updated: 11/23/00, 1/12/03 (answer to Q7 of Section 1.3 added)

Hayashi *Econometrics*: Answers to Selected Review Questions

Chapter 1

Section 1.1

1. The intercept is increased by $\log(100)$.

2. Since $(\varepsilon_i, \mathbf{x}_i)$ is independent of $(\varepsilon_j, \mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_n)$ for $i \neq j$, we have: $\mathrm{E}(\varepsilon_i \mid \mathbf{X}, \varepsilon_j) = \mathrm{E}(\varepsilon_i \mid \mathbf{x}_i)$. So

$$\begin{aligned}
&\mathrm{E}(\varepsilon_i \varepsilon_j \mid \mathbf{X}) \\
&= \mathrm{E}[\mathrm{E}(\varepsilon_j\, \varepsilon_i \mid \mathbf{X}, \varepsilon_j) \mid \mathbf{X}] \qquad \text{(by Law of Iterated Expectations)} \\
&= \mathrm{E}[\varepsilon_j\, \mathrm{E}(\varepsilon_i \mid \mathbf{X}, \varepsilon_j) \mid \mathbf{X}] \qquad \text{(by linearity of conditional expectations)} \\
&= \mathrm{E}[\varepsilon_j\, \mathrm{E}(\varepsilon_i \mid \mathbf{x}_i) \mid \mathbf{X}] \\
&= \mathrm{E}(\varepsilon_i \mid \mathbf{x}_i)\, \mathrm{E}(\varepsilon_j \mid \mathbf{x}_j).
\end{aligned}$$

The last equality follows from the linearity of conditional expectations because $\mathrm{E}(\varepsilon_i \mid \mathbf{x}_i)$ is a function of $\mathbf{x}_i$.

3.

$$\begin{aligned}
\mathrm{E}(y_i \mid \mathbf{X}) &= \mathrm{E}(\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \mid \mathbf{X}) \qquad \text{(by Assumption 1.1)} \\
&= \mathbf{x}_i'\boldsymbol{\beta} + \mathrm{E}(\varepsilon_i \mid \mathbf{X}) \qquad \text{(since } \mathbf{x}_i \text{ is included in } \mathbf{X}) \\
&= \mathbf{x}_i'\boldsymbol{\beta} \qquad \text{(by Assumption 1.2).}
\end{aligned}$$

Conversely, suppose $\mathrm{E}(y_i \mid \mathbf{X}) = \mathbf{x}_i'\boldsymbol{\beta} \quad (i = 1, 2, \dots, n)$. Define $\varepsilon_i \equiv y_i - \mathrm{E}(y_i \mid \mathbf{X})$. Then by construction Assumption 1.1 is satisfied: $\varepsilon_i = y_i - \mathbf{x}_i'\boldsymbol{\beta}$. Assumption 1.2 is satisfied because

$$\begin{aligned}
\mathrm{E}(\varepsilon_i \mid \mathbf{X}) &= \mathrm{E}(y_i \mid \mathbf{X}) - \mathrm{E}[\mathrm{E}(y_i \mid \mathbf{X}) \mid \mathbf{X}] \qquad \text{(by definition of } \varepsilon_i \text{ here)} \\
&= 0 \qquad \text{(since } \mathrm{E}[\mathrm{E}(y_i \mid \mathbf{X}) \mid \mathbf{X}] = \mathrm{E}(y_i \mid \mathbf{X})).
\end{aligned}$$

4. Because of the result in the previous review question, what needs to be verified is Assumption 1.4 and that $\mathrm{E}(CON_i \mid YD_1, \dots, YD_n) = \beta_1 + \beta_2\, YD_i$. That the latter holds is clear from the i.i.d. assumption and the hint. From the discussion in the text on random samples, Assumption 1.4 is equivalent to the condition that $\mathrm{E}(\varepsilon_i^2 \mid YD_i)$ is a constant, where $\varepsilon_i \equiv CON_i - \beta_1 - \beta_2\, YD_i$.

$$\begin{aligned}
\mathrm{E}(\varepsilon_i^2 \mid YD_i) &= \mathrm{Var}(\varepsilon_i \mid YD_i) \qquad \text{(since } \mathrm{E}(\varepsilon_i \mid YD_i) = 0) \\
&= \mathrm{Var}(CON_i \mid YD_i).
\end{aligned}$$

This is a constant since (CON_i, YD_i) is jointly normal.

5. If $x_{i2} = x_{j2}$ for all i, j, then the rank of $\mathbf{X}$ would be one.

6. By the Law of Total Expectations, Assumption 1.4 implies

$$\mathrm{E}(\varepsilon_i^2) = \mathrm{E}[\mathrm{E}(\varepsilon_i^2 \mid \mathbf{X})] = \mathrm{E}[\sigma^2] = \sigma^2.$$

Similarly for $\mathrm{E}(\varepsilon_i \, \varepsilon_j)$.

Section 1.2

5. (b)

$$\begin{aligned}
\mathbf{e}'\mathbf{e} &= (\mathbf{M}\boldsymbol{\varepsilon})'(\mathbf{M}\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{M}\boldsymbol{\varepsilon} \qquad \text{(recall from matrix algebra that } (\mathbf{AB})' = \mathbf{B}'\mathbf{A}'\text{)} \\
&= \boldsymbol{\varepsilon}'\mathbf{M}\mathbf{M}\boldsymbol{\varepsilon} \qquad \text{(since } \mathbf{M} \text{ is symmetric)} \\
&= \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} \qquad \text{(since } \mathbf{M} \text{ is itempotent).}
\end{aligned}$$

6. A change in the unit of measurement for y means that y_i gets multiplied by some factor, say λ, for all i. The OLS formula shows that $\mathbf{b}$ gets multiplied by λ. So $\overline{y}$ gets multiplied by the same factor λ, leaving R^2 unaffected. A change in the unit of measurement for regressors leaves $\mathbf{x}_i'\mathbf{b}$, and hence R^2, unaffected.

Section 1.3

4(a). Let $\mathbf{d} \equiv \widehat{\boldsymbol{\beta}} - \mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})$, $\mathbf{a} \equiv \widehat{\boldsymbol{\beta}} - \mathrm{E}(\widehat{\boldsymbol{\beta}})$, and $\mathbf{c} \equiv \mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) - \mathrm{E}(\widehat{\boldsymbol{\beta}})$. Then $\mathbf{d} = \mathbf{a} - \mathbf{c}$ and $\mathbf{dd}' = \mathbf{aa}' - \mathbf{ca}' - \mathbf{ac}' + \mathbf{cc}'$. By taking unconditional expectations of both sides, we obtain

$$\mathrm{E}(\mathbf{dd}') = \mathrm{E}(\mathbf{aa}') - \mathrm{E}(\mathbf{ca}') - \mathrm{E}(\mathbf{ac}') + \mathrm{E}(\mathbf{cc}').$$

Now,

$$\begin{aligned}
\mathrm{E}(\mathbf{dd}') &= \mathrm{E}[\mathrm{E}(\mathbf{dd}' \mid \mathbf{X})] \qquad \text{(by Law of Total Expectations)} \\
&= \mathrm{E}\Big\{\mathrm{E}[(\widehat{\boldsymbol{\beta}} - \mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}))(\widehat{\boldsymbol{\beta}} - \mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}))' \mid \mathbf{X}]\Big\} \\
&= \mathrm{E}\Big[\mathrm{Var}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})\Big] \qquad \text{(by the first equation in the hint).}
\end{aligned}$$

By definition of variance, $\mathrm{E}(\mathbf{aa}') = \mathrm{Var}(\widehat{\boldsymbol{\beta}})$. By the second equation in the hint, $\mathrm{E}(\mathbf{cc}') = \mathrm{Var}[\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})]$. For $\mathrm{E}(\mathbf{ca}')$, we have:

$$\begin{aligned}
\mathrm{E}(\mathbf{ca}') &= \mathrm{E}[\mathrm{E}(\mathbf{ca}' \mid \mathbf{X})] \\
&= \mathrm{E}\Big\{\mathrm{E}[(\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) - \mathrm{E}(\widehat{\boldsymbol{\beta}}))(\widehat{\boldsymbol{\beta}} - \mathrm{E}(\widehat{\boldsymbol{\beta}}))' \mid \mathbf{X}]\Big\} \\
&= \mathrm{E}\Big\{(\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) - \mathrm{E}(\widehat{\boldsymbol{\beta}}))\,\mathrm{E}[(\widehat{\boldsymbol{\beta}} - \mathrm{E}(\widehat{\boldsymbol{\beta}}))' \mid \mathbf{X}]\Big\} \\
&= \mathrm{E}\Big\{(\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) - \mathrm{E}(\widehat{\boldsymbol{\beta}}))(\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) - \mathrm{E}(\widehat{\boldsymbol{\beta}}))'\Big\} \\
&= \mathrm{E}(\mathbf{cc}') = \mathrm{Var}[\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})].
\end{aligned}$$

Similarly, $\mathrm{E}(\mathbf{ac}') = \mathrm{Var}[\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})]$.

4(b). Since by assumption $\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) = \boldsymbol{\beta}$, we have $\mathrm{Var}[\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})] = \mathbf{0}$. So the equality in (a) for the unbiased estimator $\widehat{\boldsymbol{\beta}}$ becomes $\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \mathrm{E}[\mathrm{Var}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})]$. Similarly for the OLS estimator $\mathbf{b}$, we have: $\mathrm{Var}(\mathbf{b}) = \mathrm{E}[\mathrm{Var}(\mathbf{b} \mid \mathbf{X})]$. As noted in the hint, $\mathrm{E}[\mathrm{Var}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})] \geq \mathrm{E}[\mathrm{Var}(\mathbf{b} \mid \mathbf{X})]$.

7. p_i is the i-th diagonal element of the projection matrix $\mathbf{P}$. Since $\mathbf{P}$ is positive semi-definite, its diagonal elements are all non-negative. Hence $p_i \geq 0$. $\sum_{i=1}^{n} p_i = K$ because this sum equals the trace of $\mathbf{P}$ which equals K. To show that $p_i \leq 1$, first note that p_i can be written as: $\mathbf{e}_i'\mathbf{P}\mathbf{e}_i$ where $\mathbf{e}_i$ is an n-dimensional i-th unit vector (so its i-th element is unity and the other elements are all zero). Now, recall that for the annihilator $\mathbf{M}$, we have $\mathbf{M} = \mathbf{I} - \mathbf{P}$ and $\mathbf{M}$ is positive semi-definite. So

$$\begin{aligned} \mathbf{e}_i'\mathbf{P}\mathbf{e}_i &= \mathbf{e}_i'\mathbf{e}_i - \mathbf{e}_i'\mathbf{M}\mathbf{e}_i \\ &= 1 - \mathbf{e}_i'\mathbf{M}\mathbf{e}_i \quad (\text{since } \mathbf{e}_i'\mathbf{e}_i = 1) \\ &\leq 1 \quad (\text{since } \mathbf{M} \text{ is positive semi-definite}). \end{aligned}$$

Section 1.4

6. As explained in the text, the overall significance increases with the number of restrictions to be tested if the t test is applied to each restriction without adjusting the critical value.

Section 1.5

2. Since $\partial^2 \log L(\boldsymbol{\zeta})/(\partial\widetilde{\boldsymbol{\theta}}\,\partial\widetilde{\boldsymbol{\psi}}') = \mathbf{0}$, the information matrix $\mathbf{I}(\boldsymbol{\zeta})$ is block diagonal, with its first block corresponding to $\boldsymbol{\theta}$ and the second corresponding to $\boldsymbol{\psi}$. The inverse is block diagonal, with its first block being the inverse of

$$-\mathrm{E}\left[\frac{\partial^2 \log L(\boldsymbol{\zeta})}{\partial\widetilde{\boldsymbol{\theta}}\,\partial\widetilde{\boldsymbol{\theta}}'}\right].$$

So the Cramer-Rao bound for $\boldsymbol{\theta}$ is the negative of the inverse of the expected value of (1.5.2). The expectation, however, is over $\mathbf{y}$ *and* $\mathbf{X}$ because here the density is a joint density. Therefore, the Cramer-Rao bound for $\boldsymbol{\beta}$ is $\sigma^2\,\mathrm{E}[(\mathbf{X}'\mathbf{X})]^{-1}$.

Section 1.6

3. $\mathrm{Var}(\mathbf{b} \mid \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathrm{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$.

Section 1.7

2. It just changes the intercept by b_2 times $\log(1000)$.

5. The restricted regression is

$$\log\left(\frac{TC_i}{p_{i2}}\right) = \beta_1 + \beta_2\,\log(Q_i) + \beta_3\,\log\left(\frac{p_{i1}}{p_{i2}}\right) + \beta_5\,\log\left(\frac{p_{i3}}{p_{i2}}\right) + \varepsilon_i. \tag{1}$$

The OLS estimate of $(\beta_1, \ldots, \beta_5)$ from (1.7.8) is $(-4.7, 0.72, 0.59, -0.007, 0.42)$. The OLS estimate from the above restricted regression should yield the same point estimate and standard errors. The SSR should be the same, but R^2 should be different.

6. That's because the dependent variable in the restricted regression is different from that in the unrestricted regression. If the dependent variable were the same, then indeed the R^2 should be higher for the unrestricted model.

7(b) No, because when the price of capital is constant across firms we are forced to use the adding-up restriction $\beta_1 + \beta_2 + \beta_3 = 1$ to calculate β_2 (capital's contribution) from the OLS estimate of β_1 and β_3.

8. Because input choices can depend on ε_i, the regressors would not be orthogonal to the error term. Under the Cobb-Douglas technology, input shares do not depend on factor prices. Labor share, for example, should be equal to $\alpha_1/(\alpha_1+\alpha_2+\alpha_3)$ for all firms. Under constant returns to scale, this share equals α_1. So we can estimate α's without sampling error.