Name: Answers

EC380: Exam 2

Scott Hiller

Spring 2024

For multiple choice questions, choose the answer that *best* answers the question. For other questions, show all of your work and explain where appropriate. Good luck!

Multiple Choice Questions (5 points each)

1. If you are testing the statistical significance of a parameter, and Stata gives you a p-value of 0.06 for that parameter, it must be that

   (a) the |t-statistic| is ***less than*** the t-critical value at $\alpha = 0.05$ level of significance.

   (b) the |t-statistic| is ***greater than*** the t-critical value at $\alpha = 0.05$ level of significance.

   (c) the |t-statistic| is ***equal to*** the t-critical value at $\alpha = 0.05$ level of significance.

   (d) Impossible to tell!

2. Which test statistic should one use when comparing a restricted and an unrestricted model with *different* dependent variables?

   (a) $F = \frac{R^2/m}{(1-R^2)/(n-k)}$

   (b) $F = \frac{(R^2_{UR}-R^2_R)/m}{(1-R^2_{UR})/(n-k)}$

   (c) $F = \frac{(RSS_R-RSS_{UR})/m}{RSS_{UR}/(n-k)}$

   (d) $t = \frac{\hat{\beta}-\beta}{se(\hat{\beta})}$

3. The assumption that $u_i \sim N(0, \sigma^2)$ implies

   (a) $E(u|x) = 0$

   (b) the errors are homoskedastic

   (c) symmetric distribution of the errors around their mean

   (d) all of the above

   (e) none of the above

4. The "dummy variable trap" is an example of

   (a) omitted variable bias

   (b) pure heteroskedasticity

   (c) perfect multicollinearity

   (d) locking yourself in your own car

5. Suppose the dependent variable in your model takes either the value of zero or one for every observation. Specifically, you are regressing college attendance (yes or no) on family income. OLS is inappropriate for estimation in this case because

   (a) OLS allows for a greater than 100% predicted probability of attending college.

   (b) OLS fits a straight line to a nonlinear relationship between family income and college attendance.

   (c) Neither a nor b.

   (d) Both a and b.

6. A die roll has a uniform distribution (the probability of rolling any particular number is 1/6). However, collecting 1000 sample means for 1000 sets of rolling a die 10 times will yield an approximately normal distribution of those sample means. This is an example of

   (a) an unbiased estimator

   (b) the Gauss-Markov theorem

   (c) the Central Limit theorem

   (d) a consistent estimator

7. If your model exhibits a high degree of multicollinearity,

   (a) it will be easier to find statistical significance of a parameter.

   (b) it will be more difficult to find statistical significance of a parameter.

   (c) you will be more likely to reject a null hypothesis of $\beta = 0$.

   (d) you are likely to have a low R^2 but many individually significant parameters.

8. (**10 points**) Test whether the effect of education on growth is ***exactly the opposite*** the effect of child mortality on growth.

$Growth_i = \beta_0 + \beta_1 Educ_i + \beta_2 Mort_i + u_i$

(a) Set up the null and alternate hypotheses.

$H_0: \beta_1 = -\beta_2$

$H_a: \beta_1 \neq -\beta_2$

(b) Write down the *unrestricted* model.

$Growth_i = \beta_0 + \beta_1 Educ_i + \beta_2 Mort_i + u_i$

(c) Write down the *restricted* model.

$Growth_i = \beta_0 + \beta_1 Educ_i - \beta_1 Mort_i + u_i$

$= \beta_0 + \beta_1 (Educ_i - Mort_i) + u_i$

(d) Which test statistic would you use to test your hypothesis?

$$F = \frac{\frac{R^2_{UR} - R^2_R}{m}}{\frac{1-R^2_{UR}}{n-k}}$$

(e) Suppose $R^2 = 0.35$ for your restricted model. What do you conclude about your hypothesis?

$$F = \frac{\frac{R^2_{UR} - .35}{1}}{\frac{1-R^2_{UR}}{n-k}}$$

If R^2_{UR} is sufficiently large s.t. $F > 2.2 \Rightarrow$ reject H_0 at $\alpha = .05$.

1. **Model Specification (16 points)** Consider the following regression model:

$$Price_{Tequila_i} = \beta_0 + \beta_1 Price_{Agave} + \beta_2 Price_{Limes} + \beta_3 Price_{Rum} + u_i$$

(a) Suppose you realize that $Price_{Limes}$ and $Price_{Rum}$ are ***irrelevant*** but decide to leave them in the regression. $Price_{Rum}$ is correlated with $Price_{Agave}$, while $Price_{Limes}$ is NOT correlated with $Price_{Agave}$.

i. Is $\hat{\beta}_1$, the coefficient on $Price_{Agave}$, an unbiased estimate of β_1? EXPLAIN.

Yes, unbiased because $cov(Price_{agave}, u) = 0$

Irrelevant vars. do not affect the estimate.

ii. Suppose you find the coefficient on $Price_{Agave}$ to be statistically significant at the 5% level. Do you trust your finding that $Price_{Agave}$ has a systematic impact on the price of tequila? WHY OR WHY NOT?

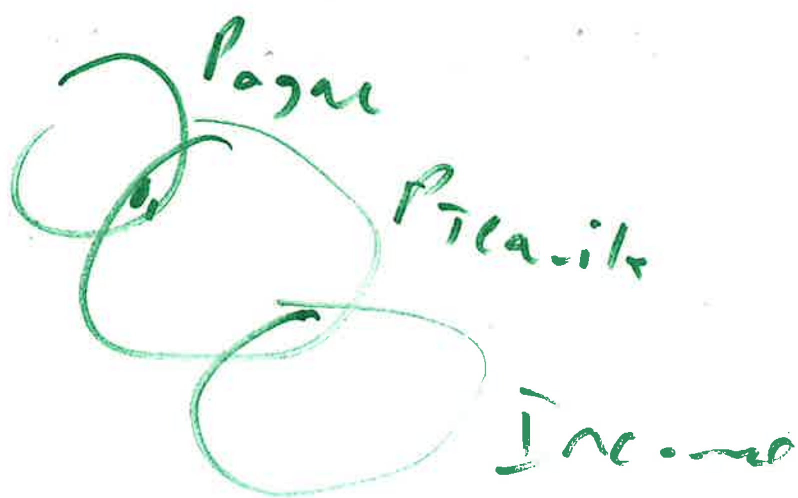
Yes, irrelevant but correlated vars introduce multicollinearity. This makes finding significance more difficult.

$$var(\hat{\beta}_1) = \frac{\frac{\sum u_i^2}{n-k}}{\sum_i (x_i - \bar{x})^2 (1-r^2)}$$

$(n-k)\downarrow \Rightarrow var(\hat{\beta}_1)\uparrow \Rightarrow$ t-stat $\downarrow$

iii. Suppose you realize that income is a **relevant** variable that you have omitted from the model. Income is uncorrelated with $Price_{Agave}$. Do you trust your finding that $Price_{Agave}$ has a systematic impact on the price of tequila? WHY OR WHY NOT?

Yes, we have an omitted var, which will increase our variance. However the omitted var is not correlated with Pagave, so no bias in the coeff.



2. You are using annual data 1990-2010 for the US to test for a relationship between quality of new music (x) and CD sales (y). Digital music probably changed this relationship around 1999.

How would you alter your basic model to test that hypothesis? Explain.

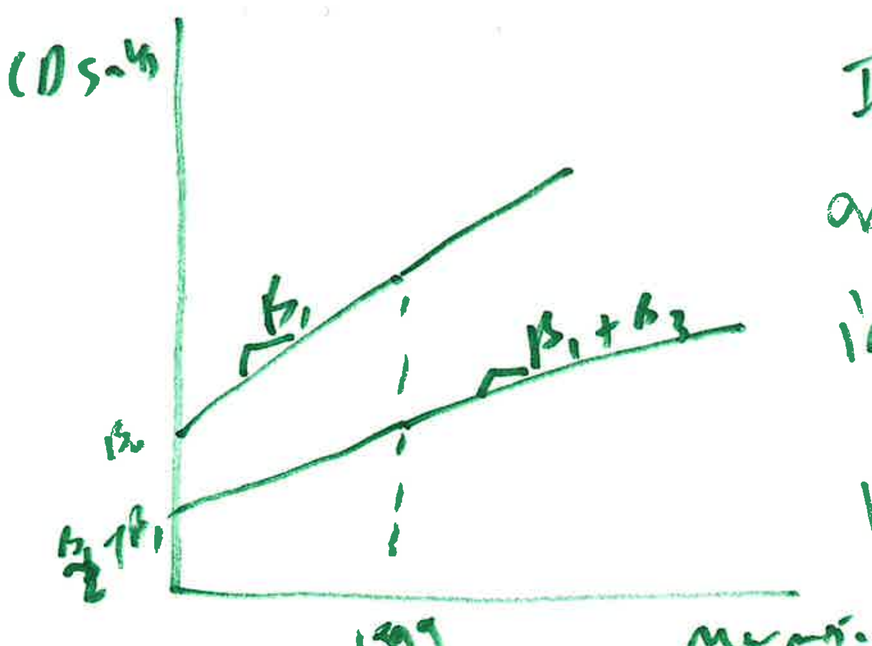
Basic model:

$$CDsales_t = \beta_0 + \beta_1 newmusic_t + u_t$$

Alteration:

$$CDsales_t = \beta_0 + \beta_2 1999 + \beta_1 newmusic_t + \beta_3 newmusic_t \cdot 1999 + u_t$$

where 1999 is a dummy var that $=1$ if $t \geq 1999$.



I would anticipate the intercept and slope to be less after 1999.

3. **Dummy Variables (12 points)** Suppose you run the following regression relating happiness to how many days of sunshine a state has, and its income.

$$Happiness_{it} = \beta_1 + \beta_2 DaysSunny_{it} + \beta_3 Income_{it} + u_i$$

Suppose you suspect that the *average happiness* varies depending on the region, specifically, North, South, East, and West.

(a) How would you allow for such "regional" differences in the model? Write the new model specification. Be sure to define any new variables.

$N_i = 1$ if north $S_i = 1$ if south

$E_i = 1$ if east ≥ 0 otherwise

We exclude west as our benchmark.

$$Happiness_{it} = \beta_0 + \beta_1 N_i + \beta_2 S_i + \beta_3 E_i + \beta_4 DaysSunny_{it} + \beta_5 Income_{it} + u_i$$

(b) Write down the estimated fitted line for each region.

$\hat{Happiness}^{South} = \hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_4 DaysSunny_{it} + \hat{\beta}_5 Income_{it}$

$\hat{Happiness}^{North} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_4 DaysSunny_{it} + \hat{\beta}_5 Income_{it}$

4. (15 points) You are using 1000 years of **annual** data for the US on real GDP and inflation to regress $Inflation_t = \beta_0 + \beta_1 GDPgap_t + u_t$. You suspect there is first-order autocorrelation present.

(a) Write an equation describing the AR(1) process.

$u_t = \rho u_{t-1} + \varepsilon_t$

(b) Suppose you run a Durbin-Watson test for AR(1) and get a d-statistic of 0.2. What do you conclude about AR(1) in the model? Explain thoroughly.

There is positive autocorrelation present. The numerator of the DW is $\sum(\hat{u}_t - \hat{u}_{t-1})^2$, the closer to zero this is, the more related is u_t to u_{t-1}, indicating positive autocorrelation.

(c) Suppose you choose to correct your model for AR(1) by using feasible generalized least squares.

i. Describe one way to estimate ρ.

Because I already have d, I could use it, $d \simeq 2(1-\hat{\rho})$, therefore $\hat{\rho} \simeq 1 - \frac{d}{2}$ or $\hat{\rho} \simeq .9$

ii. Use this estimate $\hat{\rho}$ to transform the model to rid it of AR(1) and show that it is free of autocorrelation.

$$\text{Inflation}_t - \hat{\rho}_0 \text{Inflation}_{t-1} = (\beta_0 - .9\beta_0) + (\beta_1 X_t - .9\beta_1 X_{t-1}) + (u_t - .9u_{t-1})$$

$$u_t = .9u_{t-1} + \varepsilon_t$$

$$u_t - .9u_{t-1} = \varepsilon_t$$

$$\text{Inflation}^* = \beta_0^* + \beta_1^* X_t^* + \varepsilon_t$$

No autocorrelation.

5. **Bonus** Suppose you are trying to figure out what characteristics contribute to the probability of passing econometrics. You estimate the following model using data on 5,000 people:

$$passing_i = \beta_0 + \beta_1 math_i + \beta_2 attendance_i + \beta_3 iq_i + \beta_4 crazyenoughtoenjoyeconometrics_i + u_i$$

(a) What type of test would you conduct to determine whether craziness and attendance really belong in the model?

3

i. F-test using $F = \frac{R^2/m}{(1-R^2)/(n-k)}$

ii. F-test using $F = \frac{(R^2_{UR}-R^2_R)/m}{(1-R^2_{UR})/(n-k)}$

iii. F-test using $F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)}$

iv. t-test using $t = \frac{\hat{\beta}-\beta}{se(\hat{\beta})}$

(b) What kind of variable is the dependent variable? How would you estimate this regression? Why?

3

I would estimate with a logit or probit because the dep. var. is binary, meaning we can predict values outside of 0 and 1.

T and F critical values

T value at $\alpha = .05$ is 1.684 (one tail)
T value at $\alpha = .05$ is 2.021 (two tail)
F value at $\alpha = .05$ is 2.2