

Impact of the COVID-19 Vaccination on Bike Sharing Demand in NYC: ARIMAX and Bayesian Time Series Analyses

Transportation Research Record
1–16© National Academy of Sciences:
Transportation Research Board 2023
Article reuse guidelines:sagepub.com/journals-permissions
DOI: 10.1177/03611981231165514

journals.sagepub.com/home/trr

Justin Drummond¹  and Md Sami Hasnine¹ 

Abstract

The objective of this study is to investigate the impact of the COVID-19 vaccine on bike sharing demand in New York City (NYC). The vaccination's impact was also compared with other known influences (COVID-19 cases counts, COVID-19 deaths, weather data, trip purpose, and more) to help improve bike share demand modeling in a pandemic setting. Autoregressive integrated moving average (ARIMAX) time series models were estimated for Brooklyn and Manhattan in both the pre-vaccine and post-vaccine periods. Mean absolute percentage error (MAPE) was used for model evaluation, and the average MAPE was below 5, suggesting a high-level accuracy in modeling. The Bayesian time series analysis results were very similar to the ARIMAX model results (with the exception of some parameters being significantly underestimated). The results of the time series analyses showed that vaccination did not have a significant effect on bike sharing demand in Manhattan but had a significant effect leading to increases in bike sharing demand in Brooklyn. Despite this, vaccination was not the main influence on bike sharing demand. For instance, in Brooklyn, grocery and retail/recreational shopping were strong influences both pre- and post-vaccine, indicating that regardless of vaccination, shopping is essential. COVID-19 cases counts had opposing effects in Brooklyn and Manhattan. Other findings include higher temperatures leading to increased bike demand, and precipitation and stronger winds leading to decreased bike sharing demand.

Keywords

planning and analysis, attitudes/attitudinal data, behavior analysis, behavioral process, behaviors, mode choices, pattern (behavior, choices, etc.)

Introduction and Literature Review

The COVID-19 pandemic radically changed the normal operations of the transportation sector. Owing to the varying degrees of risk for virus contraction, the usage of each transportation mode varied. For instance, public transportation experienced the greatest decrease in travel demand and was generally perceived as the riskiest mode of travel (1, 2). This is because there was a higher chance of contracting the virus while sitting in an enclosed environment over an extended period with other passengers that could potentially be infected (3). As a result, travel via private car was generally perceived as the safest mode of transportation because there was almost no risk of virus contraction (2). Biking was generally perceived as significantly less risky than public transportation (1). This is because when biking an individual is not trapped

in an enclosed environment for an extended period, as they are when riding public transportation.

A series of previous studies looked at the factors affecting the bike sharing demand (4–10). These factors can be briefly classified into four groups as follows.

1. Weather-related factors: temperature, humidity, precipitation, and wind
2. Built-in environmental factor: bike sharing stations, transit stations, universities, bike lanes

¹Department of Civil and Environmental Engineering, Howard University, Washington, DC

Corresponding Author:

Md Sami Hasnine, mdsami.hasnine@howard.edu

3. Travel-related factors: trip purpose, trip departure time
4. Census-related factors: population, the median income

During the pandemic, the influence of the typical factors above was reduced, giving rise to new factors such as COVID-19 cases, news coverage of the virus, and lockdown/restriction policies (11–13). One factor that has received little attention in the research field thus far is the impact of vaccination on transportation. Only a few studies have analyzed this issue, and their findings are reported here.

In one of the studies, linear regression was used to estimate the influence of the COVID-19 vaccine on public transport and found that transit demand in the United States increased by 0.439% for every percentage increase in first-dose vaccination rates and 0.676% for every percentage increase in complete vaccination rates (14). Another study in London, UK, using a multinomial logit model (MNL) on survey data found that vaccination increased individuals' likelihood of riding transit systems (15). The final study used multiple indicator multiple cause (MIMIC) and probity modeling on survey data from the Greater Toronto Area, Canada, finding that being vaccinated did increase current public transit usage, but the estimates were lower than that of other COVID-19 risk perception factors, indicating that vaccination was not the leading influence of transit usage (16).

The main contribution of this research to the existing literature is summarized as follows.

1. Most of the previous studies investigated the impact of vaccination on transit mode choice only. This study contributes to the existing literature by investigating the impact of COVID-19 vaccination on bike sharing only.
2. This study combined a series of open-source data sets, such as the New York State Health Department, Google Community Mobility Reports, Google Trend, Apple Mobility Trends, National Oceanic, Atmospheric Administration (NOAA), and Citi Bike data. This data fusion helped to explore how a series of factors affected bike share demand during COVID-19 in the NYC area.
3. Previous literature only modeled the impact of vaccination independent of other influences, not allowing comparison. This study shows a detailed comparison between the pre-vaccine and post-vaccine periods.
4. While this research topic may appear insignificant because the COVID-19 pandemic is seemingly near an end, there is the risk of an outbreak of other COVID-19 variants, and it is estimated that

there will be more global pandemics and health crises in the future (14). Therefore, thoroughly understanding how COVID-19 impacted transportation will help us to prepare for these health crises and to mitigate their effects on society. The pandemic did not only have a one-directional effect on transportation, but transportation also significantly shaped the pandemic (15). For example, public transportation usage significantly contributed to the initial outbreak of COVID-19 in NYC (16–20). Understanding the impact of COVID-19 and vaccination on transportation can help policy makers make informed decisions to control the spread of future viruses.

Methods

Data Aggregation

The bike sharing data used for this study was from Citi Bike, a privately owned bike sharing service in the NYC and New Jersey Area. Since 2014, Citi Bike has made its bike sharing data for every trip publicly accessible. The predictor variables for bike sharing demand (Figure 1) were from the New York State Health Department, Google Community Mobility Reports, Google Trend, Apple Mobility Trends, NOAA, and the Oxford Covid-19 government response tracker (OxCGRT). The data included in this study were chosen according to the following criterion.

1. The data are available at the city-wide level
2. The data are available at daily or weekly intervals
3. All data points are available from February 15, 2020, to January 27, 2022
4. The underlying variable is known to affect decision making according to previous literature

Data Processing

Citi Bike. Citi Bike is a bike sharing platform located in the NYC area that has provided open access to all bike sharing trips since 2014 (21). Citi Bike provides the following information: the start and end date and time; the start and end location; the type of ride if the rider was a member or not; and other information. There is no option for filtering the data by location; therefore, researchers typically reverse the geocode based on the start station's longitude and latitude to determine if the bike trip was in Manhattan or Brooklyn. The data were separated, then aggregated daily based on the start trip date.

Citi Bike provides limited information about the types of bikers, such as user type (customer = 24-h pass or 3-day pass user; subscriber = annual member), gender

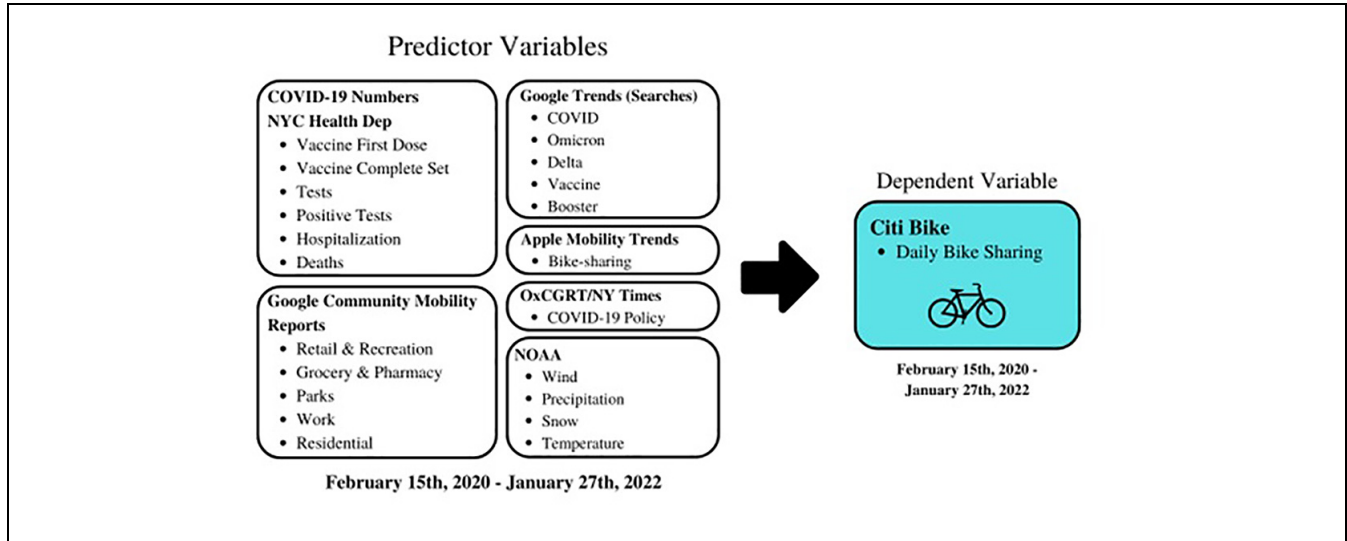


Figure 1. Data sources and predictor variables for Citi Bike data.

(zero = unknown; 1 = male; 2 = female), and year of birth. Citi Bike shares gender information from February 2020 to January 2021. It is found that 59.96% of Citi Bike users are male, and 28.73% of them are female. Around 11.31% of users declined to reveal their gender. For user type, 23.79% of users are 24-h pass or 3-day pass users. Of users, 76.21% are annual members (subscribers). The mean age of the users is 40.813, and the standard deviation is 12.4.

New York State Health Department. The New York State Health Department website provided open access to daily COVID-19 vaccination numbers as well as the number of COVID-19 tests, positive tests, hospitalizations, and deaths, at the county level (22). A data set was extracted for New York County (Manhattan) and Kings County (Brooklyn). The number of tests, positive tests, deaths, and hospitalization data were already expressed as a daily total for the time series model. Vaccination data had two variables: first dose and series complete. First dose refers to the number of individuals with one dose of an mRNA vaccine, while series complete refers to the number of individuals fully vaccinated either by one dose of the Johnson&Johnson vaccine or two doses of the mRNA vaccine.

Google Community Mobility Reports. As a result of the COVID-19 pandemic, Google has released open access to data concerning the changes in visits to specific locations (23). Google has classified these locations into the following categories: work; parks; retail/recreation; transit; residential; and grocery/pharmaceutical. Residential data are measured differently, though; rather than measuring the number of visits to residential locations, the

amount of time spent at home is measured. Google contrasts these numbers daily against the baseline (0%) mobility, which was the median value from January 3 to February 6, 2020.

Apple Mobility Trends Reports. Apple granted open access to navigation request data generated through Apple Maps (24). Apple has been anonymously counting the number of times its users request navigation assistance for car rides, transit trips, and walking. Apple Mobility Trends used January 13, 2020, as the baseline count (100%, standardized to 0% to match Google’s baseline). Each day the number of navigation requests is compared with this baseline number. The mobility trends can be filtered at a county level, so data were collected for both New York County (Manhattan) and Kings County (Brooklyn). The data were normalized so that the baseline was 0% to match the scale of the Google Community Mobility Reports’ data.

Google Trends. Google Trends provides access to searches made on Google’s search engine (25). It does not give the total number of searches but rather a relative ranking from 0 to 100 based on the popularity of a word or topic over a period. Google Trends grants an option for users to search for specific words or topics. Word searches only account for searches using the specific word entered, while topical searches account for words related to that topic chosen by Google. In this study, topical searches were done. Google Trends also allows users to specify results by location, time range, and other factors. The topics searched for were: “Coronavirus disease 2019,” “Coronavirus Delta variant,” “Omicron,” “COVID-19

vaccine,” and “Booster dose.” The “New York, NY” location option was selected and used for modeling travel demand in both Manhattan and Brooklyn.

NOAA. NOAA provides weather data, including wind, precipitation, snowfall, and temperature (26). Data were collected only from John F. Kennedy (JFK) airport, New York, because it was the closest location to Manhattan and Brooklyn without missing segments of data for all the variables. There were limited data points for the snowfall variable; therefore, its impact on transportation numbers may not be accurately captured.

OxCGRT. The OxCGRT has provided a list of different COVID-19 policies, along with maximum scores depending on the influence of the policy (27). *The New York Times* provided an article containing a timeline of the major COVID-19-related policies released over 2 years (March 2020–March 2022) (28). Based on the OxCGRT, scores were assigned to each of the policies. Before any COVID-19 policy was implemented, the score for each policy was 0 (indicating no policy in place). On the date when a policy was implemented (according to *The New York Times*), a score was added to the corresponding OxCGRT category. Likewise, when the policy was removed, the score was subtracted. The scores were aggregated on a daily level. In this study, various policy categories were used as a predictor variable as well as the cumulative policy score (the sum of the scores for every policy implemented on a given day). The policy categories used as predictors (and their maximum scores) were: School closing: 3; Workplace closing: 3; Cancel public events: 2; Restrictions on gathering: 4; Close public transit: 2; Stay at home: 3; Public information campaign: 2; and Facial coverings: 4.

The data were further processed into four scenarios, both Manhattan and Brooklyn, pre-vaccine (February 15, 2020–December 14, 2020) and post-vaccine (December 15, 2020–January 27, 2022) (Figure 2). December 15, 2020, was chosen as the post-vaccine period because the first record of vaccination from the vaccination data set was on December 15, 2020 (taken from the New York State Health Department). Various online articles show that the first vaccine was given on December 14, 2020. The New York State Health Department’s data set may have missed the record on December 14, 2020.

Modeling Approach

ARIMAX Time Series Analysis

The ARIMAX model has the following structure (where y_t is the current bike demand):

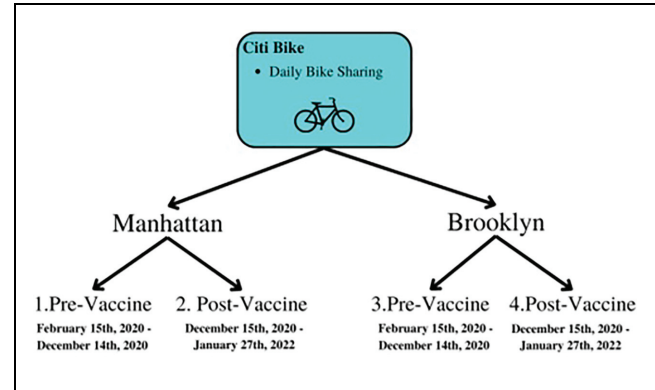


Figure 2. Four modeling scenarios: Manhattan and Brooklyn (pre- and post-vaccine).

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \delta_j \varepsilon_{t-j} + \sum_{k=1}^n \beta_k X_k + \varepsilon_t \quad (1)$$

- $\phi_i, \delta_j, \beta_k$ = parameters for the autoregressive, moving average, and predictor terms respectively
- p, q = lag term in the number of days for the autoregressive and moving average terms respectively, and n is the number of exogenous regressors term
- y_{t-i} = the previous period’s bike demand, ε_{t-j} = the previous period’s error unexplained by the model and X_k = predictor variables
- ε_t = error term, unexplained by the model

The autoregressive term uses previous demand as a predictor for the current demand; the number of previous demands used depends on p . The moving average term uses previous error (unexplained variance) as a predictor for the current demand and the number of previous errors depends on q . ARIMA models cannot model data with trends (a steady overall increase or decrease in demand) or seasonality (patterns in demand that occur weekly, monthly, or consistently over any given period). Data with trends and seasonality are called non-stationary and require differencing. This is what the “I” term in ARIMAX refers to. The Augmented Dickey-Fuller (ADF) test can be applied to the dependent variable to determine stationarity. If the data fail the ADF test, they must be differenced until they are stationary. Since ARIMAX models have exogenous predictor variables that follow linear regression, the correlation among the predictors must be taken into consideration because it threatens multicollinearity. In this study, the threshold for correlation was 0.75 (29, 30). Variables were removed to ensure that there was no correlation above 0.75 between any two variables. Therefore, all the variables in Figure 1 were not necessarily included in all the models.

The correlation matrices for the Manhattan models are shown in Tables 1 and 2.

Bayesian Times Series Analysis

A typical structural time series model has the following structure (31):

$$y_t = \mu_t + \tau_t + \sum_{k=1}^n \beta_k X_k + \varepsilon_t \quad (2)$$

where y_t is the observed bike demand, τ_t is the seasonal component, $\beta_k X_k$ is parameters and predictor variables and ε_t error term (normally distributed). The trend component reflects a steady increase or decrease in the data over the given period and is defined as

$$\mu_{t+1} = \mu_t + v_t \quad (3)$$

where v_t is the error term (normally distributed). The seasonal component reflects patterns repeating on a daily, weekly, monthly, etc., basis. The seasonal component is

$$\tau_{t+1} = - \sum_{s=1}^{S-1} \tau_{t-s} + w_t \quad (4)$$

where S is the seasonal division where $S = 52$ and $S = 12$ are weekly and monthly seasonality respectively. In this study, the seasonal component was set bi-annually $S = 2$. While this is not intuitive, it produced a lower MAPE than a daily, weekly, monthly, or quarterly seasonality. τ_t contains dummy variables for $S - 1$ number of seasons and w_t is the error term, normally distributed. In our study, Bayes theorem is applied as defined in the following equation (32):

$$H(\beta, \sigma^2 | Y_t) = \frac{F(Y_t | \beta, \sigma^2) * P(\beta, \sigma^2)}{F(Y)} \quad (5)$$

and the posterior distribution can be simplified as

$$H(\beta, \sigma^2 | Y_t) \propto F(Y_t | \beta, \sigma^2) * P(\beta, \sigma^2) \quad (6)$$

where $P(\beta, \sigma^2)$ is a prior distribution of estimates for the β parameters based on previous knowledge. In many cases as well as in this study, there was no prior information concerning the predictors used to create informed prior distributions. Therefore, the prior distributions followed a spike and slab normal distribution where: $\mu = 0$ and $\sigma^2 = \sigma_{y_t}^2$ as specified by the Bayesian structural time series-“BSTS” and “spike.slabs.prior” packages in R. $F(Y_t | \beta, \sigma^2)$ is a likelihood function of the bike demand based on the data for each predictor variable (33, 34). $F(Y_t | \beta, \sigma^2) * P(\beta, \sigma^2)$ is proportional to the posterior distribution. The mean of the posterior distribution can

be taken to obtain estimates for each parameter. The estimates can be standardized by multiplying the estimate by the standard deviation of the predictor and dividing it by the standard deviation of the dependent variable.

Model Evaluation

The accuracy of predictor variables used in both the ARIMAX and Bayesian models was determined using the MAPE, which has the following formula (35):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_t - y_t|}{y_t} \times 100 \quad (7)$$

where the term n refers to the sample size, y_t is the actual value, and \hat{y}_t is the models' estimated value. MAPE is the average deviation of the models' demand from the actual demand. MAPE values < 10 indicate an accurate forecast, 10 to 20 indicate a good forecast, 20 to 50 indicate a reasonable forecast, and > 50 indicate an inaccurate forecast (35, 36). In this study, we are looking for the impact of the predictor variables on bike demand. If the estimated model is accurately representing the actual bike demand, then the coefficient for the predictor variables (β) is accurately representative of the impact of these predictors.

Results and Discussion

ARIMAX Times Series Analysis

In this section, the results from the ARIMAX time series analysis are analyzed and discussed for Brooklyn and Manhattan, pre- and post-vaccine (Table 3). The significance level in this study was $\alpha = 0.05$. All statistically insignificant variables (below the critical values) were removed in the modeling process (except for complete vaccination numbers). The parameters or coefficients (β) were standardized for the predictor variables so that they could be compared. In Table 3, “AR” and “MA” refer to the autoregressive and moving average terms respectively.

Manhattan. Before the vaccine roll-out ($\beta = -0.18$, $t = -8.23$), increased precipitation decreased bike sharing demand while increased temperature led to increases in bike sharing ($\beta = 0.20$, $t = 3.68$). Work-related travel ($\beta = -0.09$, $t = -2.63$) led to decreased bike sharing demand while park-related activity was the leading influence for bike travel ($\beta = 0.66$, $t = 14.91$). This probably indicates that work-related travel was not a primary motivation for bike sharing; rather, it seems that leisurely travel (e.g., to parks) may have been the main reason for bike share usage. The negative β for work seems to

Table 3. ARIMAX Time Series Results

Variables	Manhattan: pre-vaccine ARIMAX (1,1,2) ^a		Manhattan: post-vaccine ARIMAX (1,0,0) ^b		Brooklyn: pre-vaccine ARIMAX (1,0,0) ^c		Brooklyn: post-vaccine ARIMAX (5,0,0) ^d	
	β	t	β	t	β	t	β	t
Complete vaccination	-	-	0.04	0.94	-	-	0.11	3.20
COVID-19 tests	-	-	0.24	4.92	-	-	-	-
COVID-19 cases	-	-	-0.31	-4.26	0.34	5.76	-	-
COVID-19 deaths	-	-	0.15	2.15	-	-	-	-
Wind	-	-	-0.10	-3.18	-	-	-0.05	-2.36
Precipitation	-0.18	-8.23	-0.12	-3.91	-0.16	-6.69	-	-
Snow	-	-	-0.43	-9.15	0.31	8.60	-0.32	-9.57
Temperature	0.20	3.68	0.23	3.56	-0.50	-4.97	0.09	2.09
Work	-0.09	-2.63	-	-	-	-	-0.32	-4.71
Park	0.66	14.91	0.30	5.83	-	-	-	-
Transit (Apple)	-	-	-	-	0.20	2.37	-	-
Retail/Recreation	-	-	-	-	0.23	2.06	0.19	2.64
Grocery trip	-	-	-	-	0.26	5.38	0.17	3.46
Residential activity	-	-	-	-	-0.99	-9.97	-0.56	-7.00
Policy score	-0.19	-2.67	-	-	-	-	-	-
Gathering restriction	-	-	-	-	0.46	6.50	-	-
Public transit closure	-	-	-	-	0.10	2.00	-	-
Searches: COVID-19	-	-	-	-	-0.16	-3.23	-	-
AR1	0.99	78.52	0.34	6.48	0.16	2.67	0.35	6.36
AR2	-	-	-	-	-	-	0.03	0.45
AR3	-	-	-	-	-	-	0.00	0.00
AR4	-	-	-	-	-	-	-0.03	-0.44
AR5	-	-	-	-	-	-	0.07	1.21
MA1	-1.90	-58.79	-	-	-	-	-	-
MA2	0.91	28.03	-	-	-	-	-	-
Intercept	-	-	7.15	9.19	7.65	86.02	8.68	37.30

Note:

^aMAPE = 1.31, sigma squared = 0.0343, AIC (Akaike Information Criterion) = 148.29, BIC (Bayesian Information Criterion) = 114.9, likelihood = 83.15.^bMAPE = 4.63, sigma squared = 0.2505, AIC = 609.01, BIC = 657.2, likelihood = -292.5.^cMAPE = 1.72, sigma squared = 0.0383, AIC = 113.65, BIC = 61.66, likelihood = 70.82.^dMAPE = 3.44, sigma squared = 0.1193, AIC = 307.62, BIC = 367.86, likelihood = -138.81.

indicate that the more individuals were engaged in work-related travel, the less they were likely to use bike sharing platforms. Other Google Community Mobility Reports variables (Retail/Recreational travel, Grocery trip, Residential activity) were removed because of high correlation threats and the Apple Mobility Transit variable was statistically insignificant. The COVID-19 Policy score variable negatively influenced bike sharing demand, indicating that stronger COVID-19 policies all led to decreases in bike sharing.

After the vaccine roll-out in Manhattan, complete vaccination ($\beta = 0.04$, $t = 0.94$) positively influenced bike sharing demand, but the t -statistic was below the critical value, indicating that vaccination was not significant. The number of COVID-19 tests ($\beta = 0.24$, $t = 4.92$) and deaths ($\beta = 0.15$, $t = 2.15$) led to increases in bike sharing demand, while the number of COVID-19 cases ($\beta = -0.31$, $t = 4.26$) led to decreases in bike sharing. These results seem contradictory but are justifiable. The COVID-19 tests variable refers to the number of COVID-19 tests taken, regardless of the result. If an individual received a positive test, then they would be less inclined to travel, which is why the COVID-19 cases variable had a negative β value. On the other hand, if an individual's test was negative, then they would be confident that they do not have COVID-19 and would be more inclined to travel. Therefore, if more people were receiving negative test results than positive test results, then it makes sense that the COVID-19 tests variable would have a positive β . The COVID-19 deaths variable's β was probably positive because individuals were moving away from public transportation as COVID-19 was breaking out (40). Therefore, individuals were moving toward modes that were perceived as less risky, such as travel by car, walking, and biking. Stronger wind speed ($\beta = -0.10$, $t = -3.18$), increased precipitation ($\beta = -0.12$, $t = -3.91$) and increased snowfall ($\beta = -0.43$, $t = -9.15$) all led to decreases in bike sharing demand. Snowfall is the strongest influence on bike sharing demand and wind is the weakest influence of the three. Increased temperature ($\beta = 0.23$, $t = 3.56$) led to increases in bike sharing as well. Travel to parks ($\beta = 0.30$, $t = 5.83$) was also a significant influence for bike sharing, similar to the pre-vaccine model.

Brooklyn. Before vaccination, COVID-19 cases led to increases in bike sharing numbers in Brooklyn ($\beta = 0.34$, $t = 5.76$). The β value is opposite that found in the Manhattan: post-vaccine model but is still justifiable.

Brooklyn is a lower-income borough with more individuals dependent on transit during the pandemic since there are more individuals working essential jobs that require in-person work, and it has a denser population

(37–39). Therefore, increasing COVID-19 cases during the pandemic (pre-vaccine) probably led to a larger shift away from public transit to bike sharing in Brooklyn than in Manhattan. Like Manhattan, before the vaccine roll-out, increased precipitation in Brooklyn ($\beta = -0.16$, $t = -6.69$) led to decreases in bike sharing while higher temperature ($\beta = 0.31$, $t = 8.60$) led to increases. Also, similar to Manhattan: pre-vaccine, work-related travel in Brooklyn was a negative influence on bike sharing ($\beta = -0.50$, $t = -4.97$). Unlike in the Manhattan models, in the Brooklyn models, Retail/Recreational travel, Grocery trip, and Residential activity variables did not threaten high multicollinearity and were included. Also, the Transit variable was statistically significant, unlike in the Manhattan model. It can be observed that transit ($\beta = 0.20$, $t = 2.37$), retail/recreational travel ($\beta = 0.23$, $t = 2.06$) and grocery trips ($\beta = 0.26$, $t = 5.38$) were all positive influences for bike sharing trips, indicating that grocery-related travel was the main travel purpose for bike sharing trips. It is important to note that transit trips can still be correlated with other trip purposes. For instance, an individual can bike to a transit station to go grocery shopping or retail shopping. Even if this was the case, the nature of the Google Mobility Trends data collection process would count both the transit and grocery or retail trip. Residential activity ($\beta = -0.99$, $t = 8.60$) was a negative influence on bike sharing, which makes sense because individuals are not home while using bike sharing platforms. Gathering restrictions ($\beta = 0.46$, $t = 6.5$) and public transit closures ($\beta = 0.10$, $t = 2.00$) both led to increases in bike sharing demand, which makes sense because individuals were resorting to biking as an alternative to regular activity that was canceled because of the pandemic (40). Lastly, increased searches for COVID-19 ($\beta = -0.16$, $t = 3.23$) led to decreases in bike sharing trips, suggesting that news of COVID-19 and fears of the virus spreading discouraged bike sharing.

After the vaccine, in Brooklyn, COVID-19 vaccination numbers ($\beta = 0.11$, $t = 3.20$) significantly impacted bike sharing demand, unlike in Manhattan: post-vaccine. Increased wind speeds ($\beta = -0.05$, $t = -2.36$) and snowfall ($\beta = -0.32$, $t = 9.57$) decreased bike demand, while higher temperatures ($\beta = 0.09$, $t = 2.09$) increased bike demand. Work-related travel ($\beta = -0.32$, $t = 4.71$) and residential activity ($\beta = -0.56$, $t = -7.00$) were still negative influences for bike sharing. Retail/recreational travel ($\beta = 0.2.64$, $t = 8.60$) and grocery trips ($\beta = 0.17$, $t = 3.46$) still led to increases in bike sharing demand. Post-vaccine, the COVID-19 policies of gathering restrictions and public transit closure no longer impacted bike sharing numbers (they were not statically significant). This could indicate that, post-vaccine, individuals were less sensitive to COVID-19 policies.

Table 4. Bayesian and ARIMAX Time Series Comparison Results

	Manhattan: pre-vaccine		Manhattan: post-vaccine		Brooklyn: pre-vaccine		Brooklyn: post-vaccine	
	Bayesian	ARIMAX	Bayesian	ARIMAX	Bayesian	ARIMAX	Bayesian	ARIMAX
MAPE (%)	13.37	1.31	13.82	4.63	22.96	1.72	18.38	3.44
Complete vaccination	–	–	0.02	0.04	–	–	0.001	0.11
COVID-19 tests	–	–	0.21	0.24	–	–	–	–
COVID-19 cases	–	–	0.00 ^b	–0.31	0.00 ^c	0.34	–	–
COVID-19 death	–	–	0.003	0.15	–	–	–	–
Wind	–	–	–	–0.10	–	–	0.00 ^h	–0.05
Precipitation	–0.14	–0.18	–0.16	–0.12	–0.16	–0.16	–	–
Snow	–	–	–	–0.43	–	–	–	–
Temperature	0.13	0.20	0.17	0.23	0.00 ^d	0.31	0.17	0.09
Work	–0.09	–0.09	–	–	–0.004	–0.50	–0.33	–0.32
Park	0.69	0.66	0.34	0.30	–	–	–	–
Transit (Apple)	–	–	–	–	0.31	0.20	–	–
Retail/Recreation	–	–	–	–	0.00 ^e	0.23	0.33	0.19
Grocery trip	–	–	–	–	0.12	0.26	0.002	0.17
Residential activity	–	–	–	–	–0.51	–0.99	–0.52	–0.56
Policy score	0.00 ^a	–0.19	–	–	–	–	–	–
Gathering restriction	–	–	–	–	0.28	0.46	–	–
Public transit closure	–	–	–	–	0.00 ^f	0.10	–	–
Searches: COVID-19	–	–	–	–	0.00 ^g	–0.16	–	–

Note:

^a5.73E-04.

^b–3.45E-04.

^c3.36E-05.

^d2.81E-05.

^e1.35E-04.

^f1.21E-05.

^g1.80E-05.

^h–1.26E-04.

Table 5. Various Diagnostics for the Bayesian Model

Model	Sample size	Simulations (MCMC)	Burn-ins	Potential Scale Reduction Factor (\hat{R})
Manhattan: pre-vaccine	303	1,000	491	0.997
Manhattan: post-vaccine	350	1,000	478	0.998
Brooklyn: pre-vaccine	303	1,000	405	0.997
Brooklyn: post-vaccine	350	1,000	497	0.998

Bayesian Times Series Analysis

In this section, the estimates for the predictor variables and MAPE from the Bayesian time series analysis with the estimates and MAPE from the ARIMAX time series analysis are presented (Table 4). The Bayesian model produces a distribution of values for estimates. In this case, the mean value of the distribution was taken as the estimate and the values were standardized for comparison. Table 5 shows various diagnostics for the Bayesian model, such as sample size, number of simulations (Markov Chain Monte Carlo-MCMC), Burn-ins, and \hat{R} .

Before commenting on the similarities and differences in Bayesian and ARIMAX results, it is important to note that the overall accuracy of ARIMAX models was

significantly better than the Bayesian models. All of the ARIMAX models' MAPE was below 5, indicating that the bike demand estimates from the model accurately represented the actual data. Three of the Bayesian models' MAPE fell between 10 and 20, a "good" representation of the actual data, and the pre-vaccine model in Brooklyn had a MAPE of 22.96, just falling outside of the "good" representation zone (35, 36). The estimates of the Bayesian time series model were still relatively similar to that of the ARIMAX model, thus cross-validating the ARIMAX results. We will look at the Manhattan models, both pre- and post-vaccine, to compare the findings. In the Manhattan: pre-vaccine period, except for the Policy score variable, the average deviation between the

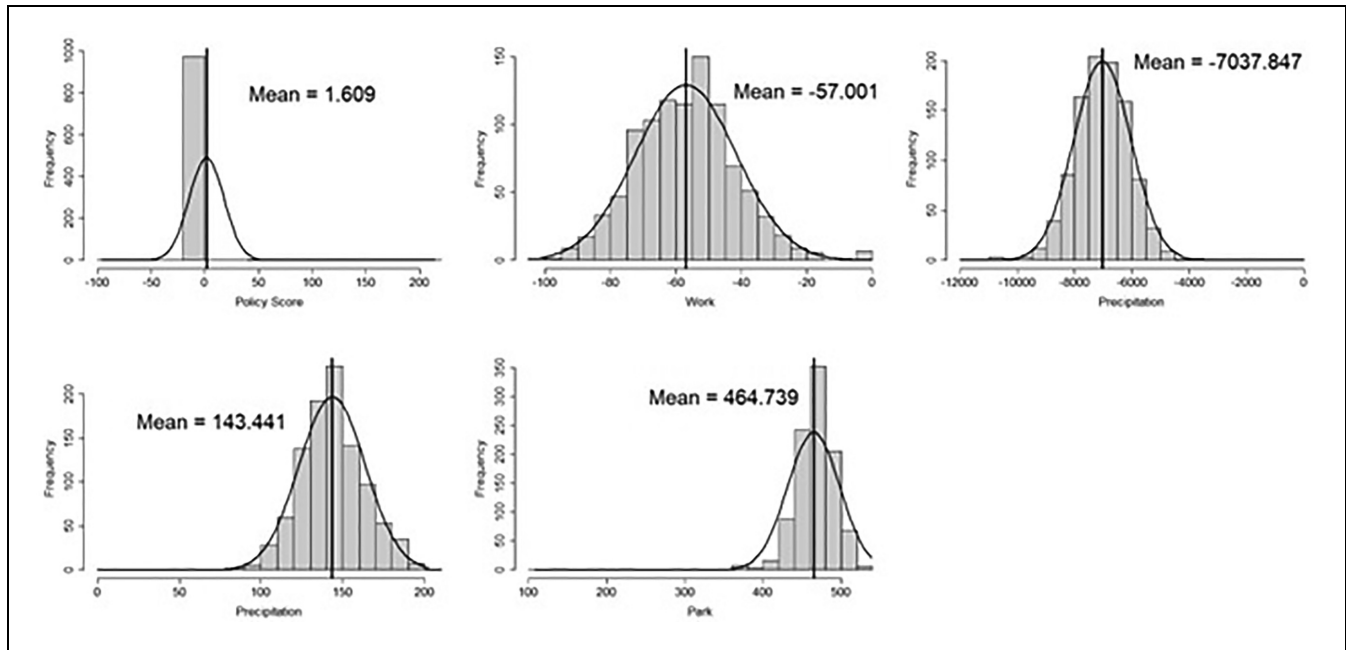


Figure 3. Distribution of Manhattan: pre-vaccine predictors.

estimates was about 16%. The estimate of the Work variables was -0.09 for both models; for the Park variables the estimates were 0.69 and 0.66 for the Bayesian and ARIMAX models respectively. In the Manhattan: post-vaccine models, most estimates were relatively close. It is also important to note that the sign (\pm) of every variable was the same, except for the Policy score variable in Manhattan: pre-vaccine and the Searches: COVID-19 variable in Brooklyn: pre-vaccine.

Overall, the Bayesian models were significantly underestimating some parameters but there were no significant overestimations. This is probably because the original prior distribution assigned to each of the parameters was a normal distribution with the mean set to zero. The original normal distribution around a zero mean had a strong pull on certain estimates. It can be seen in Figures 3 to 6 that the variables that were underestimated did not have a pronounced distribution, rather a spike and slab near zero.

The un-standardized estimates along with their posterior distributions are displayed in Figures 3 to 6.

Conclusion

ARIMAX and Bayesian time series analyses were performed on Citi Bike data in NYC to understand how the COVID-19 vaccination impacted bike sharing. The analyses were performed in two boroughs in NYC (Manhattan and Brooklyn), both during a pre-vaccine period (February 15, 2020–December 14, 2020) and a post-vaccine period (December 15, 2020–January 27, 2022). The predictors for bike sharing demand were

taken from the New York State Health Department (COVID-19 vaccination included), Apple Mobility Trends Reports, Google Community Mobility, Google Trends, NOAA, OxCGRT, and *The New York Times*.

The general findings were that COVID-19 vaccination did not significantly impact bike sharing in Manhattan but it did in Brooklyn. Various previous research found that the effects of the pandemic were perceived differently in urban (e.g., Manhattan) and suburban (e.g., Brooklyn) areas (41, 42). These studies indicate that various socioeconomic attributes (e.g., income), population density, and transit availability are the inherent reasons for the bike share demand variation in Manhattan and Brooklyn.

Even though vaccination was a statistically significant influence in Brooklyn, shopping (grocery and retail/recreational) were already strong influences for bike sharing pre-vaccine and post-vaccine, and they remained stronger influences than vaccination. Increases in COVID-19 numbers generally pushed individuals toward bike sharing in Brooklyn but away from bike sharing in Manhattan. Further research can be carried out on the socioeconomic differences in the two boroughs to better understand how COVID-19 and vaccination influenced transportation.

Other findings show that weather consistently played a key role in travel demand. In nearly every model, increased temperatures led to increased bike sharing demand, while increased precipitation and wind led to decreases in bike sharing demand. Lastly, there was a negative association between work-related travel and bike sharing both in Manhattan and Brooklyn, while

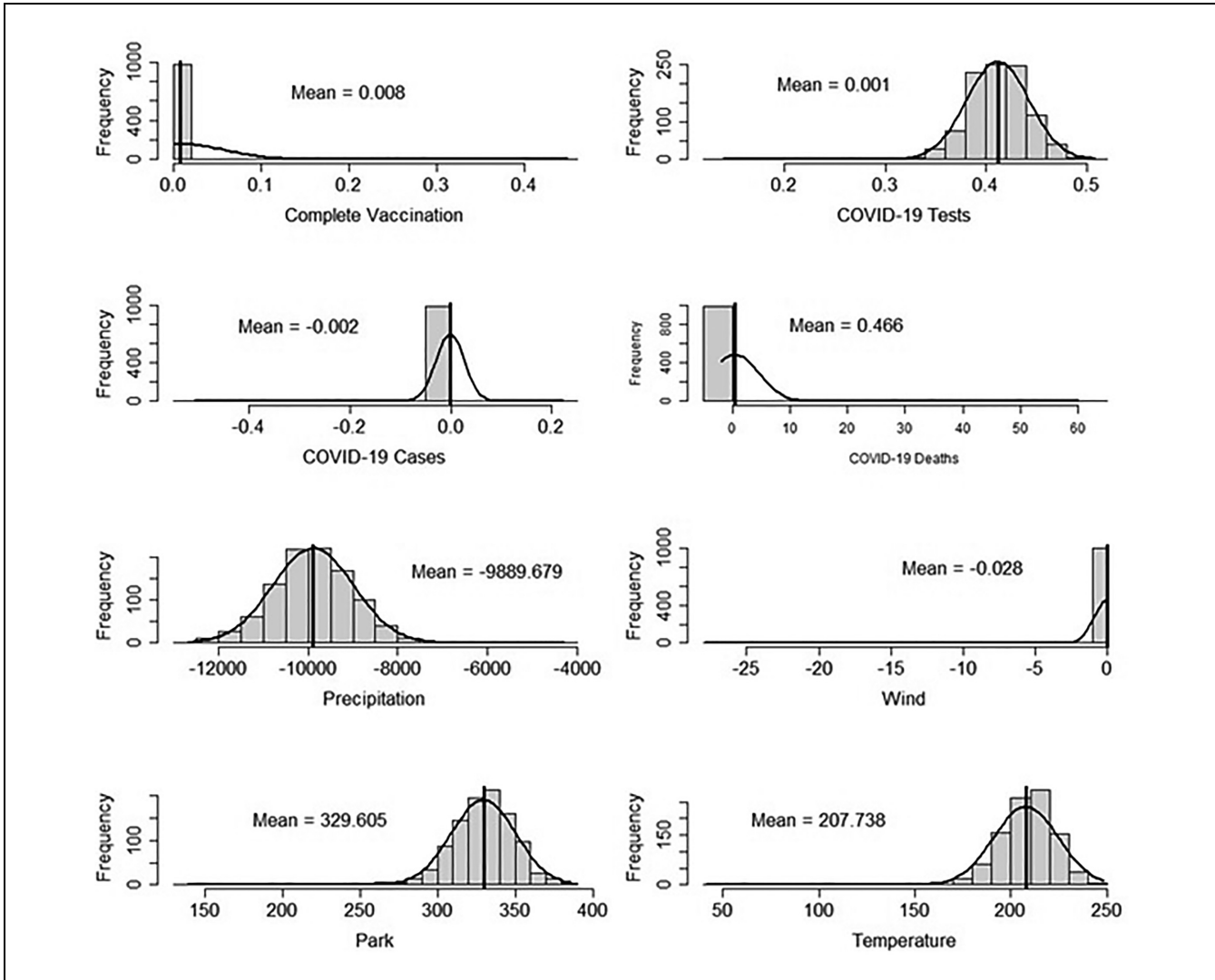


Figure 4. Distribution of Manhattan: post-vaccine predictors.

park-related travel was a positive influence for bike sharing. The average MAPE for the ARIMAX models was below 5, indicating the models were accurately representing the data; therefore, these findings can be presented with a high level of confidence. When comparing the standardized estimates from the ARIMAX and Bayesian models, the Bayesian estimates generally confirmed the findings of the ARIMAX model. The signs of nearly every parameter were the same, and the strength of the influence of most parameters were roughly the same. The Bayesian model significantly underestimated some parameters, indicating the modeling approach can be improved.

Limitations and Recommendations

One limitation in the Bayesian time series analysis was not incorporating previous knowledge about the

predictor variables in the prior distribution. Along with this, the BSTS package in R did not specify how to use other prior distributions except for the spike and slab distribution or the default distribution, which was a normal distribution around 0. It is recommended to explore further the BSTS package or use another package or program to specify different prior distributions.

Another limitation was in the inclusion process. Variables were removed if their correlation was above 0.75 with other variables. As a result, many variables that were aggregated had to be removed. Of the variables that remained, some were removed from the models because they were statistically insignificant. Variables because they were not statistically significant could have eliminated the correlation threats for other variables. For instance, in the Manhattan models, COVID-19 hospitalizations and COVID-19 cases counts had a high

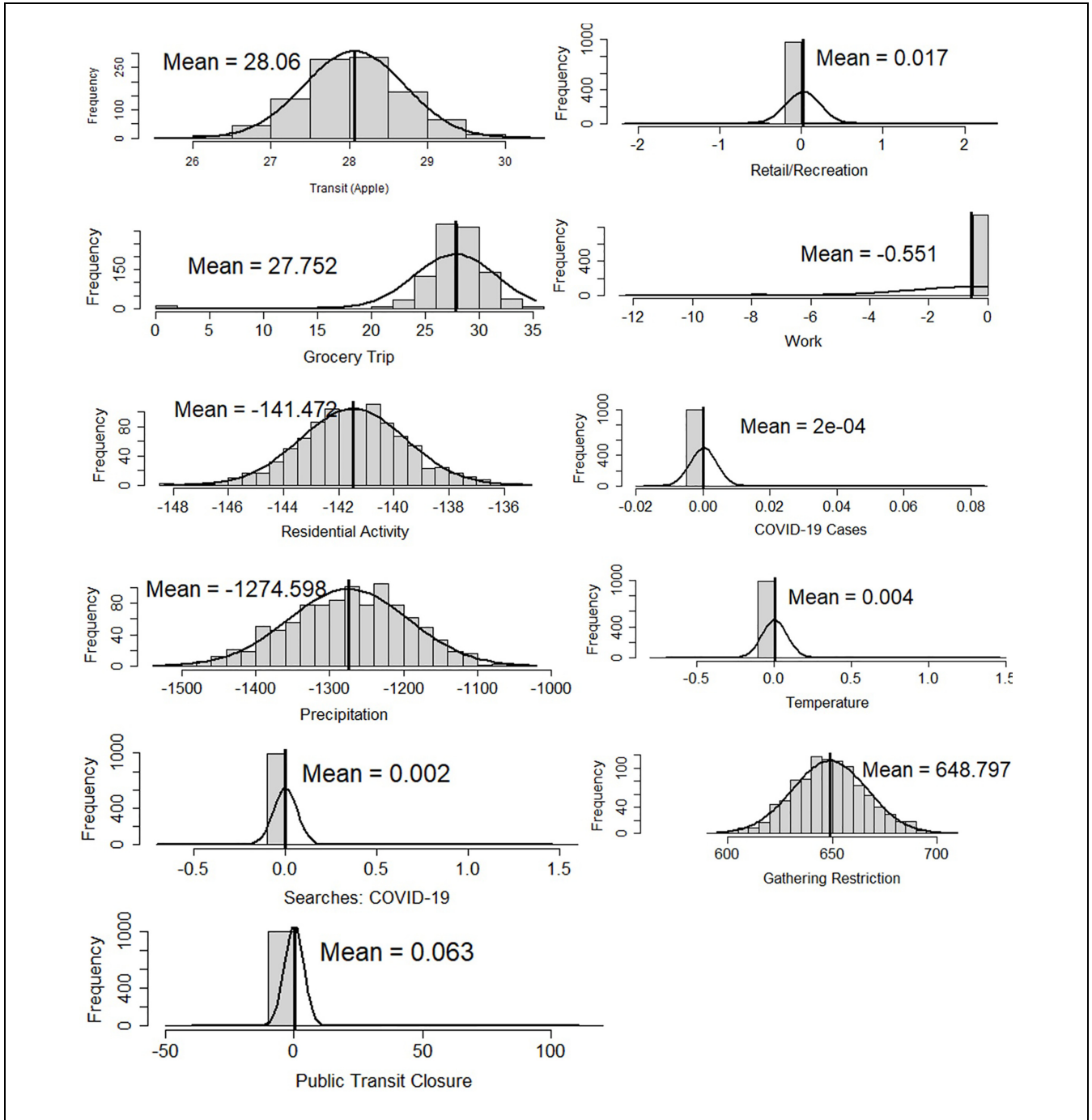


Figure 5. Distribution of Brooklyn: pre-vaccine predictors.

correlation. COVID-19 hospitalizations were removed but, when modeling, COVID-19 cases counts were not statistically significant and were removed. This would then allow for COVID-19 hospitalization to be included as a variable and modeled. The issue is that with a large number of variables, running through every combination of variable correlation before modeling, their significance after modeling, and re-running correlation is practically

impossible. Developing an algorithm that can automate this process will allow for the largest number of statistically significant variables included in each model.

This research explicitly looked at the bike share demand. However, it would be interesting to look at the relationship between the vaccination impacts on transit and bike share. Since transit users are less likely to own a car, they will be more likely to choose bike share because

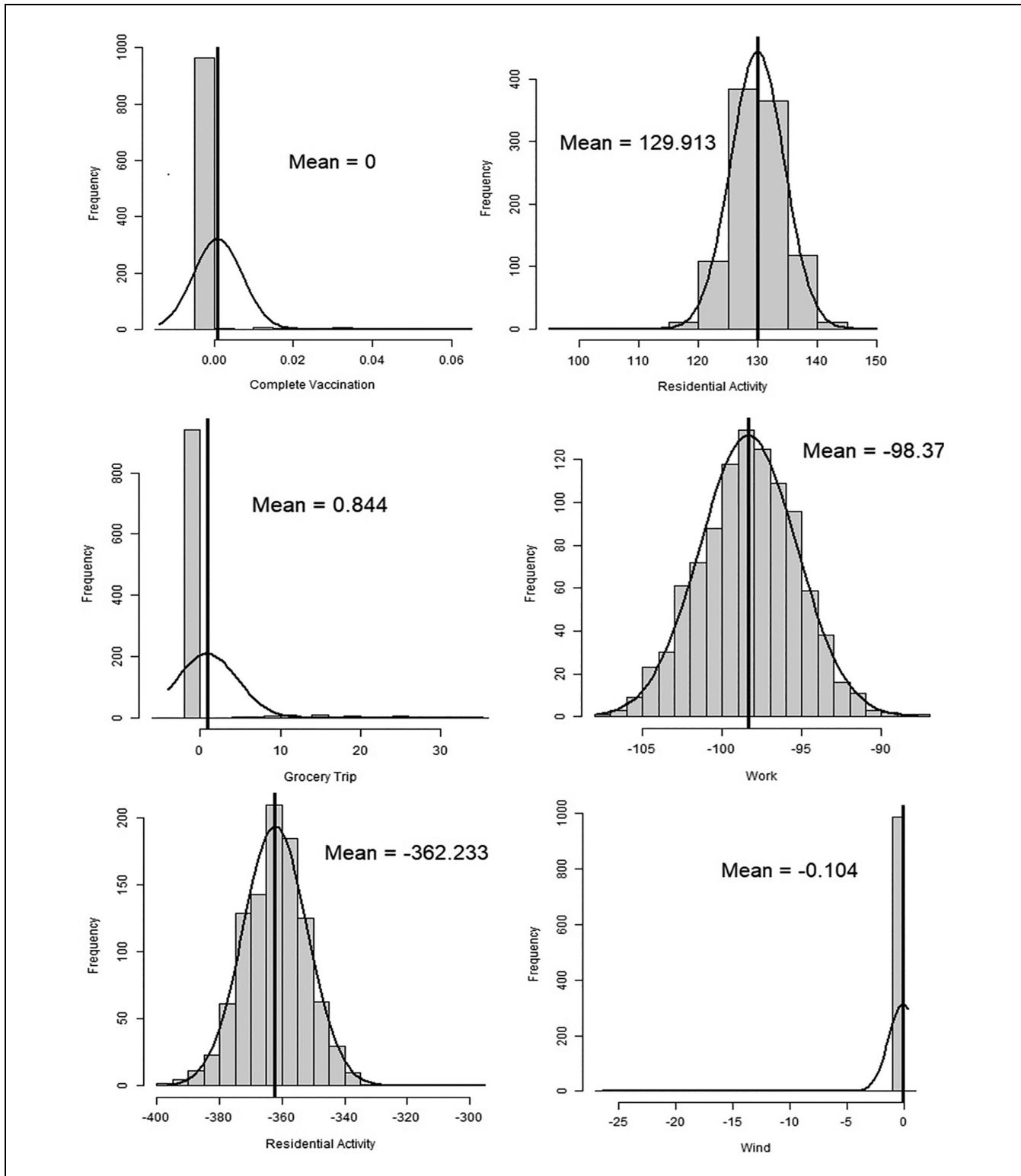


Figure 6. Distribution of Brooklyn: post-vaccine predictors.

of the fear of virus transmission. In addition, it would be interesting to explore the factors affecting this mode shift, such as trip purpose (recreation versus commuting), income (high versus low), and so forth. Given that this study uses a time series analysis over nearly 2 years, spatial factors that do not provide data changing regularly, such as population and built-in environmental factors, were not included.

In this study, we used the NOAA data set. NOAA provides weather data, including wind, precipitation, snowfall, and temperature. There are not separate weather stations in Queens and Bronx. Therefore, we decided only to compare Manhattan and Brooklyn.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: J. Drummond, M. S. Hasnine; data process: J. Drummond; analysis and interpretation of results: J. Drummond, M. S. Hasnine; draft manuscript preparation: J. Drummond, M. S. Hasnine. Author. All authors reviewed and approved the paper.


Declaration of Conflicting Interests


The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research is supported by Howard University New Faculty Start-up Funding and National Science Foundation (NSF) Graduate Research Fellowships Program (GRFP). All the interpretations are the authors' only.

ORCID iDs

Justin Drummond  <https://orcid.org/0000-0003-1134-1081>

Md Sami Hasnine  <https://orcid.org/0000-0003-4110-5047>

References

1. Barbieri, D. M., B. Lou, M. Passavanti, C. Hui, I. Hoff, D. A. Lessa, G. Sikka, et al. Impact of COVID-19 Pandemic on Mobility in Ten Countries and Associated Perceived Risk for All Transport Modes. *PLoS One*, Vol. 16, No. 2, 2021, p. e0245886. <https://doi.org/10.1371/journal.pone.0245886>.
2. Khaddar, S., and M. R. Fatmi. COVID-19: Are You Satisfied with Traveling During the Pandemic? *Transportation Research Interdisciplinary Perspectives*, Vol. 9, 2021, p. 100292. <https://doi.org/10.1016/j.trip.2020.100292>.
3. Gkiotsalitis, K., and O. Cats. Public Transport Planning Adaption Under the COVID-19 Pandemic Crisis: Literature Review of Research Needs and Directions. *Transport Reviews*, Vol. 41, No. 3, 2021, pp. 374–392. <https://doi.org/10.1080/01441647.2020.1857886>.
4. Abdullah, M., N. Ali, A. Bilal Aslam, M. Ashraf Javid, and S. Arif Hussain. Factors Affecting the Mode Choice Behavior Before and During COVID-19 Pandemic in Pakistan. *International Journal of Transportation Science and Technology*, Vol. 11, No. 1, 2022, pp. 174–186. <https://doi.org/10.1016/j.ijtst.2021.06.005>.
5. El-Assi, W., M. Salah Mahmoud, and K. Nurul Habib. Effects of Built Environment and Weather on Bike Sharing Demand: A Station Level Analysis of Commercial Bike Sharing in Toronto. *Transportation*, Vol. 44, No. 3, 2017, pp. 589–613. <https://doi.org/10.1007/s11116-015-9669-z>.
6. Martinez, M. S. The Impact Weather Has on NYC Citi Bike Share Company Activity. *Journal of Environmental and Resource Economics at Colby*, Vol. 4, No. 1, 2017, p. 12.
7. Guo, Y., and S. Y. He. Built Environment Effects on the Integration of Dockless Bike-Sharing and the Metro. *Transportation Research Part D: Transport and Environment*, Vol. 83, 2020, p. 102335. <https://doi.org/10.1016/j.trd.2020.102335>.
8. Faghieh-Imani, A., N. Eluru, A. M. El-Geneidy, M. Rabbat, and U. Haq. How Land-Use and Urban Form Impact Bicycle Flows: Evidence from the Bicycle-Sharing System (BIXI) in Montreal. *Journal of Transport Geography*, Vol. 41, 2014, pp. 306–314.
9. Faghieh-Imani, A., and N. Eluru. Analyzing Bicycle-Sharing System User Destination Choice Preferences: Chicago's Divvy System. *Journal of Transport Geography*, Vol. 44, 2015, pp. 53–64.
10. Faghieh-Imani, A., and N. Eluru. Incorporating the Impact of Spatio-Temporal Interactions on Bicycle Sharing System Demand: A Case Study of New York CitiBike System. *Journal of Transport Geography*, Vol. 54, 2016, pp. 218–227.
11. Chen, C., T. Feng, and X. Gu. Role of Latent Factors and Public Policies in Travel Decisions Under COVID-19 Pandemic: Findings of a Hybrid Choice Model. *Sustainable Cities and Society*, Vol. 78, 2022, p. 103601. <https://doi.org/10.1016/j.scs.2021.103601>.
12. Tokey, A. I. Spatial Association of Mobility and COVID-19 Infection Rate in the USA: A County-Level Study Using Mobile Phone Location Data. *Journal of Transport & Health*, Vol. 22, 2021, p. 101135. <https://doi.org/10.1016/j.jth.2021.101135>.
13. Chang, H.-H., and C. D. Meyerhoefer. COVID-19 and the Demand for Online Food Shopping Services: Empirical Evidence from Taiwan. *American Journal of Agricultural Economics*, Vol. 103, No. 2, 2021, pp. 448–465. <https://doi.org/10.1111/ajae.12170>.
14. Zhong, H., G. Wang, and T. Dai. *Wheels on the Bus: Impact of Vaccine Rollouts on Demand for Public Transportation*. Publication ID 3874150. Social Science Research Network, Rochester, NY, 2021.
15. Bansal, P., R. Kessels, R. Krueger, and D. J. Graham. Preferences for Using the London Underground During the COVID-19 Pandemic. *Transportation Research Part A: Policy and Practice*, Vol. 160, 2022, pp. 45–60. <https://doi.org/10.1016/j.tra.2022.03.033>.

16. Mashrur, S. M., B. Lavoie, K. Wang, P. Loa, and K. N. Habib. Joint Model of Transit Usage Frequency and In-Vehicle Safety Perception During the COVID-19 Pandemic. *Transportation Research Record*, 2022, pp. 1–21. [10.1061/1981221100241](https://doi.org/10.1061/1981221100241).
17. Michael, P. Statistics Say Large Pandemics Are More Likely Than We Thought. *Duke Global Health Institute*. <https://globalhealth.duke.edu/news/statistics-say-large-pandemics-are-more-likely-we-thought>. Accessed August 1, 2022.
18. Habib, Y., E. Xia, S. H. Hashmi, and Z. Fareed. Non-Linear Spatial Linkage Between COVID-19 Pandemic and Mobility in Ten Countries: A Lesson for Future Wave. *Journal of Infection and Public Health*, Vol. 14, No. 10, 2021, pp. 1411–1426. <https://doi.org/10.1016/j.jiph.2021.08.008>.
19. Harris, J. E. *The Subways Seeded the Massive Coronavirus Epidemic in NYC*. Publication 27021. Working Paper 27021. National Bureau of Economic Research, Inc, 2020. <https://doi.org/10.3386/w27021>.
20. Zhao, S., Z. Zhuang, J. Ran, J. Lin, G. Yang, L. Yang, and D. He. The Association Between Domestic Train Transportation and Novel Coronavirus (2019-NCoV) Outbreak in China from 2019 to 2020: A Data-Driven Correlational Report. *Travel Medicine and Infectious Disease*, Vol. 33, 2020, p. 101568. <https://doi.org/10.1016/j.tmaid.2020.101568>.
21. Citi Bike. Citi Bike System Data. *Citi Bike NYC*. <http://ride.citibikenyc.com/system-data>. Accessed March 15, 2022.
22. New York State Health Department. COVID-19 Data in New York. Department of Health. <https://coronavirus.health.ny.gov/covid-19-data-new-york>. Accessed March 15, 2022.
23. Google. COVID-19 Community Mobility Report. <https://www.google.com/covid19/mobility?hl=en>. Accessed March 15, 2022.
24. Apple. COVID-19—Mobility Trends Reports. *Apple*. <https://www.apple.com/covid19/mobility>. Accessed March 15, 2022.
25. Google. Google Trends. <https://trends.google.com/trends/?geo=US>. Accessed March 15, 2022.
26. NOAA. Datasets. Climate Data Online (CDO). National Climatic Data Center (NCDC). <https://www.ncdc.noaa.gov/cdo-web/datasets>. Accessed March 15, 2022.
27. Hale, T., N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, et al. A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, Vol. 5, No. 4, 2021, pp. 529–538. <https://doi.org/10.1038/s41562-021-01079-8>.
28. New York Times. Two Years of the Pandemic in New York, Step by Awful Step. *The New York Times*, March 15, 2022.
29. Kumari, S. Multicollinearity: Estimation and Elimination. *Journal of Contemporary Research in Management*, Vol. 3, No. 1, 2012, pp.87–95.
30. Salih, A. M. Comparison Among Different Shrinkage Covariance Estimators Under Multicollinearity and High Dimensions Conditions. *Al Kut Journal of Economics Administrative Sciences*, Vol. 31, 2019, pp. 201–210.
31. Scott, S., and H. Varian. Predicting the Present with Bayesian Structural Time Series. *International Journal of Mathematical Modelling and Numerical Optimisation*, Vol. 5, 2014, pp. 4–23. <https://doi.org/10.1504/IJMMNO.2014.059942>.
32. Foley, D. A Bayesian Approach to Time Series Forecasting. *Medium*. <https://towardsdatascience.com/a-bayesian-approach-to-time-series-forecasting-d97dd4168cb7>. Accessed July 31, 2022.
33. Depaoli, S., S. D. Winter, and M. Visser. The Importance of Prior Sensitivity Analysis in Bayesian Statistics: Demonstrations Using an Interactive Shiny App. *Frontiers in Psychology*, Vol. 11, 2020, p. 608045.
34. Yabe, T., Y. Zhang, and S. V. Ukkusuri. Quantifying the Economic Impact of Disasters on Businesses Using Human Mobility Data: A Bayesian Causal Inference Approach. *EPJ Data Science*, Vol. 9, No. 1, 2020, pp. 1–20. <https://doi.org/10.1140/epjds/s13688-020-00255-6>.
35. Meade, N. *Industrial and Business Forecasting Methods*, Lewis, C.D., Borough Green, Sevenoaks, Kent: Butterworth, 1982. Price: £9.25. Pages: 144. *Journal of Forecasting*, Vol. 2, No. 2, 1983, pp. 194–196. <https://doi.org/10.1002/for.3980020210>.
36. Montaña, J., A. Palmer, A. Sesé, and B. Cajal. Using the R-MAPE Index as a Resistant Measure of Forecast Accuracy. *Psicothema*, Vol. 25, 2013, pp. 500–506. <https://doi.org/10.7334/psicothema2013.23>.
37. Das, S., A. Boruah, A. Banerjee, R. Raoniari, S. Nama, and A. K. Maurya. Impact of COVID-19: A Radical Modal Shift from Public to Private Transport Mode. *Transport Policy*, Vol. 109, 2021, pp. 1–11. <https://doi.org/10.1016/j.tranpol.2021.05.005>.
38. Hu, S., and P. Chen. Who Left Riding Transit? Examining Socioeconomic Disparities in the Impact of COVID-19 on Ridership. *Transportation Research Part D: Transport and Environment*, Vol. 90, 2021, p. 102654. <https://doi.org/10.1016/j.trd.2020.102654>.
39. United States Census Bureau. United States Census Bureau: QuickFacts. 2021. <https://www.census.gov/quickfacts/US>
40. Plisic, A. Increasing Physical Activity in Children and Adolescents During the COVID-19 Pandemic. *Doctor of Nursing Practice Final Manuscripts*, 2021. <https://doi.org/10.22371/07.2021.044>.
41. Chauhan, R. S., D. C. da Silva, D. Salon, A. Shamshiripour, E. Rahimi, U. Sutradhar, S. Khoeini, A. K. Mohammadian, S. Derrible, and R. Pendyala. COVID-19 Related Attitudes and Risk Perceptions Across Urban, Rural, and Suburban Areas in the United States. *Findings*, 2021. <https://doi.org/10.32866/001c.23714>.
42. Paydar, M., and A. Kamani Fard. The Hierarchy of Walking Needs and the COVID-19 Pandemic. *International Journal of Environmental Research and Public Health*, Vol. 18, No. 14, 2021, p. 7461.