Measuring the Impact of COVID-19 on Public Transit Demand in US with Mobile App Data

1. Introduction

A novel coronavirus disease (COVID-19) has evolved into a global health crisis for its high contagiousness, high fatality rate, and lack of known treatments. To deal with the pandemic, Centers for Disease Control and Prevention (CDCs) and governments in different places recommended social distancing, self-quarantine, and work from home policies to stop the spread of the virus. This sudden shutdown caused recessions in the domain of mobility, economic, and social activities. Different modes of transportation, especially mass transit, suffered considerable loss of passengers due to self-quarantine recommendation and contagion concerns. For example, as countries are closing their borders amidst the fear of transnational spread, the final week of March 2020 witnessed commercial flights dropped by 55% compared to the same date in 2019 (Weber 2020).

This sudden meltdown especially impacted the public transit system across the United States compared with other modes. First, just like other transportation modes, fewer passengers need to take the transit systems due to the suspension of all unessential businesses and work from home orders. Moreover, compared with other transportation modes, public transit systems are perceived to have higher risk. It is natural and common to assume that sharing a narrow and unventilated space with numerous strangers is a bad idea during the pandemic, thus people will generally avoid taking transit trips. According to an online survey (sample of n = 1000 Americans and n = 1000 Canadians), about 48% of Americans and 40% of Canadians feel that riding transit poses a high health risk due to the virus (Yellin 2020). Therefore, most transit systems has witnessed major decrease of transit demand even before the outbreak of the pandemic. For larger transit systems, for example, Washington Metropolitan Area Transit Authority reported that metrorail ridership has been reduced by 90% and bus ridership has been reduced by up to 75% by the end of March (WMATA 2020); for smaller transit system such as El Dorado transit in El Dorado county, California, the ridership also decreased by up to 75% (Christensen 2020). This unprecedented widespread recession has created major economic and stuff health difficulties for all public transit systems across the nation. It is necessary and urgent to measure the impact of the COVID-19 pandemic on the transit systems. This includes the impacts’ extent, geographic and temporal variation, and its relevance with different social-economic, demographic factors, and the development of the pandemic.

However, most transit systems will not or have yet to release their ridership data, and it will take a long time to collect these data for each system. In this paper, we use the transit demand decrease data obtained from the Transit app to infer the change of ridership. We use logistic model to fit the data of each transit system and get the key parameters from the models: **floor value**, which represents the limit of social distancing**; decay rate and half-life**, which represents the speed of the recession; **divergent date**, which represents the initial date when the transit demand began to decrease. We **also calculate the distance between the transit demand decrease curve and the disease cases number curve.**

1. Results

For all systems during the first few days, the demand change would generally oscillate around the base line of 0%. As the epidemic and quarantines progressed, the demand decreased until reached a low level and then stabilized. The shape of the curves can be well fit using a logistic model:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Where: is the curves’ minimum decreased value or *floor value*, which represents the ratio of transit users that continued to use transit systems during the pandemic. k is the logistic growth rate or steepness of the curve or *decay rate*, which represents the speed of demand decline. t represent the time (day) and is the day when the function reaches the midpoint; is the curves’ baseline value.

We fit each transit system’s daily demand data using logistic model individually and calculate these three parameters for all 113 transit systems in the US.

**Floor value and correlated factors**. Floor value represents the ratio of public transit users in this system that still would not or cannot stop using it regardless of the pandemic. For different systems in different metro areas, the floor value is vastly different as shown in Figure 1. Information Technology (IT) -dominating cities, such as San Francisco Bay area, and university cities, such as Ithaca and Ann Arbor generally have low floor value during the pandemic. Meanwhile, cities in the South and Midwest generally have higher floor value. This suggests that the floor value can have natural linkages with different social-economic factors of the local area.

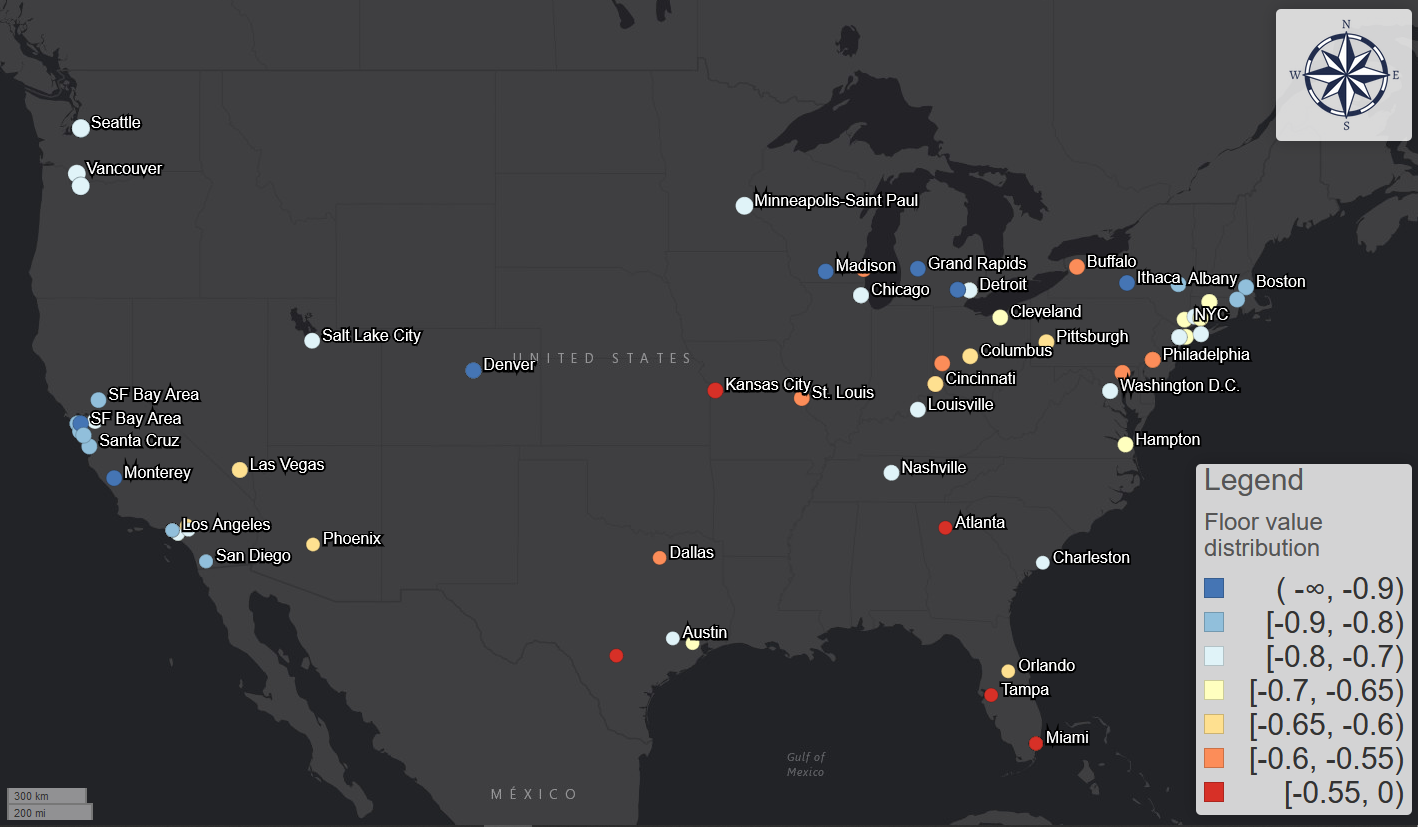


Figure 1: The distribution of floor value across the United States.

To corroborate this geographic pattern, we conducted linear regression analyses between the floor value and different social-economic and demographic factors. The regression results in Table 1 show that three factors are significant with p-value smaller than 0.05; the F test also shows the model is significant with p-value of 1.31E-10. The R-squared value is 0.36, which indicates relatively large effect size. The residuals assessment moreover shows that the residuals are subject to normal distribution and there are no multicollinearity effect and no leverage points.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Estimate coefficient | Standard Error | t value | Pr(>|t|) | VIF | R-squared increment |
| Intercept | -0.86256 | 0.13004 | -6.633 | 1.31E-09 | - | - |
| Ratio of people with non-physical occupation | -0.52727 | 0.12008 | -4.391 | 2.62E-05 | 1.03 | 0.13 |
| Ratio of African American population | 0.47421 | 0.07549 | 6.282 | 7.02E-09 | 1.05 | 0.25 |
| Ratio of population over 45 years old | 0.87797 | 0.26589 | 3.302 | 0.0013 | 1.05 | 0.077 |

Table 1: regression results

* **Ratio of population with non-physical occupations**.

Similar to *life fixity* introduced by Kim et al. (2017), the ratio of population with non-physical occupations measures the population’s degree of freedom to change the routine of their daily life. It represents how many people can work remotely, thus avoid regular transit commuting to reduce contagion risk. If an area has higher ratio of non-physical jobs, workers can easily work from home thus the transit demand will decrease further. The results validate that more transit demand decreases and higher percentage of people with non-physical occupations are correlated.

Although we do not directly add the ratio of Hispanic population into the regression model to avoid multicollinearity, the further correlation analyses between the Hispanic population and the non-physical occupation ratio indicate very strong positive correlation (p-value = 1.07E-08, cor=-0.5, 95% CIs = -0.35 and -0.63). This correlation moreover suggests the vulnerable of Hispanic population during this health crisis: if a city has more Hispanic population, it is very likely for the city to have a high floor value, which means more people will not work from home during the pandemic. This is also consistent with the occupation statistics: according to the labor force characteristics survey made by US Bureau of Labor Statistics, Hispanic population has the lowest percent (22%) of management, professional, and related occupation compared with White (41%), African American (31%), and Asian people (54%) in 2018 (U.S. Bureau of Labor Statistics 2018).

The ratio of people with non-physical occupations is also highly correlated with the median of income. In our correlation analyses, if the transit is located in an area with higher income, the floor value of the transit demand is also likely to be lower. The income has a natural linkage with the occupation composition.

These correlation results are also proven by the user surveys conducted by Transit app in April 2020. According to the survey results, the top-5 occupation categories that are most likely to work from home are computer and mathematical; life, physical and social science; education, training and library; architecture and engineering; and legal. Although the categorizations of the survey and the ACS table are different, this is generally consistent with the non-physical occupation categories we assumed in the ACS table. Meanwhile, the survey also proves that Spanish speakers are more likely to continue using the Transit app for trip planning purposes: English-language users dropped 71% from early February while Spanish-language users dropped by 50% over the same time period (Transit app 2020d). The income correlation is also confirmed by the survey results: compared with the survey results conducted by American Public Transportation Association (APTA) in 2017, active users skew towards lower income brackets during the pandemic, especially for those whose annual income is less than $15000 (Transit app 2020d). The survey results provide a first-hand proof and reaffirm the correlation results about the vulnerability of Hispanic population and low-income population.

The lockdown resembles a sieve: only the most essential, the most resilient, and the most desperate would stay.