Problem Set 5

MGMT 737

1. **Poisson Regression**. This analysis will use the dataset `Detroit.csv`. You should use built-in pacakges for this – in `R`, use the `fixest` library. In `Stata`, use the `reghdfe` and `ppmlhdfe` packages.

   (a) Run an OLS regression of `flows` (the number of workers who work in `home_ID` and work in `work_ID`) on `distance_Google_miles`, and include `home_ID` and `work_ID` fixed effects (absorb them), and cluster on `home_ID`. Report the coefficient and standard error on `distance_Google_miles`.

   (b) Run an OLS regression of `log(flows)` on `log(distance_Google_miles)` and include `home_ID` and `work_ID` fixed effects , omitting the cells with zero flows. Cluster on `home_ID`. Report the coefficient and standard error on `log(distance_Google_miles)`.

   (c) Repeat part 1b, but instead of omitting the zero cells, run the OLS regression of `c+ log(flows)` for c = 0.1, 1 and 10. Compare how your coefficients change.

   (d) Finally, repeat part 1a using Poisson regression, and contrast the estimates to Part b and c.

2. **Duration Modeling**. This analysis will use the dataset `acs_duration.csv`. The `acs_duration.csv` dataset is from the American Community Survey in 2019, and has heads of households' responses to the question "How long have you lived in this home?" (`moving_approx` – in reality, this value is given as a range in the public data – I have imputed using the midpoint. A fun exercise left to the reader is to think about how to generalize this problem using ranges.) and homeownership (`homeowner` vs. `renter`).

   (a) Using the ACS data,write down how to estimate the unconditional probability that a household stays in a home for T or more years, using the available data. Estimate this for $T = 7$ and report the value.

   (b) Calculate the hazard rate for each observed value of `moving_approx`. Report this value for $T = 7$.

   (c) Recalculate these hazard values for $T = 7$ for homeowners and renters. Contrast the difference in hazard rates over time.