Michael Burrage II
Econometrics Assignment 4

1. Using the dataset (dataset assignment 4.dta) provided, answer the following questions:
    a. Suppose you are estimating a model of the dependent variable y and you are interested in β_1, β_2, and β_3. Estimate the model using OLS and with all the x variables as controls. Report the coefficients of β_1, β_2, and β_3 along with their standard errors.

$\beta 1$ (x1): Coefficient = -27,038.11, Standard Error = 28,871.93, p = 0.349
$\beta 2$ (x2): Coefficient = 1,190.91, Standard Error = 64.03, $p < 0.001$
$\beta 3$ (x3): Coefficient = 1,067.89, Standard Error = 63.10, $p < 0.001$

    b. Use robust standard errors instead of the usual standard errors. Does it change the coefficient estimates of β_1, β_2, and β_3? How about the standard errors? Is the robust standard error larger or smaller than the usual standard errors?

$\beta 1$ (x1): Coefficient = -27,038.11, Robust Standard Error = 28,763.17
$\beta 2$ (x2): Coefficient = 1,190.91, Robust Standard Error = 136.06
$\beta 3$ (x3): Coefficient = 1,067.89, Robust Standard Error = 132.98

The coefficient estimates remain the same for $\beta 1$, $\beta 2$, and $\beta 3$. The standard errors for $\beta 2$ and $\beta 3$ are larger with robust standard errors, indicating potential heteroskedasticity. For $\beta 1$, the robust standard error is slightly smaller than the robust standard error.

    c. The dataset provide you with the id and t, which are the individual id and time period. Rerun the estimation in part (a) with individual fixed effects and time fixed effects. What is the coefficient of x1? Report the coefficient and the robust standard errors. (Hint: If you are generating dummy variables for each individual, your computer may not be able to handle the large number of variables. Instead, use reghdfe to handle fixed effects estimation)

Coefficient of x1: 10.5
Robust Standard Error: 8.45

    d. Suppose now we have extra information about the value of true $\beta 1$ and $\beta 2$, which are -3 and 2. You probably have noticed that the coefficient estimates of $\beta 1$ and $\beta 2$ are not close to the true values. The reason is that some of the control variables are problematic. Loop through different combinations of control variables and estimate the model. Report the coefficient estimates of $\beta 1$ and $\beta 2$ that are closest to the true values and which set of control variables you used to get the closest estimates.

Closest Estimates: $\beta 1$: -3.01; $\beta 2$: 2.05
Control Variable Used: x3, x4

    e. Rather than using the time fixed effects, we can also use a linear time trend. Rerun the model with a linear time trend. Rerun the model with the set of control variables that you found in the previous part and report the coefficient estimates

of β1 and β2. Are these coefficient estimates more precise than the previous estimates?

Coefficients linear time trend and the control variables (x3, x4):
β1: −3.02
β2: 2.01
No time trend:
β1: SE=0.12
β2: SE=0.09
Linear time trend:
β1:SE=0.10
β2:SE=0.08

The coefficient estimates with the linear time trend are slightly more precise, as indicated by smaller standard errors compared to the previous estimates.

f. How about a nonlinear time trend? Repeat the previous part with a quadratic time trend.

Coefficients quadratic time trend and the same control variables (x3, x4):
β1: −3.01
β2: 2.00
Linear time trend:
β1: SE=0.10
β2: SE=0.08
Quadratic time trend:
β1: SE=0.11
β2: SE=0.09

The quadratic time trend does not improve precision compared to the linear time trend because the standard errors for the coefficients are slightly larger.