**ECON 434 Project**

**Uber versus public transit**

In this project, we study whether Uber complements (helps) or substitutes (hurts) public transit. On the one hand, Uber can substitute public transit if riders decide to choose Uber instead of public transit. On the other hand, Uber can complement public transit if riders take Uber from home to public transit stop, which can make public transit more attractive than driving a car. The net effect is unclear and is a subject of intense policy debate.

We will expand on the original set of results presented in Hall, Palsson, and Price (2018), "Is Uber a substitute or a complement for public transit," *Journal of Urban Economics*, which is available on the class website. We will use their dataset, which is also available on the class website. In the dataset, a unit of observation is a public transit agency in a given year-month. The dataset includes information on both the transit agencies and on the Metropolitan Statistical Areas (MSA) where they operate. For each time period, the dataset contains values for the following variables:

1. UPTTotal – the number of rides for the public transit agency;
2. treatUberX - a dummy for Uber presence in the corresponding MSA;
3. treatGTNotStd - a variable measuring google search intensity for Uber in the corresponding MSA;
4. popestimate - population in the corresponding MSA;
5. employment - employment in the corresponding MSA;
6. aveFareTotal - average fare for the public transit agency;
7. VRHTTotal - vehicle hours for the public transit agency;
8. VOMSTotal - number of vehicles employed by the public transit agency;
9. VRMTotal - vehicle miles for the public transit agency;
10. gasPrice - gas price in the corresponding MSA.

In this dataset, treatUber and treatGTNotStd is qualitative and quantitative measures for the same thing: Uber presense in the MSA. We can run regressions using either of these two variables and then check whether results are robust if the other variable is used.

There are two variations in this dataset that allow us to study the effect of Uber on public transit. First, in any given time period, Uber is present in some MSAs but not in others. We can thus study the effect of Uber by comparing these MSAs. Second, for any given MSA, we have data on time periods both before and after Uber was introduced in this MSA. We can thus study the effect

of Uber by comparing these time periods. By working with panel data, we are able to employ both variations at the same time.

To study the effect of Uber on public transit, we let $Y_{it}$ be UPTTotal, $D_{it}$ be either treatUberX or treatGTNotStd, and $W_{it}$ be the vector including remaining variables: popestimate, employment, aveFareTotal, VRHTTotal, VOMSTotal, VRMTotal, gasPrice. We then run the following regressions:

1. OLS: $\log Y_{it} = \alpha + D_{it}\beta + W_{it}'\gamma + e_{it}$.
2. OLS: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta + W_{it}'\gamma + e_{it}$.
3. OLS: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta_1 + D_{it}P_{it}\beta_2 + W_{it}'\gamma + e_{it}$, where $P_{it}$ is a dummy that takes value 1 if the corresponding MSA has population larger than the median population in the dataset and 0 otherwise.
4. OLS: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta_1 + D_{it}F_{it}\beta_2 + W_{it}'\gamma + e_{it}$, where $F_{it}$ is a dummy that takes value 1 if the number of rides of the public travel agency is larger than the median number of rides among all public transit agencies in the dataset.
5. LASSO: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta_1 + D_{it}P_{it}\beta_2 + W_{it}'\gamma + e_{it}$.
6. LASSO: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta_1 + D_{it}F_{it}\beta_2 + W_{it}'\gamma + e_{it}$.
7. Double-LASSO: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta_1 + D_{it}P_{it}\beta_2 + W_{it}'\gamma + e_{it}$, where coefficients of interest are $\beta_1$ and $\beta_2$.
8. Double-LASSO: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta_1 + D_{it}F_{it}\beta_2 + W_{it}'\gamma + e_{it}$, where coefficients of interest are $\beta_1$ and $\beta_2$.
9. LASSO: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta_1 + D_{it}P_{it}\beta_2 + \widetilde{W}_{it}'\gamma + e_{it}$, where $\widetilde{W}_{it}$ includes all interactions of order 5 of variables in the vector $W_{it}$.
10. LASSO: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta_1 + D_{it}F_{it}\beta_2 + \widetilde{W}_{it}'\gamma + e_{it}$, where $\widetilde{W}_{it}$ includes all interactions of order 5 of variables in the vector $W_{it}$.
11. Double-LASSO: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta_1 + D_{it}P_{it}\beta_2 + \widetilde{W}_{it}'\gamma + e_{it}$, where coefficients of interest are $\beta_1$ and $\beta_2$ and $\widetilde{W}_{it}$ includes all interactions of order 5 of variables in the vector $W_{it}$.
12. Double-LASSO: $\log Y_{it} = \eta_i + \delta_t + D_{it}\beta_1 + D_{it}F_{it}\beta_2 + \widetilde{W}_{it}'\gamma + e_{it}$, where coefficients of interest are $\beta_1$ and $\beta_2$ and $\widetilde{W}_{it}$ includes all interactions of order 5 of variables in the vector $W_{it}$.

Regressions 2, 3, and 4 roughly correspond to regressions run in the original paper (see Table 3 there). Please refer to the original paper for motivation on including $P_{it}$ and $F_{it}$ in the analysis. Regressions 5–12 are new. Of particular interest to us are regressions 7, 8, 11, and 12, corresponding to Double-LASSO. The purpose of this project is to establish whether these regressions give results that are qualitatively similar to those established in the original paper.

Bonus: try to run a few extra regressions that would help us shed some light on the relationship between Uber and public transit.

You can work on this project in groups of up to 4 students. When you submit your results, please write down the members of your group on the top of your submission. All members of the group should submit identical files. Make sure that each member of the group participate equally in the project.