Problem Set 2 MGMT 737

For both analyses this week, you will be using data from Mian and Sufi's 2014 Econometrica article, *What Explains the 2007-2009 Drop in Employment?*. The analyses will not match the exact numbers in the paper, but the full replication set is available if you are interested in exploring it.

1. **Standard Errors** For this problem, use the dataset `networth_delta_elas.csv`, where `county_fips` is the county FIPS code, `statename` is the state FIPS code, `elasticity` is the Saiz elasticity measure, `total` is the number of households in each county, and `netwp_h` is the change in net worth within a county from 2006 to 2009.

   (a) Write a function to esitmate the linear regression of networth change against a constant and the Saiz elasticiy. Report the coefficient on the elasticity.

   (b) Next, estimate the homoskedastic SE, heteroskedasticity-robust SE, HC2, and HC3 standard errors for the elasticity estimate.

   (c) Now, we will estimate the three standard errors from Abadie et al. (2020) [see section 4]. I will walk you through the estimation. Let $V^{\text{causal}} = n^{-1}\Gamma^{-1}(\rho\Delta^{\text{cond}}+(1-\rho)\Delta^{\text{ehw}})\Gamma^{-1}$, $V^{\text{causal,sample}} = \Gamma^{-1}\Delta^{\text{cond}}\Gamma^{-1}$, $V^{descr} = n^{-1}(1-\rho)\Gamma^{-1}\Delta^{\text{ehw}}\Gamma^{-1}$, and $V^{\text{ehw}} = n^{-1}\Gamma^{-1}\Delta^{\text{ehw}}\Gamma^{-1}$. Our elasticity measure is X_i, and the outcome is Y_i. Let Z denote our constant (and potentially an additional control).

   - Estimate $\hat{\epsilon}_i$ as the standard residual from the linear regression of Y on X and Z
   - Estimate the short regression of X on Z (X as the outcome, Z as the right hand side) to calculate $\hat{\gamma}$, the projection of X on Z (note that when Z is a constant, this is just the mean of X).
   - Estimate $\hat{\Gamma} = n^{-1}\sum_i(X_i - \hat{\gamma}Z_i)^2$
   - Estimate $\hat{\Delta}^{\text{ehw}} = n^{-1}\sum_i(X_i - \hat{\gamma}Z_i)\hat{\epsilon}_i^2(X_i - \hat{\gamma}Z_i)$.
   - Now, estimate $V^{EHW} = (1/n) * \hat{\Gamma}^{-1}\hat{\Delta}^{\text{ehw}}\hat{\Gamma}^{-1}$. Check that this coincides with your previous EHW estimates (it should differ slightly b/c of no degree of freedom corrections, but quite close).
   - Estimate $\rho = n/N$ using the following fact: the data is observed at the county level, and in the United States, there are 3,006 counties. Recall that (in my notation) n is the number of observations in the sample, and N is the "population." Using this measure, estimate $V^{descr} = (1-\rho)V^{EHW}$ and report the standard error. Note the relative size of each.
   - Next, calculate:

     $$\hat{G} = \left(n^{-1}\sum_i(X_i - \hat{\gamma}Z_i)\hat{\epsilon}_i Z_i'\right)\left(n^{-1}\sum_i Z_i Z_i'\right)^{-1} \tag{1}$$

     Note that this is exactly zero in our current case.
   - Now, let $\hat{\Delta}^Z = n^{-1}\sum_i((X_i - \hat{\gamma}Z_i)\hat{\epsilon}_i - \hat{G}Z_i)^2$. Note that this should be equal to $\hat{\Delta}^{EHW}$.
   - Finally, calculate V^{causal} and $V^{\text{causal,sample}}$ using $\hat{\Delta}^Z$ in the place of Δ^{cond} and and report the standard errors. (We cannot estimate Δ^{cond} feasibly, so we use Δ^Z in its place.) Note that in this setting, we have identical estimates for the causal estimates b/c we cannot do better than the EHW estimate.
   - Now reimplement this approach, but include state fixed effects as controls in Z. Report your estimates for the standard errors using V^{EHW}, V^{descr}, V^{causal} and $V^{causal,sample}$ in this setting.

2. **Binscatter**: For this problem, use the dataset `networth_delta_elas.csv`, where `county_fips` is the county FIPS code, `statename` is the state FIPS code, `year` is the year, `elasticity` is the Saiz elasticity measure, total is the number of households in each county, and `hpi` is Zillow House Price value. For this problem, you may use your own regression estimate function, or a pre-existing function to estimate the regression.

(a) Calculate annual house price appreciation (hpa) within each county, and regress HPA on the elasticity measure interacted with year, using your constructed function. I.e. $\texttt{hpa}_{it} = \alpha_t + \sum_s \texttt{elasticity}_i \times 1(year_t = s)\beta_t$ Plot the β_t coefficient for each year across time. Report the coefficient measuring the effect of elasticity in 2008.

(b) Construct 10 decile dummies for the elasticity and reestimate the regression, pooling the years 2008-2010, and using the ten dummies in the place of the continuous elasticity measure. Plot these decile effects such that each point reflects an approximation to the conditional expectation function. Report the value for the first decile.