included the Stata commands and output directly in the text. No doubt there are Stata users and users of other software packages who will, at least in some cases, see more efficient or more elegant ways to compute estimates and test statistics.

Some of the solutions are fairly long. In addition to filling in all or most of the algebraic steps, I have tried to offer commentary about why a particular problem is interesting, why I solved the problem the way I did, or which conclusions would change if we varied some of the assumptions. Several of the problems offer what appear to be novel solutions to situations that can arise in actual empirical work.

My progress in finishing this manual was slowed by a health problem in spring and summer of 2010. Fortunately, several graduate students came to my aid by either working through some problems or organizing the overall effort. I would like to thank Do Won Kwak, Cuicui Lu, Myoung-Jin Keay, Shenwu Sheng, Iraj Rahmani, and Monthien Satimanon for their able assistance.

I would appreciate learning about any mistakes in the solutions and also receiving suggestions for how to make the answers more transparent. Of course I will gladly entertain suggestions for how the text can be improved, too. I can be reached via email at wooldri1@msu.edu.

Solutions to Chapter 2 Problems

2.1. a. Simple partial differentiation gives

$$\frac{\partial \mathrm{E}(y|x_1,x_2)}{\partial x_1} = \beta_1 + \beta_4 x_2$$

and

$$\frac{\partial \mathrm{E}(y|x_1,x_2)}{\partial x_2} = \beta_2 + 2\beta_3 x_2 + \beta_4 x_1$$

b. By definition, $\mathrm{E}(u|x_1 x_2) = 0$. Because x_2^2 and $x_1 x_2$ are functions of (x_1,x_2), it does not matter whether or not we also condition on them: $\mathrm{E}(u|x_1,x_2,x_2^2,x_1 x_2) = 0$.

c. All we can say about $\mathrm{Var}(u|x_1,x_2)$ is that it is nonnegative for all x_1 and x_2: $\mathrm{E}(u|x_1,x_2) = 0$ in no way restricts $\mathrm{Var}(u|x_1,x_2)$.

2.2. a. Because $\partial \mathrm{E}(y|x)/\partial x = \delta_1 + 2\delta_2(x - \mu)$, the marginal effect of x on $\mathrm{E}(y|x)$ is a linear function of x. If δ_2 is negative then the marginal effect is less than δ_1 when x is above its mean. If, for example, $\delta_1 > 0$ and $\delta_2 < 0$, the marginal effect will eventually be negative for x far enough above μ. (Whether the values for x such that $\partial \mathrm{E}(y|x)/\partial x < 0$ represents an interesting segment of the population is a different matter.)

b. Because $\partial \mathrm{E}(y|x)/\partial x$ is a function of x, we take the expectation of $\partial \mathrm{E}(y|x)/\partial x$ over the distribution of x: $\mathrm{E}[\partial \mathrm{E}(y|x)/\partial x] = \mathrm{E}[\delta_1 + 2\delta_2(x - \mu)] = \delta_1 + 2\delta_2 \mathrm{E}[(x - \mu)] = \delta_1$.

c. One way to do this part is to apply Property LP.5 from Appendix 2A. We have

$$\begin{aligned} \mathrm{L}(y|1,x) = \mathrm{L}[\mathrm{E}(y|x)] &= \delta_0 + \delta_1 \mathrm{L}[(x-\mu)|1,x] + \delta_2 \mathrm{L}[(x-\mu)^2|1,x] \\ &= \delta_0 + \delta_1(x-\mu) + \delta_2(\gamma_0 + \gamma_1 x), \end{aligned}$$

because $\mathrm{L}[(x-\mu)|1,x] = x - \mu$ and $\gamma_0 + \gamma_1 x$ is the linear projection of $(x-\mu)^2$ on x. By assumption, $(x-\mu)^2$ and x are uncorrelated, and so $\gamma_1 = 0$. It follows that

$$L(y|x) = (\delta_0 - \delta_1\mu + \delta_2\gamma_0) + \delta_1 x$$

2.3. a. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$, where u has a zero mean given x_1 and x_2: $E(u|x_1,x_2) = 0$. We can say nothing further about u.

b. $\partial E(y|x_1,x_2)/\partial x_1 = \beta_1 + \beta_3 x_2$. Because $E(x_2) = 0$, $\beta_1 = E[\partial E(y|x_1,x_2)/\partial x_1]$, that is, β_1 is the average partial effect of x_1 on $E(y|x_1,x_2)/\partial x_1$. Similarly, $\beta_2 = E[\partial E(y|x_1,x_2)/\partial x_2]$.

c. If x_1 and x_2 are independent with zero mean then $E(x_1 x_2) = E(x_1)E(x_2) = 0$. Further, the covariance between $x_1 x_2$ and x_1 is $E(x_1 x_2 \cdot x_1) = E(x_1^2 x_2) = E(x_1^2)E(x_2)$ (by independence) $= 0$. A similar argument shows that the covariance between $x_1 x_2$ and x_2 is zero. But then the linear projection of $x_1 x_2$ onto $(1,x_1,x_2)$ is identically zero. Now just use the law of iterated projections (Property LP.5 in Appendix 2A):

$$\begin{aligned} L(y|1,x_1,x_2) &= L(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 | 1,x_1,x_2) \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 L(x_1 x_2 | 1,x_1,x_2) \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2. \end{aligned}$$

d. Equation (2.47) is more useful because it allows us to compute the partial effects of x_1 and x_2 at *any* values of x_1 and x_2. Under the assumptions we have made, the linear projection in (2.48) does have as its slope coefficients on x_1 and x_2 the partial effects at the population average values of x_1 and x_2 – zero in both cases – but it does not allow us to obtain the partial effects at any other values of x_1 and x_2. Incidentally, the main conclusions of this problem go through if we allow x_1 and x_2 to have nonzero population means.

2.4. By assumption,

$$E(u|\mathbf{x}, v) = \delta_0 + \mathbf{x}\boldsymbol{\delta} + \rho_1 v$$

for some scalars δ_0, ρ_1 and a column vector $\boldsymbol{\delta}$. Now, it suffices to show that $\delta_0 = 0$ and $\boldsymbol{\delta} = \mathbf{0}$. One way to do this is to use LP.7 in Appendix 2A, and in particular, equation (2.56). This says that $(\delta_0, \boldsymbol{\delta}')'$ can be obtained by first projecting $(1, \mathbf{x})$ onto v, and obtaining the population residual, $\mathbf{r}$. Then, project u onto $\mathbf{r}$. Now, since v has zero mean and is uncorrelated with $\mathbf{x}$, the first step projection does nothing: $\mathbf{r} = (1, \mathbf{x})$. Thus, projecting u onto $\mathbf{r}$ is just projecting u onto $(1, \mathbf{x})$. Since u has zero mean and is uncorrelated with $\mathbf{x}$, this projection is identically zero, which means that $\delta_0 = 0$ and $\boldsymbol{\delta} = 0$.

2.5. By definition and the zero conditional mean assumptions, $\text{Var}(u_1|\mathbf{x}, \mathbf{z}) = \text{Var}(y|\mathbf{x}, \mathbf{z})$ and $\text{Var}(u_2|\mathbf{x}) = \text{Var}(y|\mathbf{x})$. By assumption, these are constant and necessarily equal to $\sigma_1^2 \equiv \text{Var}(u_1)$ and $\boldsymbol{\sigma}_2^2 \equiv \text{Var}(u_2)$, respectively. But then Property CV.4 implies that $\sigma_2^2 \geq \sigma_1^2$. This simple conclusion means that, when error variances are constant, the error variance falls as more explanatory variables are conditioned on.

2.6. a. By linearity of the linear projection,

$$\text{L}(q|1, \mathbf{x}) = \text{L}(q^*|1, \mathbf{x}) + \text{L}(e|1, \mathbf{x}) = \text{L}(q^*|1, \mathbf{x}),$$

where the last inequality follows because $\text{L}(e|1, \mathbf{x}) = 0$ when $\text{E}(e) = 0$ and $\text{E}(\mathbf{x}'e) = \mathbf{0}$. Therefore, the parameters in the linear projection of q onto (1,$\mathbf{x}$) are the same as the linear projection of q^* onto (1,$\mathbf{x}$). This fact is useful for studying equations with measurement error in the explained or explanatory variables.

b. $r = q - \text{L}(q|1, \mathbf{x}) = (q^* + e) - \text{L}(q|1, \mathbf{x}) = (q^* + e) - \text{L}(q^*|1, \mathbf{x})$ (from part a)

$= [q^* - \text{L}(q^*|1, \mathbf{x}) + e = r^* + e$.

2.7. Write the equation in error form as

$$y = g(\mathbf{x}) + \mathbf{z}\boldsymbol{\beta} + u$$
$$\text{E}(u|\mathbf{x}, \mathbf{z}) = 0.$$

Take the expected value of the first equation conditional only on $\mathbf{x}$:

$$\mathrm{E}(y|\mathbf{x}) = g(\mathbf{x}) + [\mathrm{E}(\mathbf{z}|\mathbf{x})]\boldsymbol{\beta}$$

and subtract this from the first equation to get

$$y - \mathrm{E}(y|\mathbf{x}) = [\mathbf{z} - \mathrm{E}(\mathbf{z}|\mathbf{x})]\boldsymbol{\beta} + u$$

or

$$\tilde{y} = \tilde{\mathbf{z}}\boldsymbol{\beta} + u$$

Because $\tilde{\mathbf{z}}$ is a function of $(\mathbf{x},\mathbf{z})$, $\mathrm{E}(u|\tilde{\mathbf{z}}) = 0$ [since $\mathrm{E}(u|\mathbf{x},\mathbf{z}) = 0$], and so $\mathrm{E}(\tilde{y}|\tilde{\mathbf{z}}) = \tilde{\mathbf{z}}\boldsymbol{\beta}$.

This basic result is fundamental in the literature on estimating *partial linear models*. First, one estimates $\mathrm{E}(y|\mathbf{x})$ and $\mathrm{E}(\mathbf{z}|\mathbf{x})$ using very flexible methods (typically, *nonparametric methods*). Then, after obtaining residuals of the form $\tilde{y}_i \equiv y_i - \hat{\mathrm{E}}(y_i|\mathbf{x}_i)$ and $\tilde{\mathbf{z}}_i \equiv \mathbf{z}_i - \hat{\mathrm{E}}(\mathbf{z}_i|\mathbf{x}_i)$, $\boldsymbol{\beta}$ is estimated from an OLS regression $\tilde{y}_i$ on $\tilde{\mathbf{z}}_i, i = 1,\ldots,N$. Under general conditions, this kind of nonparametric partialling-out procedure leads to a $\sqrt{N}$-consistent, asymptotically normal estimator of $\boldsymbol{\beta}$. See Robinson (1988) and Powell (1994).

In the case where $\mathrm{E}(y|\mathbf{x})$ and the elements of $\mathrm{E}(\mathbf{z}|\mathbf{x})$ are approximated as linear functions of a common set of functions, say $\{h_1(\mathbf{x}),\ldots,h_Q(\mathbf{x})\}$, the partialling out is equivalent to estimating a linear model

$$y = \alpha_0 + \alpha_1 h_1(\mathbf{x}) + \ldots + \alpha_Q h_Q(\mathbf{x}) + \mathbf{x}\boldsymbol{\beta} + error$$

by OLS.

2.8. a. By exponentiation we can write $y = \exp[g(\mathbf{x}) + u] = \exp[g(\mathbf{x})]\exp(u)$. It follows that

$$\mathrm{E}(y|\mathbf{x}) = \exp[g(\mathbf{x})]\mathrm{E}[\exp(u)|\mathbf{x}] = \exp[g(\mathbf{x})]a(\mathbf{x})$$

Using the product rule gives

$$\begin{aligned}\frac{\partial \mathrm{E}(y|\mathbf{x})}{\partial x_j} &= \frac{\partial g(\mathbf{x})}{\partial x_j}\exp[g(\mathbf{x})]a(\mathbf{x}) + \exp[g(\mathbf{x})]\frac{\partial a(\mathbf{x})}{\partial x_j} \\ &= \frac{\partial g(\mathbf{x})}{\partial x_j}\mathrm{E}(y|\mathbf{x}) + \mathrm{E}(y|\mathbf{x})\frac{\partial a(\mathbf{x})}{\partial x_j}\cdot\frac{1}{a(\mathbf{x})}\end{aligned}$$

Therefore,

$$\frac{\partial \mathrm{E}(y|\mathbf{x})}{\partial x_j}\cdot\frac{x_j}{\mathrm{E}(y|\mathbf{x})} = \frac{\partial g(\mathbf{x})}{\partial x_j}\cdot x_j + \frac{\partial a(\mathbf{x})}{\partial x_j}\cdot\frac{x_j}{a(\mathbf{x})}$$

We can establish this relationship more simply by assuming $\mathrm{E}(y|\mathbf{x}) > 0$ for all $\mathbf{x}$ and using equation (2.10).

b. Write $z_j \equiv \log(x_j)$ so $x_j = \exp(z_j)$. Then, using the chain rule,

$$\frac{\partial g(\mathbf{x})}{\partial \log(x_j)} = \frac{\partial g(\mathbf{x})}{\partial z_j} = \frac{\partial g(\mathbf{x})}{\partial x_j}\cdot\frac{\partial x_j}{\partial z_j} = \frac{\partial g(\mathbf{x})}{\partial x_j}\cdot\exp(z_j) = \frac{\partial g(\mathbf{x})}{\partial x_j}\cdot x_j$$

c. From $\log(y) = g(\mathbf{x}) + u$ and $\mathrm{E}(u|\mathbf{x}) = 0$ we have $\mathrm{E}[\log(y)|\mathbf{x})] = g(\mathbf{x})$. Therefore, using (2.11), the elasticity would be simply

$$\frac{\partial g(\mathbf{x})}{\partial \log(x_j)} = \frac{\partial g(\mathbf{x})}{\partial x_j}\cdot x_j$$

which, compared with the definition based on $\mathrm{E}(y|\mathbf{x})$, omits the elasticity of $a(\mathbf{x})$ with respect to x_j.

2.9. This is easily shown by using iterated expectations:

$$\mathrm{E}(\mathbf{x}'y) = \mathrm{E}[\mathrm{E}(\mathbf{x}'y|\mathbf{x})] = \mathrm{E}[\mathbf{x}'\mathrm{E}(y|\mathbf{x})] = \mathrm{E}[\mathbf{x}'\mu(\mathbf{x})]$$

Therefore,

$$\boldsymbol{\delta} = [\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}\mathrm{E}(\mathbf{x}'y) = [\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}\mathrm{E}[\mathbf{x}'\mu(\mathbf{x})]$$

and the latter equation is the vector of parameters in the linear projection of $\mu(\mathbf{x})$ on $\mathbf{x}$.

2.10. a. As given in the hint, we can always write

$$\mathrm{E}(y|\mathbf{x},s) = (1-s)\cdot\mu_0(\mathbf{x}) + s\cdot\mu_1(\mathbf{x})$$

Now condition only on s and use iterated expectations:

$$\begin{aligned}\mathrm{E}(y|s) &= \mathrm{E}[\mathrm{E}(y|\mathbf{x},s)|s] = \mathrm{E}[(1-s)\cdot\mu_0(\mathbf{x}) + s\cdot\mu_1(\mathbf{x})|s] \\ &= (1-s)\mathrm{E}[\mu_0(\mathbf{x})|s] + s\mathrm{E}[\mu_1(\mathbf{x})|s]\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{E}(y|s=1) &= \mathrm{E}[\mu_1(\mathbf{x})|s=1] \\ \mathrm{E}(y|s=0) &= \mathrm{E}[\mu_0(\mathbf{x})|s=0]\end{aligned}$$

and so, by adding and subtracting $\mathrm{E}[\mu_0(\mathbf{x})|s=1]$, we get

$$\begin{aligned}\mathrm{E}(y|s=1) - \mathrm{E}(y|s=0) &= \mathrm{E}[\mu_1(\mathbf{x})|s=1] - \mathrm{E}[\mu_0(\mathbf{x})|s=0] \\ &= \{\mathrm{E}[\mu_1(\mathbf{x})|s=1] - \mathrm{E}[\mu_0(\mathbf{x})|s=1]\} + \{\mathrm{E}[\mu_0(\mathbf{x})|s=1] - \mathrm{E}[\mu_0(\mathbf{x})|s=0]\}\end{aligned}$$

b. Use part a and linearity of the conditional means:

$$\begin{aligned}\mathrm{E}(y|s=1) - \mathrm{E}(y|s=0) &= [\mathrm{E}(\mathbf{x}|s=1)\boldsymbol{\beta}_1 - \mathrm{E}(\mathbf{x}|s=1)\boldsymbol{\beta}_0] + [\mathrm{E}(\mathbf{x}|s=1)\boldsymbol{\beta}_0 - \mathrm{E}(\mathbf{x}|s=0)\boldsymbol{\beta}_0] \\ &= \mathrm{E}(\mathbf{x}|s=1)\cdot(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + [\mathrm{E}(\mathbf{x}|s=1) - \mathrm{E}(\mathbf{x}|s=0)]\cdot\boldsymbol{\beta}_0\end{aligned}$$

This decomposition attributes the difference in the unconditional means, $\mathrm{E}(y|s=1) - \mathrm{E}(y|s=0)$, to two pieces. The first part is due to differences in the regression parameters, $\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0$ – where we evaluate the difference at the average of the covariates from the $s = 1$ subpopulation. The second part is due to a difference in means of the covariates from the two subpopulations – where we apply the regression coefficients from the $s = 0$ subpopulation. If, for example, the two regression functions are the same – that is, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$ – then any difference in the subpopulation means $\mathrm{E}(y|s=0)$ and $\mathrm{E}(y|s=1)$ is due to a difference in averages of the covariates across the subpopulations. If the covariate means are the same – that is, $\mathrm{E}(\mathbf{x}|s=1) = \mathrm{E}(\mathbf{x}|s=0)$ – then $\mathrm{E}(y|s=0)$ and $\mathrm{E}(y|s=0)$ can still differ if

$\beta_1 \neq \beta_0$. In many applications, both pieces in $E(y|s = 1) - E(y|s = 0)$ are present.

Incidentally, the approach in this problem is not the only interesting way to decompose $E(y|s = 1) - E(y|s = 0)$. See, for example, T.E. Elder, J.H. Goddeeris, and S.J. Haider, "Unexplained Gaps and Oaxaca–Blinder Decompositions," *Labour Economics*, 2010.

Solutions to Chapter 3 Problems

3.1. To prove Lemma 3.1, we must show that for all $\varepsilon > 0$, there exists $b_\varepsilon < \infty$ and an integer N_ε such that $P[|x_N| \geq b_\varepsilon] < \boldsymbol{\varepsilon}$, all $N \geq N_\varepsilon$. We use the following fact: since $x_N \xrightarrow{p} a$, for any $\varepsilon > 0$ there exists an integer N_ε such that $P[|x_N - a| > 1] < \varepsilon$ for all $N \geq N_\varepsilon$. [The existence of N_ε is implied by Definition 3.3(1).] But $|x_N| = |x_N - a + a| \leq |x_N - a| + |a|$ (by the triangle inequality), and so $|x_N| - |a| \leq |x_N - a|$. It follows that $P[|x_N| - |a| > 1] \leq P[|x_N - a| > 1]$. Therefore, in Definition 3.3(3) we can take $b_\varepsilon \equiv |a| + 1$ (irrespective of the value of ε) and then the existence of N_ε follows from Definition 3.3(1).

3.2. Each element of the $K \times 1$ vector $\mathbf{Z}_N'\mathbf{x}_N$ is the sum of J terms of the form $Z_{Nji}x_{Nj}$. Because $Z_{Nji} = o_p(1)$ and $x_{Nj} = O_p(1)$, each term in the sum is $o_p(1)$ from Lemma 3.2(4). By Lemma 3.2(1), the sum of $o_p(1)$ terms is $o_p(1)$.

3.3. This follows immediately from Lemma 3.1 because $\mathbf{g}(\mathbf{x}_N) \xrightarrow{p} \mathbf{g}(\mathbf{c})$.

3.4. Both parts follow from the continuous mapping theorem and basic properties of the normal distribution.

a. The function defined by $\mathbf{g}(\mathbf{z}) = \mathbf{A}'\mathbf{z}$ is clearly continuous. Further, if $\mathbf{z} \sim \text{Normal}(\mathbf{0}, \mathbf{V})$ then $\mathbf{A}'\mathbf{z} \sim \text{Normal}(\mathbf{0}, \mathbf{A}'\mathbf{V}\mathbf{A})$. By the continuous mapping theorem, $\mathbf{A}'\mathbf{z}_N \xrightarrow{d} \mathbf{A}'\mathbf{z} \sim \text{Normal}(\mathbf{0}, \mathbf{A}'\mathbf{V}\mathbf{A})$.

b. Because $\mathbf{V}$ is nonsingular, the function $g(\mathbf{z}) = \mathbf{z}'\mathbf{V}^{-1}\mathbf{z}$ is continuous. But if $\mathbf{z} \sim \text{Normal}(\mathbf{0}, \mathbf{V})$, $\mathbf{z}'\mathbf{V}^{-1}\mathbf{z} \sim \chi^2_K$. So $\mathbf{z}_N'\mathbf{V}^{-1}\mathbf{z}_N \xrightarrow{d} \mathbf{z}'\mathbf{V}^{-1}\mathbf{z} \sim \chi^2_K$.

3.5. a. Because $\text{Var}(\bar{y}_N) = \sigma^2/N$, $\text{Var}[\sqrt{N}(\bar{y}_N - \mu)] = N(\sigma^2/N) = \sigma^2$.

b. By the CLT, $\sqrt{N}(\bar{y}_N - \mu) \overset{a}{\sim} \text{Normal}(0, \sigma^2)$, and so $\text{Avar}[\sqrt{N}(\bar{y}_N - \mu)] = \sigma^2$.

c. We obtain $\text{Avar}(\bar{y}_N)$ by dividing $\text{Avar}[\sqrt{N}(\bar{y}_N - \mu)]$ by N. Therefore, $\text{Avar}(\bar{y}_N) = \sigma^2/N$.