Impact of TNCs on Road Safety – San Francisco Case Study

Abstract

Transport Network Companies (TNCs) are increasingly establishing themselves as one of the dominant mode choices in the urban areas. Impact of TNCs on congestion is widely documented, however their contribution towards minimizing road safety is largely contradicting. Using San Francisco city road network information available along with road crash data for the year 2010 in which TNCs were not present to 2016 data when TNCs services were fully established, the study attempts to quantify their impact on road safety. The classwork finds that the pick-up and drop-off services (PUDO) indeed have a positive relationship with occurrence of road crash, while TNCs volume are negatively related.

Research Question and Hypothesis

Past research focussing on understanding the impact of TNCs on road safety and congestion have produced contrasting conclusions. The current proposal aims to use San Francisco road network and traffic crash details to identify the relationship between TNCs and its impact on Road Safety. Primarily, the proposed research aims to understand whether operations of TNCs are contributing towards the Cities own Vision Zero.

Methods and Data

Data

I utilized already calibrated data made available as part of the supplementary material of Erhardt et al (2019) for the San Francisco city. The supplementary data consists of

a) Calibrated San Francisco's travel demand model (SF-CHAMP) for base year 2010 and 2016 which provides estimates of traffic volumes on all roads in San Francisco and is sensitive to changes in population and demographics, employment, transportation networks, and congestion

b) Associated data reflecting TNC volumes, pickups, and drop-offs to each road segment in city by time of day (TOD).

Further, as a part of this exercise, I extracted SF crash data from the TransBASE website which is the central repository maintaining traffic (current and historical) crashes as defined under San Francisco and California public records laws. Under no circumstances the sensitive data related to the accused or victim is been seeked or sought for my analysis.

Method

The study follows before-and-after assessment (comparison) structure for the year 2010 and 2016 conditions. By doing so, the study estimates the relationship between the TNC operations activity and its impact on the overall road safety scenario.

Part I: Processing the road network data

QGIS (open source GIS) was used to convert the CSV delimited file containing SF road network into a proper projection and then later on as GeoJSON files.

Part II: Processing road crash data

The second setup for my analysis was to analyze the basic trends in road crashes in San Francisco City. I operationalized the traffic crashes in two ways

1. Total Collisions is sets of records where a crash is reported.
2. Each crash record was introduced a dummy attribute for the year of collision. 0 for 2010 and 1 for 2016

Part III: Data manipulation

1. Crash records for both year 2010 and 2016 were spatially (respectively) joined and attached to the nearest road segment. The resultant join layer was a new vector (point) layer with the same geometry type and coordinate reference system as the input (road crash) layer.
2. The road network was then joined with resultant layer as derived in step 1. The obtained layer had attributes of both the segment and crash data.
3. Aggregation of total number of crashes, total fatality count and total injury counts was performed on the layer obtained in step 2.
4. The available sets of data were then further aggregated to suit experiments requirement using Jupyter notebook. The final code is made available as part of the research to support verification and authenticity.

Findings

A panel data model was setup to compare the impact of TNC services on the road safety between 2010 and 2016. Non-TNC vehicle volume, TNC Volume along with TNCs pick-up and drop-off (PUDOs) were used as independent variable to contemplate their impact on total crashes. The dependent variable was transformed by taking log of the variable to handle the skewness of traffic crashes to the right and which are at best random in nature. This gives the percent increase (or decrease) in the response for every one-unit increase in the independent variable.

Below is the snippet of the PanelOLS model arrangement along with the output

```
# basic log model for bus, all clusters
mod=PanelOLS.from_formula('np.log(1+total_crashes) \
                    ~ NON_TNC_VOL \
                    + TNC_VOL \
                    + PUDO \
                    + YEAR_DUMMY \
                    + EntityEffects',
                    data=df3)
res=mod.fit()
print(res)
```

Figure 1: PanelOLS model

```
                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:     np.log(1 + total_crashes)   R-squared:                 0.0023
Estimator:                         PanelOLS   R-squared (Between):       0.1399
No. Observations:                     84516   R-squared (Within):        0.0023
Date:                  Wed, Apr 29 2020       R-squared (Overall):       0.0945
Time:                          17:09:39       Log-likelihood             230.77
Cov. Estimator:                Unadjusted
                                              F-statistic:               24.552
Entities:                           42258     P-value                    0.0000
Avg Obs:                           2.0000     Distribution:           F(4,42254)
Min Obs:                           2.0000
Max Obs:                           2.0000     F-statistic (robust):      24.552
                                              P-value                    0.0000
Time periods:                             2   Distribution:           F(4,42254)
Avg Obs:                       4.226e+04
Min Obs:                       4.226e+04
Max Obs:                       4.226e+04

                             Parameter Estimates
==============================================================================
              Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
NON_TNC_VOL   1.666e-05  3.276e-06     5.0862     0.0000   1.024e-05   2.309e-05
TNC_VOL      -7.106e-05  1.292e-05    -5.5019     0.0000  -9.638e-05  -4.575e-05
PUDO             0.0016     0.0003     5.8200     0.0000      0.0010      0.0021
YEAR_DUMMY       0.0058     0.0027     2.1139     0.0345      0.0004      0.0112
==============================================================================

F-test for Poolability: 1.5691
P-value: 0.0000
Distribution: F(42257,42254)

Included effects: Entity
```

Figure 2: Results of PanelOLS model

First thing and foremost is the fact that R-squared value is exceptionally small value and it is likely due to the presence of large number of 0 road crash reported road network segments. Under the existing limitation, the non-tnc volume is significant meaning that they are positively related to road crashes. Same goes with PUDO and Congested speed. However, coefficient of TNC volume is negative and opposite of what I expected. I was assuming that both PUDO and TNC-volume would go in the same direction, given that they are directly proportional in nature i.e. a greater number of TNC service means a larger number of pick-up and drop-off services. The T-stat for these variables are also more than 1.96 suggesting that the impacts of these variables is significant.

The analysis suggests that there exist important limitations in the way the research exercise was conducted and the variables should be further refined so to have better causal relationship supporting the research question idea.

Acknowledgement

The author acknowledges Prof. Greg Erhardt's for his significant help and guidance in this project exercise.

References

Buckley, Michael. n.d. "Safety Perceptions of Transportation Network Companies (TNCs) by the Blind and Visually Impaired," 84.

Erhardt, Gregory, Sneha Roy, Drew Cooper, Bhargava Sana, Mei Chen, and Joe Castiglione. 2019. "Do Transportation Network Companies Decrease or Increase Congestion?" Science Advances 5 (May): eaau2670. https://doi.org/10.1126/sciadv.aau2670.

February 21, Gerald H. Baker |, and 2018 at 05:00 PM. n.d. "New Act Regulates Transportation Network Companies." New Jersey Law Journal. Accessed March 23, 2020. https://www.law.com/njlawjournal/2018/02/21/new-act-regulates-transportation-network-companies/.

Miller, Derek. n.d. "The Sharing Economy and How It Is Changing Industries." The Balance Small Business. Accessed March 23, 2020. https://www.thebalancesmb.com/the-sharing-economy-and-how-it-changes-industries-4172234.

"Shared Economies_1pgr_07.06.17_v6.Pdf." n.d. Accessed March 23, 2020. https://www.naco.org/sites/default/files/documents/Shared%20Economies_1pgr_07.06.17_v6.pdf.

"Sharing Economy Business Models: Focus on Ride-Sharing." n.d. California Management Review. Accessed March 23, 2020. https://cmr.berkeley.edu/2018/09/ride-sharing/.

"Transportation Network Company (TNC) Legislation." 2016. Transportation Policy Research. June 6, 2016. https://policy.tti.tamu.edu/technology/tnc-legislation/.