ECN 102: Analysis of Economics Data

Homework 4 Answer Key

Remy Beauregard

Question 1: Log interpretation

AED (2022) Ch.9 Question 8

For each of the following, say what model (i.e. write down $\hat{y} = b_1 + b_2 x$) leads to the given interpretation of the slope coefficient (your intercept term will be arbitrary):

a) A one unit change in x is associated with a 0.3 unit change in y

$$\hat{y} = b_1 + 0.3x$$

b) A one unit change in x is associated with a 0.3 percent change in y

$$\widehat{\ln y} = b_1 + 0.003x$$

c) A one percent change in x is associated with a 0.3 percent change in y

$$\widehat{\ln y} = b_1 + 0.3 \ln(x)$$

d) A one percent change in x is associated with a 0.3 unit change in y

$$\hat{y} = b_1 + 30 \ln(x)$$

Question 2: Population model

a) What are the four population assumptions for multivariate regression? How do these differ from our bivariate population assumptions? **Our assumptions are largely the same, except now include regressors $x_2, ..., x_k$ instead of just a single regressor. Our interpretation for linearity, unbiasedness, homoskedasticity, and independence now must hold across all regressors simultaneously.**

b) Explain in words what it means for our multivariate OLS estimator to be:

1. unbiased $E[b_j] = \beta_j$
2. consistent $V[b_j] \to 0$ as $n \to \infty$
3. BLUE **Our estimator has the minimum variance of all linear unbiased estimators in its class**

Question 3: Estimation

a) Explain two ways our multivariate OLS coefficient β_j could be more precisely estimated compared to an alternative setting. **Options include: 1) Having a better model fit to reduce RMSE, 2) having a larger sample size n, or 3) having a regressor that is less explained by our other regressors (so $\tilde{x}_{ji}$ is larger)**

Suppose we run a multivariate regression in Stata with 4 regressors and a constant with $n = 50$. Interpret the following output:

b) t = 3.3 for b_2 **Since** $3.3 > 2.01$**, we can reject the default null that** $\beta_2 = 0$ **at** $\alpha = 0.05$**.**

```
di invttail(50-5,0.025)
```

```
2.0141034
```

c) P>|t| = 0.09 for b_3 **Since 0.09 > 0.05, we fail to reject our default null that** $\beta_3 = 0$ **at** $\alpha = 0.05$ **and thus cannot say our coefficient is different from zero.**

d) Prob > F = 0.04 **Since the p-value of our default F-test for overall significance is below** $\alpha = 0.05$**, we have sufficient evidence to reject the null that all regressors are not different from zero and assert that at least one is non-zero.**

For the overall test yielding $F(v_1, v_2) = 2.74$,

e) what will v_1 and v_2 be equal to?

$$v_1 = k - 1 = 5 - 1 = 4$$
$$v_2 = n - k = 50 - 5 = 45$$

f) how would we find $F^*_{v_1,v_2,\alpha}$?

```
di invFtail(4,45,0.05)
```

```
2.5787392
```

Question 4: Adding regressors

Suppose we start with a given multivariate regression and then want to add one more regressor. In each case, discuss (1) what each result means and then (2) whether we have sufficient proof to say our coefficient is statistically significantly different from zero at $\alpha = 0.05$.

a) Including the regressor increased our R-squared **Our model fit improved; not enough evidence.**

b) Including the regressor increased our adjusted R-squared **Our measure of model fit that penalizes additional regressors improved; still not enough evidence.**

c) Including the regressor decreased our RMSE **Our non-standardized measure of model fit improved; still not enough evidence.**

d) A t-test of our regressor yielded $p = 0.10$ **Our p-value is greater than α; cannot reject the null.**

If we instead conducted an F-test of our single regressor,

e) what would be the p-value for that computed F-statistic? How do we know? **Since our F-statistic for a test of a single regressor will be exactly equal to the square of the t-statistic (and the resulting F-distribution the exact square of the T-distribution) our p-values will be identical: $p = 0.10$.**

f) what would our F-statistic be equal to? Why? **While we may not know our t-statistic for a given regressor in this case, we know the F-statistic for the test of a single regressor will be equal to the square of the t-statistic from a 2-sided test of association on that regressor; $F_{v_1,v_2,\alpha} = (t_{v_1,\alpha})^2$**