Department of Economics
University of Maryland

John C. Chao
Spring 2013

Economics 422

Midterm Examination Solution Sheet

QUESTION 1

(a) Let $\overline{R}^2$ denote adjusted R^2, and note that

$$\overline{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)\frac{RSS}{TSS},$$

where k here denotes the number of regressors not counting the one associated with the intercept. Moreover, the unadjusted R^2 is given by the formula

$$R^2 = 1 - \frac{RSS}{TSS} \Longrightarrow RSS = TSS \times \left(1 - R^2\right)$$

which implies that

$$RSS = TSS \times \left(1 - R^2\right) = 552.36 \times (1 - 0.7911) \approx 115.39$$

Hence,

$$\begin{aligned} \overline{R}^2 &= 1 - \left(\frac{n-1}{n-k-1}\right)\frac{RSS}{TSS} \\ &= 1 - \left(\frac{20-1}{20-2-1}\right)\left(\frac{115.39}{552.36}\right) \\ &\approx 0.7665 \end{aligned}$$

(b)

$$\begin{aligned} SER &= \sqrt{\frac{RSS}{n-k-1}} = \sqrt{\frac{115.39}{20-2-1}} \approx 2.6053, \\ RMSE &= \sqrt{\frac{RSS}{n}} = \sqrt{\frac{115.39}{20}} \approx 2.4020 \end{aligned}$$

QUESTION 2

(a) The estimated slope of

$$\widehat{\beta}_1 = 8$$

means that an increase in temperature of one degree Fahrenheit leads to an additional 8 sodas being sold. Interpreted literally, the estimated intercept

$$\widehat{\beta}_0 = -240$$

means that when the temperature is $0°F$, the number of sodas sold is -240. While the estimated slope makes sense, the estimated intercept does not, since soda sales cannot be negative. This somewhat absurd result is probably due to the fact that we unwittingly extrapolated the estimated linear relationship beyond the range of our sample.

(b)

$$\widehat{Sodas} = -240 + 8 \times 80 = 400.$$

(c) Note that

$$\widehat{Sodas} = -240 + 8 \times Temp = 0 \Longrightarrow Temp = 30°F$$

Hence, when temperature is $30°F$ or below, the sales drop to zero.

QUESTION 3

(a) Note that, in this case,

$$\begin{aligned} \{\text{person } i \text{ is left-handed}\} &= \{LH_i = 1\} \\ \{\text{person } i \text{ is right-handed}\} &= \{LH_i = 0\} \end{aligned}$$

Hence,

$$\begin{aligned} & E\left[TS_i | \text{person } i \text{ is right-handed}\right] \\ = \; & E\left[TS_i | LH_i = 0\right] \\ = \; & E\left[\beta_0 | LH_i = 0\right] + E\left[\beta_1 LH_i | LH_i = 0\right] + E\left[u_i | LH_i = 0\right] \\ = \; & \beta_0 + \beta_1 E\left[LH_i | LH_i = 0\right] + E\left[u_i | LH_i = 0\right] \\ = \; & \beta_0 + \beta_1 E\left[LH_i | LH_i = 0\right] + 0 \\ & \text{(by assumption)} \\ = \; & \beta_0 \end{aligned}$$

where the fourth equality above follows from the assumption that $E[u_i|LH_i] = 0$ with probability one. In addition,

$$\begin{aligned}
& E[TS_i|\text{person } i \text{ is left-handed}] \\
= \; & E[TS_i|LH_i = 1] \\
= \; & E[\beta_0|LH_i = 1] + E[\beta_1 LH_i|LH_i = 1] + E[u_i|LH_i = 1] \\
= \; & \beta_0 + \beta_1 E[LH_i|LH_i = 1] + E[u_i|LH_i = 1] \\
= \; & \beta_0 + \beta_1 E[LH_i|LH_i = 1] + 0 \\
= \; & \beta_0 + \beta_1
\end{aligned}$$

where the fourth equality above again follows from the assumption that $E[u_i|LH_i] = 0$ with probability one. Let

$$\tau = E[TS_i|\text{person } i \text{ is right-handed}] - E[TS_i|\text{person } i \text{ is left-handed}],$$

and it follows that in this case

$$\tau = \beta_0 - \beta_0 - \beta_1 = -\beta_1$$

Hence, to obtain an unbiased estimator of τ, we estimate β_1 by OLS to obtain the estimator

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n \left(LH_i - \overline{LH}\right)\left(TS_i - \overline{TS}\right)}{\sum_{i=1}^n \left(LH_i - \overline{LH}\right)^2}$$

where

$$\overline{LH} = \frac{1}{n}\sum_{i=1}^n LH_i,$$

and set

$$\widehat{\tau} = -\widehat{\beta}_1.$$

Since $\widehat{\beta}_1$ is an unbiased estimator of β_1 under the standard assumptions, i.e.,

$$E\left[\widehat{\beta}_1\right] = \beta_1,$$

it follows that

$$E[\widehat{\tau}] = -E\left[\widehat{\beta}_1\right] = -\beta_1 = \tau,$$

so that $\widehat{\tau}$ is an unbiased estimator of τ.

(b) Note that, in this case,

$$\begin{aligned}
\{\text{person } i \text{ is left-handed}\} &= \{LH_i = 1\} \cap \{RH_i = 0\} \\
\{\text{person } i \text{ is right-handed}\} &= \{LH_i = 0\} \cap \{RH_i = 1\}
\end{aligned}$$

Hence,

$$
\begin{aligned}
& E\left[TS_i|\text{person } i \text{ is right-handed}\right] \\
= \ & E\left[TS_i|LH_i = 0, RH_i = 1\right] \\
= \ & E\left[\beta_1 LH_i|LH_i = 0, RH_i = 1\right] + E\left[\beta_2 RH_i|LH_i = 0, RH_i = 1\right] + E\left[u_i|LH_i = 0, RH_i = 1\right] \\
= \ & \beta_1 E\left[LH_i|LH_i = 0, RH_i = 1\right] + \beta_2 E\left[RH_i|LH_i = 0, RH_i = 1\right] + E\left[u_i|LH_i = 0, RH_i = 1\right] \\
= \ & \beta_1 E\left[LH_i|LH_i = 0, RH_i = 1\right] + \beta_2 E\left[RH_i|LH_i = 0, RH_i = 1\right] + 0 \\
= \ & 0 + \beta_2 + 0 \\
= \ & \beta_2
\end{aligned}
$$

where the fourth equality above follows from the assumption that $E\left[u_i|LH_i, RH_i\right] = 0$ with probability one. In addition,

$$
\begin{aligned}
& E\left[TS_i|\text{person } i \text{ is left-handed}\right] \\
= \ & E\left[TS_i|LH_i = 1, RH_i = 0\right] \\
= \ & E\left[\beta_1 LH_i|LH_i = 1, RH_i = 0\right] + E\left[\beta_2 RH_i|LH_i = 1, RH_i = 0\right] + E\left[u_i|LH_i = 1, RH_i = 0\right] \\
= \ & \beta_1 E\left[LH_i|LH_i = 1, RH_i = 0\right] + \beta_2 E\left[RH_i|LH_i = 1, RH_i = 0\right] + E\left[u_i|LH_i = 1, RH_i = 0\right] \\
= \ & \beta_1 E\left[LH_i|LH_i = 1, RH_i = 0\right] + \beta_2 E\left[RH_i|LH_i = 1, RH_i = 0\right] + 0 \\
= \ & \beta_1 + 0 + 0 \\
= \ & \beta_1.
\end{aligned}
$$

where the fourth equality above again follows from the assumption that $E\left[u_i|LH_i, RH_i\right] = 0$ with probability one. It follows that in this case

$$\tau = \beta_2 - \beta_1$$

Hence, to obtain an unbiased estimator of τ, we estimate β_1 and β_2 by OLS to obtain the estimators $\widehat{\beta}_1$ and $\widehat{\beta}_2$. Our estimator of τ is then

$$\widehat{\tau} = \widehat{\beta}_2 - \widehat{\beta}_1.$$

Since both $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are unbiased estimators under the standard assumptions, i.e.,

$$E\left[\widehat{\beta}_1\right] = \beta_1 \text{ and } E\left[\widehat{\beta}_2\right] = \beta_2,$$

it follows that

$$E\left[\widehat{\tau}\right] = E\left[\widehat{\beta}_2\right] - E\left[\widehat{\beta}_1\right] = \beta_2 - \beta_1 = \tau,$$

so that $\widehat{\tau}$ is an unbiased estimator of τ.

QUESTION 4

(a) Note that

$$\begin{aligned}\widetilde{\beta}_1 &= \frac{Y_2 - Y_1}{X_2 - X_1} \\ &= \frac{\beta_0 + \beta_1 X_2 + u_2 - \beta_0 - \beta_1 X_1 - u_1}{X_2 - X_1} \\ &= \beta_1 \frac{X_2 - X_1}{X_2 - X_1} + \frac{u_2 - u_1}{X_2 - X_1} \\ &= \beta_1 + \frac{u_2 - u_1}{X_2 - X_1}\end{aligned}$$

Note that since X_i is a continuous random variable $X_2 - X_1 = 0$ with probability zero. Taking expectation, we get

$$\begin{aligned}E\left[\widetilde{\beta}_1\right] &= E\left[\beta_1\right] + E\left[\frac{u_2 - u_1}{X_2 - X_1}\right] \\ &= \beta_1 + E\left[\left(\frac{1}{X_2 - X_1}\right) E\left(u_2 - u_1 | X_1, X_2\right)\right] \\ &\quad \text{(by law of iterated expectations)} \\ &= \beta_1 + E\left[\left(\frac{1}{X_2 - X_1}\right) \left\{E\left(u_2 | X_2\right) - E\left(u_1 | X_1\right)\right\}\right] \\ &\quad \text{(since } (X_1, u_1) \text{ is independent of } (X_2, u_2)\text{)} \\ &= \beta_1 + 0 \\ &\quad \text{(by Assumption A1)} \\ &= \beta_1\end{aligned}$$

Hence, $\widetilde{\beta}_1$ is an unbiased estimator of β_1.

(b) To begin, note that by the same argument as above, we have

$$E\left[\widetilde{\beta}_1 | X_1, .., X_n\right] = \beta_1 + \left(\frac{1}{X_2 - X_1}\right) E\left(u_2 - u_1 | X_1, X_2\right) = \beta_1$$

Hence,

$$\begin{aligned}
& Var\left(\widetilde{\beta}_1|X_1,..,X_n\right) \\
= & E\left[\left(\widetilde{\beta}_1 - E\left[\widetilde{\beta}_1|X_1,..,X_n\right]\right)^2|X_1,..,X_n\right] \\
= & E\left[\left(\widetilde{\beta}_1 - \beta_1\right)^2|X_1,..,X_n\right] \\
= & E\left[\left(\frac{u_2-u_1}{X_2-X_1}\right)^2|X_1,..,X_n\right] \\
= & \frac{1}{(X_2-X_1)^2}E\left[\left(u_2^2 - 2u_1u_2 + u_1^2\right)|X_1,..,X_n\right] \\
= & \frac{1}{(X_2-X_1)^2}\left(E\left[u_2^2|X_2\right] - 2E\left[u_1|X_1\right]\left[u_2|X_2\right] + E\left[u_1^2|X_1\right]\right) \\
& \text{(by Assumptions A1 and A2)} \\
= & \frac{2\sigma_u^2}{(X_2-X_1)^2} \\
& \text{(by Assumptions A3 and A5)}
\end{aligned}$$

On the other hand, the conditional variance for the OLS estimator is given by

$$\begin{aligned}
& Var\left(\widehat{\beta}_1|X_1,..,X_n\right) \\
= & \frac{\sigma_u^2}{\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2},
\end{aligned}$$

where

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i.$$

Note that $\widetilde{\beta}_1$ is a linear, unbiased estimator. Hence, by the Gauss-Markov theorem, we would expect that

$$Var\left(\widehat{\beta}_1|X_1,..,X_n\right) \leq Var\left(\widetilde{\beta}_1|X_1,..,X_n\right)$$

In fact, intuitively, we would expect the OLS estimator $\widehat{\beta}_1$ to be strictly more efficient since it uses more data (and, thus, more information) than $\widetilde{\beta}_1$.

Below, we will show explicitly that

$$Var\left(\widehat{\beta}_1|X_1,..,X_n\right) \leq Var\left(\widetilde{\beta}_1|X_1,..,X_n\right)$$

and that the inequality is strict, provided that some additional assumption (to be specified below) is satisfied. We make use of the following result.

Claim: Let a and b be real numbers; then

$$|ab| \leq \frac{a^2}{2} + \frac{b^2}{2}. \tag{0.1}$$

Proof:

Consider two cases. First, suppose that $ab < 0$. In this case, note that

$$(a+b)^2 = a^2 + 2ab + b^2 \geq 0$$

from which it follows that

$$a^2 + b^2 \geq -2ab = 2|ab| \geq 0 \tag{0.2}$$

Next, consider the case where $ab \geq 0$. In this case, note that

$$(a-b)^2 = a^2 - 2ab + b^2 \geq 0$$

from which it follows that

$$a^2 + b^2 \geq 2ab = 2|ab| \geq 0. \tag{0.3}$$

The desired result then follows immediately from (0.2) and (0.3).

We can apply the inequality (0.1) to show that

$$Var\left(\widehat{\beta}_1|X_1,..,X_n\right) \leq Var\left(\widetilde{\beta}_1|X_1,..,X_n\right).$$

To proceed, take

$$a = X_2 - \overline{X}_n \text{ and } b = X_1 - \overline{X}_n$$

It follows that

$$\begin{aligned}
& (X_2 - X_1)^2 \\
= \; & \left(X_2 - \overline{X}_n - \left[X_1 - \overline{X}_n\right]\right)^2 \\
= \; & \left(X_2 - \overline{X}_n\right)^2 - 2\left(X_1 - \overline{X}_n\right)\left(X_2 - \overline{X}_n\right) + \left(X_1 - \overline{X}_n\right)^2 \\
\leq \; & \left(X_2 - \overline{X}_n\right)^2 + 2\left|\left(X_1 - \overline{X}_n\right)\left(X_2 - \overline{X}_n\right)\right| + \left(X_1 - \overline{X}_n\right)^2 \\
\leq \; & 2\left(X_2 - \overline{X}_n\right)^2 + 2\left(X_1 - \overline{X}_n\right)^2 \\
& \text{(by applying inequality (0.1))}
\end{aligned}$$

It further follow that

$$\begin{aligned}
& Var\left(\widehat{\beta}_1|X_1,..,X_n\right) \\
= \; & \frac{\sigma_u^2}{\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2} \\
\leq \; & \frac{2\sigma_u^2}{2\left\{\left(X_2 - \overline{X}_n\right)^2 + \left(X_1 - \overline{X}_n\right)^2\right\}} \\
\leq \; & \frac{2\sigma_u^2}{(X_2 - X_1)^2} \\
= \; & Var\left(\widetilde{\beta}_1|X_1,..,X_n\right).
\end{aligned}$$

Moreover, note that the first inequality above will be strict, so that we will have

$$Var\left(\widehat{\beta}_1|X_1,..,X_n\right) < Var\left(\widetilde{\beta}_1|X_1,..,X_n\right)$$

as long as

$$\left(X_i - \overline{X}_n\right)^2 \neq 0$$

for at least one $i \geq 3$.

QUESTION 5

(a) Note that the R^2 value for the second regression is bigger than that of the first regression. However, this fact alone does not imply that *Educ* is a significant explanatory variable for RelPersInc, since we know that by adding more regressors, the unadjusted R^2 value may increase in finite sample even if the true coefficients on the extra regressors happened to be zero.

(b) Under homoskedasticity and normally distributed errors, we can make use of the following simplified expression for the F-statistic

$$\begin{aligned} F &= \frac{\left(R^2_{unrestricted} - R^2_{restricted}\right)/q}{\left(1 - R^2_{unrestricted}\right)/\left(n - k_{unrestricted} - 1\right)} \\ &= \frac{(0.775 - 0.621)/1}{(1 - 0.775)/(94 - 3 - 1)} \\ &= 61.6 \end{aligned}$$

Under the null hypothesis $H_0 : \beta_3 = 0$,

$$F \sim F_{1,90}$$

i.e., it has a F-distribution with 1 and 90 degrees of freedom. At the 5% significance level, the critical value is 3.95. Since

$$F = 61.6 > 3.95,$$

we reject $H_0 : \beta_3 = 0$.

(c) As long as homoskedasticity holds, the F-statistic will still have the form

$$F = \frac{\left(R^2_{unrestricted} - R^2_{restricted}\right)/q}{\left(1 - R^2_{unrestricted}\right)/\left(n - k_{unrestricted} - 1\right)}.$$

However, if the errors are not necessarily normally distributed, then we need to use a large sample approximation. In this case, under $H_0 : \beta_3 = 0$

$$F \overset{a}{\sim} \chi^2_1/1 \equiv \chi^2_1.$$

At the 5% significance level, the critical value is 3.84. Since

$$F = 61.6 > 3.84,$$

we still reject $H_0 : \beta_3 = 0$, so we reach the same decision as in part (b). Note that the critical values for parts (b) and (c) are not terribly different since we do have a pretty large sample; hence, it is perhaps not all that surprising that we end up reaching the same decision in both cases.