ECON 4230 Intermediate Econometric Theory
Solutions to Problem Set #1

Directions. Make sure your answers are typed and tables are appropriately formatted. Turn in a hard copy at the beginning of class on the date above.

#1. For the two-variable regression model, derive the ordinary least squares slope estimator for two cases: (1) with an intercept; and (2) when the intercept is restricted to be zero. Compare the slope estimator for the two cases and comment on the results.

SOLUTION:

For case (1), the OLS problem is

$$\min_{\hat{\beta}_1, \hat{\beta}_2} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2.$$

The first-order conditions (FOCs) are

$$\frac{\partial(SSR)}{\partial\hat{\beta}_1} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0 \text{ and}$$

$$\frac{\partial(SSR)}{\partial\hat{\beta}_2} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)X_i = 0.$$

Distributing the summation operator for the first FOC and dividing by $-2n$ gives

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}.$$

Substituting this expression into the second FOC and light algebra gives (I'll do this in class) …

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

For case (2), the OLS problem is

$$\min_{\hat{\beta}_2} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_2 X_i)^2.$$

The FOC is

$$\frac{\partial(SSR)}{\partial\hat{\beta}_2} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_2 X_i)X_i = 0.$$

Solving for $\hat{\beta}_2$ gives

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2}.$$

#2. Verify the second-order conditions for case (2) above when the intercept is suppressed.

SOLUTION:

For case (2), the second-order condition is

$$\frac{\partial^2(SSR)}{\partial(\hat{\beta}_2)^2} = 2\sum_{i=1}^{n} X_i^2 > 0,$$

which is clearly satisfied given the Classical assumptions.

#3. Using Table 2.10 from our textbook, estimate three separate regression models to predict family income using (*i*) critical reading, (*ii*) mathematics, and (*iii*) writing SAT scores. Which model is the best predictor of family income? Defend your answer.

SOLUTION:

The regression results are shown in Table 1:

Table 1. OLS regression results for household income

Variable	Math	Reading	Writing
Intercept	-589.11***	-499.90***	-514.07***
	(54.50)	(68.25)	(58.15)
SAT category	1.297***	1.141***	1.193***
	(0.109)	(0.140)	(0.121)
R^2	0.9462	0.8929	0.9235

Notes. Sample size = 10. Standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01. The significance test for the intercept is two-tailed while the significance test for the SAT category is a one-tailed test.

Based on the goodness-of-fit, the best model for predicting household income involves math SAT scores.

#4. Gujarati and Porter, 5th Edition, Exercise 3.10.

SOLUTION:

Let the deviation in the means for X and Y be represented by

$$x = X - \bar{X} \text{ and } y = Y - \bar{Y}.$$

Applying the standard OLS formula using data on x and y, and noting that $\bar{x} = \bar{y} = 0$, gives

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x} = 0 \text{ and } \hat{\beta}_2 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}.$$

This is the same OLS formula for the slope coefficient. This makes sense because measuring the data as deviation from the means should have no impact on the slope of the best-fitting regression line. It is basically a simple rescaling of the horizontal and vertical axes.

#5. Go to the Federal Reserve Economic Database (FRED) and collect U.S. time series data on current-dollar consumption and GDP over the sample period 1980:1-2017:4. Estimate a two-variable Keynesian consumption function and address the following items:

a) Report the regression results and comment on the results.
b) Provide a graph of the residuals over time.
c) Using the graph in part (b) and economic theory to guide your answers, which Classical assumptions are most dubious?
d) What is the prediction level of consumption in the first two quarters of 2017? Did the model do a good job of predicting consumption?
e) What is the R^2 value? How do you interpret the number and why do you think it is so high?
f) Re-estimate the model using growth rates of consumption and GDP. Graph the residuals and contrast the results with the consumption function estimated in levels. What is the predicted level of consumption for the first two quarters of 2017 using the model in growth rates? Contrast with the results in part (d). Which model do you prefer: levels or growth rates?
g) Finally, perform a hypothesis test that the government spending multiplier equals 10.

SOLUTION:

a) The results from the estimated Keynesian consumption function (in levels) are given in Table 2. The estimated MPC is approximately 0.70, statistically significant, and the model has a very, very good fit.

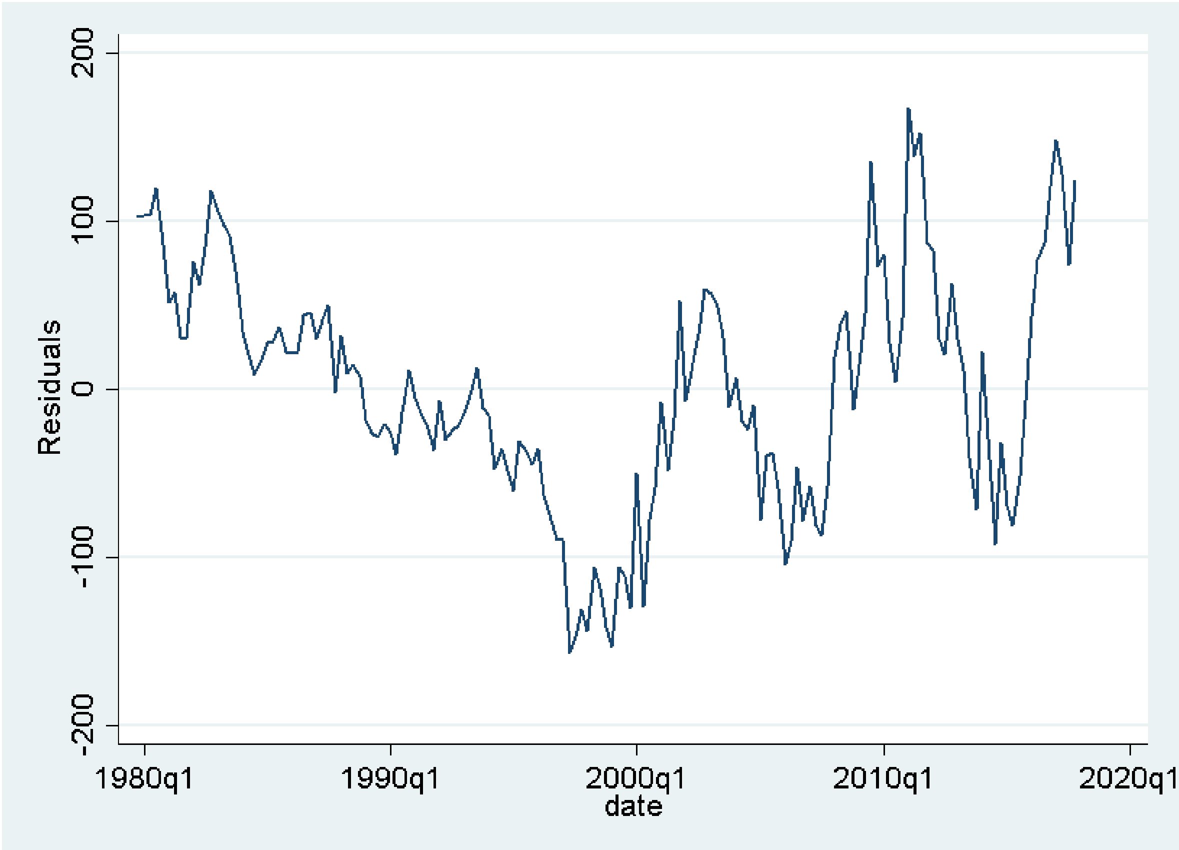
Table 2. OLS estimated Keynesian consumption function

Variable	Levels	Growth Rates
Intercept	-364.21***	0.0194***
	(12.81)	(0.0031)
GDP	0.7035***	0.6951***
	(0.0011)	(0.0495)
R^2	0.9996	0.5682

Notes. Dependent variable is U.S. consumption expenditures. Sample period 1980:1 – 2017:4. Standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01. The significance test for the intercept is two-tailed while the significance test for GDP is a one-tailed test.

b) The plot of the residuals (in levels) is given by Figure 1.

Figure 1. Residuals from an OLS estimated Keynesian consumption function (in levels)



c) The plot of the residuals shows clear signs of positive autocorrelation and possibly omitted-variable bias as there is a downward trend in the residuals between 1980 and 2000.

d) The predicted levels of consumption for the first two quarters of 2017 is $13,043.79 and $13,178.99 billion, respectively. The actual levels of consumption are $13,191.56 and $13,307.02 billion. Both predictions are too low, but off by only 1%.

e) The value of the R^2 is given in Table 2. It is probably high because (Keynesian) economic theory states that current consumption and income exhibit a stable, linear relationship. It is also high because both follow a strong trend.

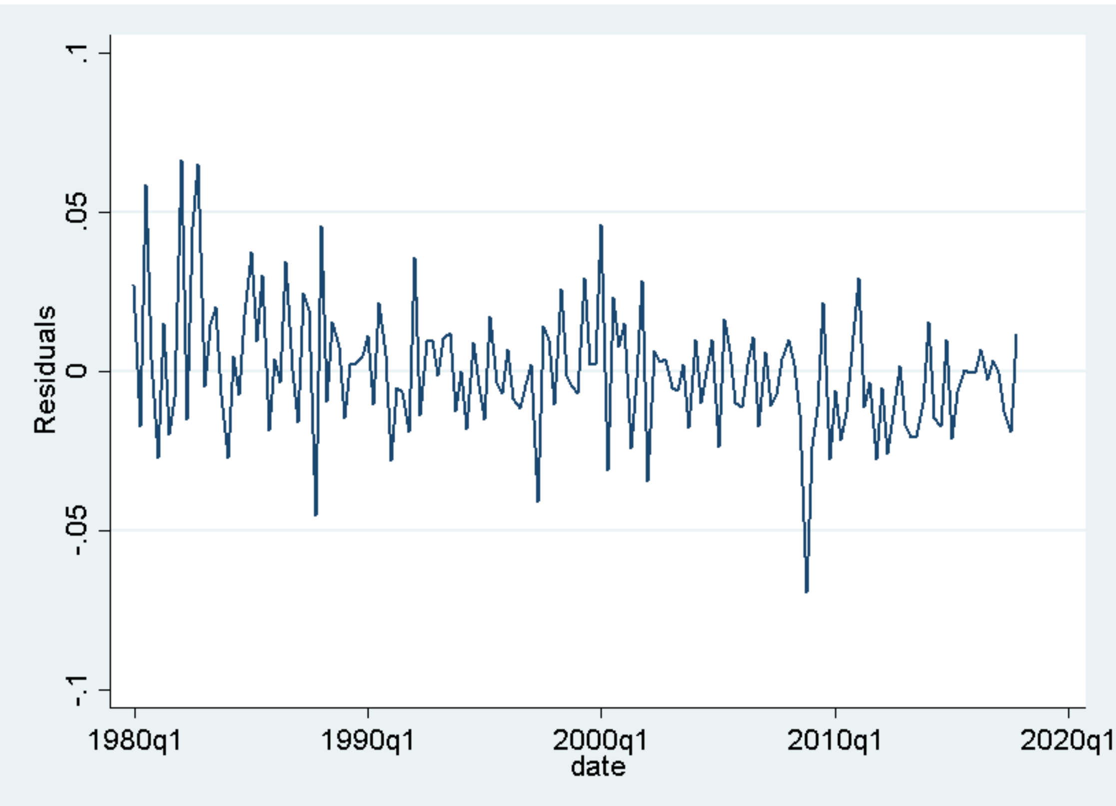
f) The results from the model (in growth rates) are also shown in Table 2 and the residuals are shown in Figure 2 below. The predicted annualized growth rates for first and second quarter consumption are 4.18% and 4.75%, respectively. Dividing by 4 and putting in decimal form, the predicted levels of consumption expenditures are…

$$\hat{C}_{2017:1} = C_{2016:4} \times (1 + \hat{g}_{2017:1}) = 13,056.90 \times (1 + 0.01045) = 13,193.34 \text{ and}$$

$$\hat{C}_{2017:2} = C_{2017:1} \times (1 + \hat{g}_{2017:2}) = 13,191.56 \times (1 + 0.01188) = 13,348.28.$$

The predicted values using the model in growth rates are closer to the actual values, but this is a little unfair given the predictions use the actual values in the previous quarter. The goodness-of-fit measures are not comparable because the dependent variables are different in the two models. I would probably prefer the model in growth rates since the residuals show less evidence of autocorrelation.

Figure 2. Residuals from the OLS Keynesian consumption function (in growth rates)



g) A government spending multiplier equal to 10 is equivalent to an MPC (β_2) of 0.9. The null and alternative hypotheses are …

H_0: $\beta_2 = 0.9$
H_1: $\beta_2 \neq 0.9$.

For a significance level of $\alpha = 0.05$ and a degrees of freedom of $n - k = 152 - 2 = 150$, the critical values are ± 1.96. The t statistic (regression in levels) is

$$t = \frac{0.7035 - 0.9}{0.0011} = 178.64,$$

which leads to a rejection of the null hypothesis that the government spending multiplier is 10.