

## Problem Set 2

Lorenzo Bartolo

### EMPIRICAL /COMPUTER WORK

4. [50 points (all questions here are worth 5 points each except part (a) which is 10 points)] Important: As usual, your answer should include a printout (can cut and paste into a file. I will show you how to do this) of relevant calculations on the computer (R or other software output) AND a write up of final answers following the sub parts of the question. The data again are the same as for Problem Set 1. Use these data to answer the following questions.

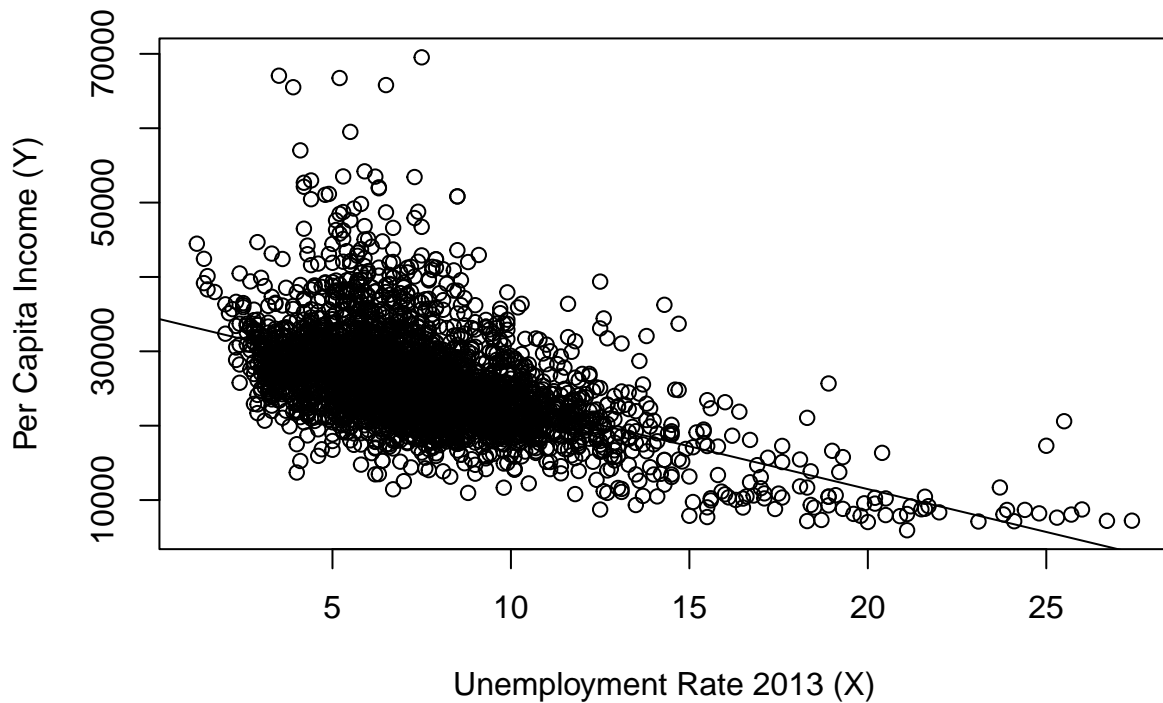
(a) Run a regression to determine the impact of the 2013 unemployment rate (*UnempRate2013*) on the per capita income (*PerCapitaInc*) in a county. What is the estimated slope? Explain what this number means in words in terms of the unemployment rate and in terms of per capita income. Also indicate if the relationship is statistically significant at the 10%, 5%, and 1% levels. For this first pass, use homoskedastic standard errors.

$$\widehat{PerCapitaInc} = 34507.06 - 1152.81 \times UnempRate2013 + u_i \quad (\text{Regression equation})$$

```
plot(PerCapitaInc ~ UnempRate2013, data = rural_atlas_merged,
     main = "Scatterplot of Per Capita Income and Unemployment Rate",
     xlab = "Unemployment Rate 2013 (X)", ylab = "Per Capita Income (Y)")

abline(a = 34507, b = -1152)
```

## Scatterplot of Per Capita Income and Unemployment Rate



```
summary.lm(model_a <- lm(formula = rural_atlas_merged$PerCapitaInc ~
  rural_atlas_merged$UnempRate2013))
```

```
##
## Call:
## lm(formula = rural_atlas_merged$PerCapitaInc ~ rural_atlas_merged$UnempRate2013)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16191  -3523   -708    2327   43668
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    34507.06     257.92   133.8  <2e-16 ***
## rural_atlas_merged$UnempRate2013 -1152.81      31.33   -36.8  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5613 on 3269 degrees of freedom
## (7 observations deleted due to missingness)
## Multiple R-squared:  0.2929, Adjusted R-squared:  0.2927
## F-statistic: 1354 on 1 and 3269 DF, p-value: < 2.2e-16
```

```
confint(model_a, level = 0.99)
```

```
##                                0.5 %    99.5 %
## (Intercept)                   33842.325 35171.801
## rural_atlas_merged$UnempRate2013 -1233.553 -1072.069
```

*# 1% level: Reject the hypothesis that the coefficient on  
# the unemployment rate is zero at the 1% level.*

```
confint(model_a, level = 0.95)
```

```
##                                2.5 %    97.5 %
## (Intercept)                   34001.368 35012.758
## rural_atlas_merged$UnempRate2013 -1214.235 -1091.387
```

*# 5% level: Reject the hypothesis that the coefficient on  
# the unemployment rate is zero at the 5% level.*

```
confint(model_a, level = 0.9)
```

```
##                                5 %      95 %
## (Intercept)                   34082.707 34931.419
## rural_atlas_merged$UnempRate2013 -1204.355 -1101.267
```

*# 10% level: Reject the hypothesis that the coefficient on  
# the unemployment rate is zero at the 10% level.*

*# Intercept = 34507.06*

*# Slope = The slope of -1152.81 means that when the  
# unemployment rate differs by 1%, on average, the per  
# capita income is lower by \$1,152.81*

*# R-squared = On average 29.29% of the variance of the per  
# capita income rate is explained by the unemployment rate*

*# SER = On average the deviation of the actual achieved per  
# capita income rate and the regression line is \$5,613*

(b) Re-run the regression from part (a) but this time use heteroskedastic standard errors. Are your coefficients the same as in part (a)? Why? Are your standard errors (of your betas) the same as in part (a)? Why?

*# Coefficients are the same as part (a)*

```
summary.lm(model_b <- lm(formula = rural_atlas_merged$PerCapitaInc ~
  rural_atlas_merged$UnempRate2013))
```

```
##
## Call:
## lm(formula = rural_atlas_merged$PerCapitaInc ~ rural_atlas_merged$UnempRate2013)
##
## Residuals:
```

```
##      Min      1Q Median      3Q      Max
## -16191 -3523   -708   2327  43668
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          34507.06    257.92   133.8   <2e-16 ***
## rural_atlas_merged$UnempRate2013 -1152.81    31.33   -36.8   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5613 on 3269 degrees of freedom
## (7 observations deleted due to missingness)
## Multiple R-squared:  0.2929, Adjusted R-squared:  0.2927
## F-statistic: 1354 on 1 and 3269 DF, p-value: < 2.2e-16
```

```
coeftest(model_b, vcov = vcovHC(model_a, type = "HC0"))
```

```
##
## t test of coefficients:
##
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          34507.063    258.767 133.352 < 2.2e-16 ***
## rural_atlas_merged$UnempRate2013 -1152.811    29.064 -39.665 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(c) Run the same regression as in part (b) but now also include the additional regressors percentage of the population that is college-educated (`Ed5CollegePlusPct`), percentage of the population that is black (`BlackNonHispanicPct2010`), and percentage of the population that is Hispanic (`HispanicPct2010`). Now, what is the estimated impact of unemployment rate in 2013 on per capita income? Also indicate if the relationship is statistically significant at the 10%, 5%, and 1% levels? Make sure that you are using heteroskedastic standard errors.

```
summary.lm(model_c <- lm(formula = rural_atlas_merged$PerCapitaInc ~
  rural_atlas_merged$UnempRate2013 + rural_atlas_merged$Ed5CollegePlusPct +
  rural_atlas_merged$BlackNonHispanicPct2010 + rural_atlas_merged$HispanicPct2010))
```

```
##
## Call:
## lm(formula = rural_atlas_merged$PerCapitaInc ~ rural_atlas_merged$UnempRate2013 +
##      rural_atlas_merged$Ed5CollegePlusPct + rural_atlas_merged$BlackNonHispanicPct2010 +
##      rural_atlas_merged$HispanicPct2010)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -20528.1 -1912.5   -44.4   1906.4  26878.9
##
## Coefficients:
##                      Estimate Std. Error t value
## (Intercept)          20905.800    272.950   76.59
## rural_atlas_merged$UnempRate2013          -500.569    24.661  -20.30
```

```
## rural_atlas_merged$Ed5CollegePlusPct          465.544      7.427    62.68
## rural_atlas_merged$BlackNonHispanicPct2010     -51.490      4.868   -10.58
## rural_atlas_merged$HispanicPct2010            -82.032      3.793   -21.63
##                                     Pr(>|t|)
## (Intercept)                                   <2e-16 ***
## rural_atlas_merged$UnempRate2013               <2e-16 ***
## rural_atlas_merged$Ed5CollegePlusPct           <2e-16 ***
## rural_atlas_merged$BlackNonHispanicPct2010     <2e-16 ***
## rural_atlas_merged$HispanicPct2010             <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3718 on 3266 degrees of freedom
## (7 observations deleted due to missingness)
## Multiple R-squared:  0.69, Adjusted R-squared:  0.6897
## F-statistic: 1818 on 4 and 3266 DF, p-value: < 2.2e-16
```

(d) Provide economic/econometric intuition as to why the impact of the unemployment rate's impact on per capita income changed between parts (b) and (c). Note that I am asking you to think about the context (and hence the “story” behind these data).

(e) Construct a 95% confidence interval for the slope coefficient on UnempRate2013 in (c). Write out your calculations. Clearly indicate how this confidence interval relates to whether UnempRate2013 is statistically significant or not in this context by relating your answer to your constructed confidence interval.

(f) You recall from problem set 1 that both the means of per capita income and of unemployment rate in 2013 are quite different across metro and nonmetro areas. You therefore want to explore this in more detail. Run the regression from (c) using only metro areas in 2013 (Metro2013==1). [Hint: You need to restrict the data based on a criterion before running the regression.] Now, what is the estimated effect of the 2013 unemployment rate on per capita income and also indicate if the relationship is statistically significant at the 10%, 5%, and 1% levels? Make sure that you are using heteroskedastic standard errors.

(g) Now, run the regression from (c) using only non-metro areas in 2013 (Metro2013==0). [Hint: You need to restrict the data based on a criterion before running the regression.] Now, what is the estimated effect of the 2013 unemployment rate on per capita income and also indicate if the relationship is statistically significant at the 10%, 5%, and 1% levels? Make sure that you are using heteroskedastic standard errors.

(h) What did you learn from the comparison between results in parts (f) and (g)? Explain your answer. Note that I again am asking you to think about the context (and hence the “story” behind these data).

(i) Return to the full sample. Now, run a regression to determine the impact of changing the percentage of the population which is college educated (Ed5CollegePlusPct) on the per capita income (PerCapitaInc) in a county. Include controls for the unemployment rate in 2010 (UnempRate2010), percentage of the population that is black (BlackNonHispanicPct2010), and now also include a dummy variable for metro status (Metro2013). Now, what is the estimated impact of percentage with a college education on per capita income? Also indicate if the relationship is statistically significant at the 10%, 5%, and 1% levels? Make sure that you are using heteroskedastic standard errors.