**ECO220Y1Y, APRIL 2016, FINAL EXAM: SOLUTIONS**

**(1) (a)** In the pre enforcement time period, on average countries with a corruption index that is 0.1 unit higher (index ranges from -2.5 to 1.6) have annual unpaid parking violations per diplomat that are about 3.8% higher. (Note: A one unit change in the corruption index is not a marginal change so the approximation that a 1 unit higher value is associated with 38% higher violations is much rougher.)

**(b)**

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{0.02/1}{(1-0.02)/(110-1-1)} = 2.2$$

The critical value (cv) from the F table for $\alpha$ = 0.10 is 2.8: the test statistic is not large enough to reject the null in favor of the research hypothesis. In the post enforcement period we cannot rule out the possibility that there is no relationship between the natural log of violations and the corruption index. However, that does *not* prove that there is no relationship. In contrast, in the pre enforcement period we did find a larger (positive) correlation that is statistically significant which means that we can reject the null that there is no relationship.

**(c)** The intercept of 24.3 means that across the 110 countries in the pre-enforcement period (11/1997 – 11/2002) there were on average 24.3 unpaid annual NYC parking violations per diplomat. This is ridiculously large: the average diplomat was raking up over 24 parking tickets (and not paying them) each year. The slope of -23.8 means that across the 110 countries there were on average 23.8 fewer unpaid annual NYC parking violations per diplomat in the post-enforcement period (11/2002 – 11/2005) compared to the pre-enforcement period: a HUGE decline from an average of 24.3 unpaid violations per diplomat per year to just 0.5.

**(d)** The graph shows an extreme violation of the homoscedasticity assumption. In the pre-enforcement period there is a lot of scatter: some countries have very large numbers of parking violations per diplomat while others have very small numbers. In contrast, in the post-enforcement period there is far less scatter around the line: all countries are much more similar in the number of parking violations per diplomat. The value of $s_e$ is NOT a good measure of the amount of scatter when there is heteroscedasticity: it clearly understates the amount of scatter in the pre-enforcement period and overstates the amount of scatter in the post-enforcement period.

**(e)** On average countries with a GNI per capita that is 10% higher have annual parking violations per diplomat that are 2% lower after controlling for the level of corruption (the corruption index) and whether it is the pre or post enforcement time period. The P-value of 0.022 means that we can reject the null hypothesis $H_0: \beta_3 = 0$ in favor of $H_1: \beta_3 \neq 0$ at the 5% significance level, but not the 1% significance level. (In other words, we have good evidence to suggest that ln_gni_pc_2005 adds to the model after controlling for corruption and an indicator for the post-enforcement period, but it is not overwhelming evidence.)

**(f)** In both Regression #4 and #5 there is a huge negative and highly statistically significant coefficient on the dummy variable for the post-enforcement period: this reflects the dramatic drop in unpaid parking violations once legal enforcement actions were begun in November of 2002. [Note that a one unit change in this variable, which only has a range of 0 to 1, is *not a marginal change* that is why our usual interpretation of coefficients when logs are involved breaks down: obviously we cannot have a 375 percent drop in parking violations even if it is a huge drop! See note below for more detail.] However, the role of corruption norms differs between these two regressions: in Regression #4 it is positive and statistically significant suggesting diplomats from more corrupt countries on average accumulate more unpaid parking tickets but in Regression #5 this coefficient is not statistically significant, which means we cannot rule out zero effect of corruption norms after controlling for GNI per capita and the pre versus post enforcement period. Given that these data are observational and corruption and GNI per capita are strongly negatively correlated, it is not surprising that we see swings in the corruption regression coefficients depending on whether GNI is controlled for or not. In Regression #4, GNI is an important lurking (unobserved/omitted/confounding) variable that causes (endogeneity) bias in the corruption coefficient. Hence, it is unclear whether corruption norms have a separate influence, but there is no denying that there was a huge drop in unpaid parking tickets once legal enforcement began.

**NOTE:** (A reminder that the interpretation of a coefficient when logs are involved is approximate) In *Logarithms in Regression Analysis* you learned very common sets of *approximate* interpretations of coefficients when logarithms are involved (on either y and/or x). These approximations are excellent for small changes but not for large ones (as arises in this example). In a case like this where the policy variable of interest is a dummy variable – a dummy for the post enforcement period – we can compute the exact interpretation of the coefficient like this:

For pre-enforcement, dummy in Reg. #4 is zero: $\ln(y0_{it}) = \alpha + \beta_c * corruption_{it} + \varepsilon_{it}$

For post-enforcement, dummy in Reg. #4 is one: $\ln(y1_{it}) = \alpha + \delta * post\_enforcement + \beta_c * corruption_{it} + \varepsilon_{it}$

The difference between post and pre-enforcement: $\ln(y1_{it}) - \ln(y0_{it}) = \delta$

$$\delta = \ln\left(\frac{y1_{it}}{y0_{it}}\right) = \ln\left\{1 + \frac{y1_{it} - y0_{it}}{y0_{it}}\right\}$$

$$\frac{y1_{it} - y0_{it}}{y0_{it}} = e^{\delta} - 1$$

In this case we have $\hat{\delta} = -3.747768$ so the estimated change is -0.9764 (=e^-3.747768 – 1), which is a 98% drop in the violations per diplomat in the post enforcement period compared to the earlier period. Note that this matches what we'd expect from Regression #3: it says violations were 23.8 lower in the post enforcement period compared to 24.3 in the earlier period: -0.9794 = (0.5 – 24.3)/24.3, a 98% drop. However, suppose we had gotten a more marginal $\hat{\delta} = -0.05$, which we would *approximately* interpret as a 5% drop in the post-enforcement period. If we use the exact formula above we get $\frac{y1_{it}-y0_{it}}{y0_{it}} = e^{\delta} - 1 = e^{-0.05} - 1 = -0.049$. Obviously the approximation is excellent in such a case because that is a marginal change. How marginal does it need to be? Well this is subjective. On page 94 of *Mastering `Metrics: The Path from Cause to Effect* (2015), the authors (prominent professors of economics at MIT and LSE) suggest a $\hat{\delta}$ between -0.2 and 0.2 is sufficiently marginal to just use the standard approximations you learned in *Logarithms in Regression Analysis*.

**(2)** $SE[\hat{P}] = \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = \sqrt{\frac{0.115(1-0.115)}{200}} = 0.0226 \qquad SE[\hat{P}] = \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = \sqrt{\frac{0.11(1-0.11)}{100}} = 0.0313$

The first white bar shows the callback rate for the 200 job ads that received an application from a fictitious Asian applicant with "no whitening" whereas the first gray bar shows the callback rate for the subsample of 100 job ads that received such an application and had explicit pro-diversity language. Because the sample size is much smaller for the gray bar we'd expect a bigger SE to reflect the greater sampling error, which is consistent with what Figure 2 shows.

**(b)** $\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = 0.11 \pm 1.96\sqrt{\frac{0.11(1-0.11)}{100}} = 0.11 \pm 1.96 * 0.0313 = 0.11 \pm 0.06$; LCL=0.05; UCL=0.17

$\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = 0.22 \pm 1.96\sqrt{\frac{0.22(1-0.22)}{100}} = 0.22 \pm 1.96 * 0.0414 = 0.22 \pm 0.08$; LCL=0.14; UCL=0.30

Hence the bands at the top of the bar would spread from 0.05 to 0.17 compared to 0.08 to 0.14 and from 0.14 to 0.30 compared to 0.18 to 0.26: in other words, the bands would be about twice as wide. This would give people the impression that these estimates are very noisy (i.e. very affected by sampling error) and that the overlapping intervals mean no statistically significant differences. However, we cannot check for statistically significant differences by checking if CI's overlap: that would often lead to the incorrect conclusion of no statistically significant differences even when there are some. Instead we must do hypothesis testing. (This will be illustrated in the next part to this question.)

**(c)** Call group 2 the whitened Asian applicants and group 1 the non-whitened Asian applicants.

$$H_0: p_2 - p_1 = 0$$

$$H_1: p_2 - p_1 > 0$$

$$\bar{P} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{0.11 * 100 + 0.22 * 100}{100 + 100} = 0.165$$

$$z = \frac{\hat{P}_2 - \hat{P}_1}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1} + \frac{\bar{P}(1-\bar{P})}{n_2}}} = \frac{0.22 - 0.11}{\sqrt{\frac{0.165(1-0.165)}{100} + \frac{0.165(1-0.165)}{100}}} = \frac{0.11}{0.0525} = 2.10$$

$$P - value = P(Z > 2.10) = 0.5 - 0.4821 = 0.0179$$

Hence at a 5% (or 10%) significance level there is sufficient evidence to conclude that whitening *causes* a higher callback rates for Asian applicants even when applying to ads that have an explicit pro-diversity message. However, the evidence does not quite meet a 1% burden of proof: while the point estimate of the difference is huge – 11 percentage points, a 22% versus 11% call back rate – the sample sizes are not that big which means there is a fair bit of sampling error. [However, note that there is still a statistically significant difference at a 5% level which is why putting ME's around the point estimates as suggested in Part (b) can be misleading.]

**(d)**

| | | | | |
|---|---|---|---|---|
| ***Panel A:*** *All job ads, both races combined* | | | | |
| Callback rate | **0.1075** (**0.0155**) | 0.187 (0.011) | **0.0795** | 3.68 |
| ***Panel B:*** *Job ads with pro-diversity language, both races combined* | | | | |
| Callback rate | 0.110 (0.022) | **0.2** (**0.0163**) | **0.09** | 2.88 |

*Panel A*: $\hat{P}_1 = \frac{0.10*200+0.115*200}{200+200} = 0.1075; SE(\hat{P}_1) = \sqrt{\frac{0.1075*(1-0.1075)}{400}} = 0.01549$; Difference = 0.0795

*Panel B*: $\hat{P}_2 = \frac{0.15*100+0.20*100+0.25*100+0.18*100+0.20*100+0.22*100}{100+100+100+100+100+100} = 0.2; SE(\hat{P}_1) = \sqrt{\frac{0.2*(1-0.2)}{600}} = 0.01633$; Difference = 0.09

**(e)** The patterns for black applicants and Asian applicants are similar – as shown in Figures 1 and 2 – so combining the two races will double the sample sizes, which reduces sampling error, gives smaller standard errors, larger z test statistics (in absolute value) and smaller P-values, which means stronger evidence. However, combining the three levels of whitening that range from mild to extensive would weaken the evidence compared to comparing the extremes of no whitening versus the most whitening. In other words, when we pool the three levels of whiting it lowers the average callback rate and makes it less different from no whitening. Two things affect the strength of the statistical evidence for a difference: the sample sizes and the size of the difference.

***Part 2 Answers***

**(1)** [2pts] If the wheel is fair, what is the chance that in 32 spins a player never wins the "Jackpot"? **(C)**

**(2)** [3pts] Notice that one prize appears twice on the wheel: the dark circle with dots, which is called a "color bomb." If the wheel is fair, what is the chance that in 20 spins a player wins a color bomb 2 times? **(B)**

**(3)** [2pts] For a Standard Normal random variable what is the value of $P(-2 < Z < 2)$? **(D)**

**(4)** [1pt] If you randomly selected a female 15 years old or older in 2015, she is most likely to be in which age range? **(D)**

**(5)** [2pts] If you randomly selected a woman who is 60 years old or older in 1995, what is the chance that she is in the labor force? **(E)**

**(6)** [2pts] For 2015, if the age distribution and the overall female participation rate were the same, but female labor force participation had nothing to do with age, then what value would be in the cell for "Not in labor force" and "15 to 19 years"? **(A)**

**(7)** [2pts] Compared to women in their 30's in 1995, what can we say about women in their 30's in 2015? **(B)**

**(8)** [2pts] Consider a lottery with an 84% chance of winning nothing, a 10% chance of $10, a 5% chance of $100 and a 1% chance of $10,000. If the random variable X records the outcome of the lottery, what is the standard deviation of X? **(B)**

**(9)** [2pts] Let $X$ be a random variable where the probability model is: $P(X = -1) = P(X = 1) = \frac{1}{2}$. Let $Y$ be a random variable where the probability model is: $P(Y = -10) = P(Y = 10) = \frac{1}{2}$. Which of these statements is true? **(E)**

**(10)** [2pts] The weight of a certain species of fish at 5 years of age follows the Normal model. The population mean is 1.7 kg and the population s.d. is 0.2 kg. What is the chance of catching a fish that weighs over 1.8 kg? **(E)**

**(11)** [3pts] A bus arrives every 10 minutes to a bus stop. Consider five independent trips. Assume that for each trip the bus is equally likely to arrive any time between 0 to 10 minutes. The expected total wait time for the five trips is 25 minutes. What is the standard deviation of total wait time for the five trips? **(D)**

**(12)** [3pts] Consider an investment portfolio with 10 shares of Stock A and 20 shares of Stock B…The expected total value of the investment portfolio is $1,600. What is the standard deviation of the total value of the investment portfolio? **(C)**

**(13)** [3pts] Consider the hypothesis test $H_0: p = 0.5$ versus $H_1: p > 0.5$. If $\alpha = 0.05$, $n = 1{,}000$ and the true proportion is $p = 0.52$ then the power $(1 - \beta)$ of the test is 0.35. Which of these would make the test more powerful? **(A)**

**(14)** [2pts] For $H_0: \mu = 0$ versus $H_1: \mu \neq 0$ with a 10% significance level and 9 degrees of freedom, how big must the test statistic be (in absolute value) to reject the null and infer the research (alternative) hypothesis is true? **(D)**

**(15)** [2pts] A random sample of 115 houses have a mean selling price of $531,042 with a s.d. of $159,189…To test the overall statistical significance of the model what is the value of the test statistic? **(A)**

**(16)** [2pts] Which of these statements are true? **(C)**

**(17)** [2pts] Which of these is NOT a plausible explanation for any differences in the conclusions across these studies? **(B)**

**(18)** [2pts] Consider a multiple regression model to assess how eating breakfast (regbreak = 1 if regularly eat breakfast and 0 otherwise) and its protein intensity (high_pro = 1 if relatively high in protein and 0 otherwise) …Which of these results would directly support the claim that high protein breakfasts are better for controlling BMI? **(D)**