Michael Burrage II
Econometrics Assignment 2

1. Summary Statistics

- **Year:**
  - Observations: 1020
  - Mean: 1990.5
  - Standard Deviation: 5.76911
  - Minimum: 1981
  - Maximum: 2000
- **State Tax:**
  - Observations: 1020
  - Mean: 24.7378
  - Standard Deviation: 16.90607
  - Minimum: 2
  - Maximum: 111
- Retail Price
  - Observations: 1020
  - Mean: 148.0146
  - Standard Deviation: 148.0146
  - Minimum: 49.4
  - Maximum: 404.7
- Federal Tax
  - Observations: 1020
  - Mean: 19.3
  - Standard Deviation: 5.978716
  - Minimum: 8
  - Maximum: 34
- Cigarette pack sold per capita
  - Observations: 1020
  - Mean: 106.6021
  - Standard Deviation: 28.29377
  - Minimum: 31.9
  - Maximum: 245.4
- Consumer Price Index
  - Observations: 1020
  - Mean: .7652439
  - Standard Deviation: .1458584
  - Minimum: .5278746
  - Maximum: 1
- Per Capita Income
  - Observations: 1020
  - Mean: 18902.59
  - Standard Deviation: 5974.892
  - Minimum: 7849
  - Maximum: 41492

Michael Burrage II
Econometrics Assignment 2

2. Deflating Nominal Variables

To adjust for inflation, nominal variables such as per capita income, state cigarette tax, and federal cigarette tax were deflated using the CPI. The following real variables were created:

- **Real Per Capita Income**: Adjusted per capita income in thousands of base year dollars.
- **Real State Tax**: Adjusted state cigarette tax per pack.
- **Real Federal Tax**: Adjusted federal cigarette tax per pack.
- **Real Tax**: The total real cigarette tax (state + federal).

3. Regression Models

Four different regression models were estimated to explore the relationship between cigarette taxes and per capita cigarette consumption. Each model progressively includes more variables to assess their impact on the relationship between cigarette taxes and consumption.

- **Model 1**: Regresses cigarette consumption on real cigarette tax.
- **Model 2**: Regresses cigarette consumption on cigarette tax and real per capita income.
- **Model 3**: Regresses cigarette consumption on cigarette tax and real per capita income and adds a continuous year variable to control for time trends.
- **Model 4**: Regresses cigarette consumption on cigarette tax and real per capita income and includes year as a set of dummy variables to control fixed effects per year.

4. Interpretation of Results

a. Expected Sign of the Coefficient on Real Tax

The coefficient for real cigarette tax is expected to be negative. Higher taxes increase the price of cigarettes, which in turn reduces demand, leading to a decrease in cigarette consumption.

b. Interpretation of Coefficients for Real Tax for each Model

- **Model 1**: The coefficient is **-0.00937**, indicating that for every 1-unit increase in real cigarette tax, cigarette packs sold per capita decrease by approximately 0.00937 units.
- **Model 2**: The coefficient is **-0.00900**, slightly smaller than in Model 1. This suggests that per capita income influences cigarette consumption, reducing the direct effect of taxes.
- **Model 3**: The coefficient is **-0.00776**, less than both Model 1 and Model 2. This implies that some of the changes in consumption are due to trends over time, not just taxes.
- **Model 4**: The coefficient is **-0.00821,** including a year as dummy variables results in controlling for fixed effects, reducing the confounding impact of unobserved changes across years.

The coefficient is different across models due to the following: omitted variable bias and collinearity. Model 1 may suffer from omitted variable bias since it does not include variables that control for income or time trends, as other factors are not accounted for in the model. Also, the coefficients are different because of multicollinearity; as more variables are added, some

variables of variables are sharing information. For example, income and time trends may both impact cigarette consumption and adding both to the model can result in the reduced effect of real tax.

c. R-squared Changes Across Models

As expected, the R-squared increases as more variables are added because each additional variable explains more of the variation in cigarette consumption. R squared always increases the more variables that are added to the regression model.

- **Model 1**: R-squared = 0.3739
- **Model 2**: R-squared = 0.3755
- **Model 3**: R-squared = 0.4546
- **Model 4**: R-squared = 0.4784

d. Statistical Significance of Real Tax

Yes, in all models, the coefficient for real_tax is highly statistically significant ($p < 0.001$) from zero reflected in the p values and the confidence interval not crossing zero. The significance of the tax variable remains consistent across models despite changes in the size of the coefficient.

- **Model 1: p-value** < .0001 and Confidence interval: [−0.0101194,−0.0086275]
- **Model 2: p-value** < .0001 and Confidence interval: [−0.0098797,−0.0081381]
- **Model 3: p-value** < .0001 and Confidence interval: [−0.008599,−0.0069214]
- **Model 4: p-value** < .0001 and Confidence interval: [−0.009061,−0.0073634]

5. Frisch-Waugh-Lovell (FWL) Theorem

The coefficients are identical whether obtained using the FWL theorem or from the original model. This confirms that whether you use the residuals or the original dependent variable, the coefficient estimate for real tax remains the same, which is exactly as expected and consistent with the Frisch-Waugh-Lovell theorem.

6. FWL Theorem for Model 4

After repeating the same procedure was repeated for Model 4, using the residuals from the regression of real_tax on real_pci and year dummy variables. The results again confirmed the robustness of the relationship between real_tax and cigarette consumption. This confirms that the FWL theorem holds true for this model as well, showing that the coefficient for real tax remains consistent whether we include the control variables directly or through residuals.

7. High-Dimensional Fixed Effects Model

The results from **reghdfe** and the previous model are slightly different but are highly consistent. The coefficient for real tax (reghdfe) is −0.0082122, and the coefficient for real tax (previous model) is −0.0082198. The difference between these two coefficients is insignificant, indicating

that the effect of real tax on cigarette consumption is robust since both models show a statistically significant negative impact of real tax on cigarette packs sold. Also the R square for each model, r-squared (reghdfe): 0.4784 and r-squared (previous model): 0.4784 are identical. Both models explain approximately 47.8% of the variation in cigarette consumption, showing that real tax, real PCI, and year effects together account for a substantial portion of the observed variation in cigarette packs sold per capita.

Thus, the use of reghdfe to absorb year fixed effects yields nearly identical results to the model with explicit year dummies, confirming that the negative and significant effect of real tax on cigarette consumption remains consistent across both approaches.

8. Relationship Between Real Tax and Real PCI

Based on this information Model 1, does not control for real PCI, so it will overstate the impact of real tax on cigarette consumption. This is because, in Model 1, the effect of real tax includes both its direct impact on cigarette consumption and its indirect effect through income (real PCI). Specifically, the regression of real PCI on real tax shows a positive coefficient of 1220.116, meaning that higher taxes are associated with higher income levels. Additionally, the relationship between real PCI and pcpacks is negative, with a coefficient of -0.0011 from Model 2, indicating that higher income leads to lower cigarette consumption. Therefore, omitting real PCI in Model 1 makes the negative impact of real tax on cigarette consumption appear stronger than it truly is, because the model fails to account for the indirect effect of higher taxes increasing income.

9. Comparison of Coefficients Between Model 1 and Model 3

Model 1, which does not control for real PCI and year, shows a more negative coefficient for real tax (−0.0094) compared to Model 3 (−0.0078). This suggests that Model 1 overstates the negative effect of real tax on cigarette consumption because it fails to account for other important variables like income (real PCI) and time trends (year). Specifically, in Model 3, the coefficient for real PCI is 4.56×10^{-7}, indicating that higher income levels slightly reduce cigarette consumption, and the coefficient for the year is −1.6196, meaning that consumption is trending downward over time.

Model 3 provides a more accurate estimate of the effect of real tax by controlling for income and time, leading to a smaller (less negative) coefficient. This is consistent with the idea that omitting relevant variables in Model 1 leads to an overstatement of the impact of real tax. Additionally, the R-squared in Model 1 is 0.3739, while in Model 3) it increases to 0.4546, indicating that the inclusion of real PCI and year improves the model's explanatory power.